

国家科学数字图书馆项目研究报告

项目类型：预研课题

项目名称：分布环境下信息系统的开放描述

项目编号：csdl2002-15

项目承担单位：国家科学数字图书馆项目管理中心

报告名称：数字图书馆建设的标准规范体系

报告编号：001

报告版本：完成稿

报告日期：2002年4月30日

报告人：张晓林（执笔）、曾蕾、李广建、冯英、刘炜

国家科学数字图书馆版权声明

本报告研究工作得到中国科学院国家科学数字图书馆项目资助，项目编号为 csdl2002-15。国家科学数字图书馆项目管理中心和报告人拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

未经报告人和国家科学数字图书馆项目管理中心明确书面许可，任何人不得将本作品复制、转载、或在电子信息系统上做镜像。个人或机构为了学习和研究的目的，按照“合理使用”原则及其限定可以使用本作品，但应注明真实作者和完整出处，并在明显地方标明“国家科学数字图书馆资助项目”的字样。

国家科学数字图书馆项目管理中心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。

项目组声明

“分布环境下信息系统的开放描述”项目组谨保证本作品中出现的文字、图片、声音、剪辑和文后引文等内容的真实性及可靠性，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。国家科学数字图书馆项目管理中心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。

本项目组授权国家科学数字图书馆项目管理中心作为发表单位，在其网站上发表此作品，同意国家科学数字图书馆项目管理中心向其他媒体推荐此作品。未经项目组和国家科学数字图书馆项目管理中心同意，其他媒体一律不得转载。在不发生重复授权的前提下，项目组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。

项目组负责人：张晓林

北京北四环西路 33 号 中国科学院文献情报中心（邮编 100080）

联系电话：010-82628347 电子邮件：zhangxl@mail.las.ac.cn

数字图书馆建设的标准与规范体系¹

张晓林²、曾蕾³、李广健²、冯英⁴、刘炜⁵

文摘：本文根据数字资源建设生命周期框架，介绍了国际上主流数字图书馆项目关于内容创建、对象描述、资源组织、服务和数字内容长期保存等方面标准规范的选择和规定，提出了国家科学数字图书馆项目应该建立的标准规范描述体系的基本要求。

关键词：数字图书馆，标准，规范，描述体系，国家科学数字图书馆

The Framework of Digital Library Standards and Guidelines

Zhang Xiaolin¹, Zeng Lei², Li Guangjian¹, Feng Ying³, Liu Wei⁴

Abstract: Based on a life cycle structure of digital resource development, this paper describes major digital library projects' requirements for and adaptations of standards and guidelines in content creation, object description, collection organization description, services, and preservation, and presents basic principles for such a framework for Chinese National Science Digital Library.

Keywords: Digital Library, Standards, Guidelines, Standard Framework, CSDL

1. 数字信息系统的标准规范体系

面对分布、异构、变化和开放的数字信息资源与服务环境，各类数字信息系统需要建立自己的标准与规范描述体系（Frameworks of Standards，以下简称“描述体系”），按照统一的原则、框架和基本方式，规定数字信息资源和服务系统建设应遵循的各个层次的标准与规范，从而支持在整个数字信息环境中有效使用、广泛获取和长期保存信息。

根据描述体系覆盖的标准与规范的范围，可以将它们归于两类：

(1) 数字信息资源建设的标准描述体系，对数字信息资源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，主要是在图书馆、博物馆、档案馆等领域，例如英国公共图书馆领域的 NOF/People's Network 项目标准与指南^[1]、英国分布国家电子资源项目（DNER）标准体系^[2]、加拿大文化在线项目（CCOP）标准与指南^[3]、美国 IMSL 数字资源建设指南框架^[4]、RLG/CMI 数字化指南^[5]、美国国会图书馆数字资源格式描述体系^[6]等。有些描述体系面向更大环境，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。例如在政府信息和电子商务领域，有英国电子政府互操作框架（e-GIF）^[7]和 ebXML 电子商务标准体系^[8]，其中 e-GIF 从信息系统角度将标准规范分为系统互联（Interconnection）、数据整合（Data integration）、信息获取（Information access）等三个方面，包括通信协议、安全机制、数据编码、数据标记、元数据、数据转换、数据交换格式等方面的标准。在图书情报领域，有英国 DNER 系统互联指南^[9]、美国亚利桑那州数字化项目指南^[10]和美国科罗拉多州数字化项目（CDP）指南^[11]，覆盖资源加工、元数据、版权管理、数字化资源选择、资源建设和使用政策等方面的标准或指南。

(2) 涉及数字信息资源建设某一方面的标准规范描述体系，尤其是对数字信息资源的描述、组织的标准与规范及其应用要求进行规定。这些体系涉及广泛领域，包括数字图书馆、

1 中国科学院文献情报中心（Library, Chinese Academy of Sciences）

2 美国肯特大学图书情报学院（School of Library and Information Science, Kent State University, USA）

3 中国高等教育保障体系管理中心（Administrative Center, Chinese Academic Library and Information System）

4 上海图书馆（Shanghai Library）

专业信息服务、科学数据、电子政务等，例如美国国会图书馆数字资源检索与互操作规范体系^[12]、RLG/CMI 描述指南^[13]、OhioLink 多媒体资源标准体系^[14]、加州数字图书馆数字图像标准^[15]、加州数字图书馆元数据与编码标准^[16]、美国 NSDL 元数据标准体系^[17]、UN/FAO 农业信息资源检索元数据框架^[18]、CEN/ISSS 元数据体系^[19]、INDECS 数字知识产权元数据框架^[20]、英国电子政府体系元数据框架^[21]、加拿大政府信息元数据框架^[22]等。

本文主要针对第一类描述体系展开分析，而在这类体系中，一般都根据自己的目的和覆盖范围，将数字资源或系统涉及的标准规范分为多个层次，形成标准规范结构体系。例如，

NOF 按照数字信息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。

综合上述结构，可以将数字信息资源建设涉及的标准规范分为内容创建 (Creation)、描述 (Description/Metadata)、组织 (Organization)、管理 (Management)、服务 (Service)、长期保存 (Preservation) 和项目建设 (Projects) 等。本文从数字信息资源建设角度，主要依据 NOF、IMLS、CCOP、RLG/CMI 体系，并参考其它描述体系，分别按内容创建、描述、组织、服务等层次介绍有关标准规范的规定，并将在另外的报告中讨论关于管理和项目建设的标准与规范。

2. 关于数字内容创建的标准规范

在数字资源建设中，数字内容包括由传统载体 (印本、图片、录音录像等) 数字化而形成的数字对象，或者是原生数字形态的内容对象 (例如直接的数字文本、数字摄像或数字录音文件等)。数字内容创建的标准规范涉及内容编码、内容对象格式、内容对象标识等方面。

2.1 内容编码

内容编码涉及具体数据内容的计算机编码形式和标记形式，是制约数字信息可使用性乃至可持续性的最基本条件。数字图书馆项目通常会要求资源内容在编码层次遵循基本的标准，例如以下方面标准：

(1) 基本编码标准，国际上普遍要求遵循 ISO/IEC 10646^[23]/UNICODE^[24]。在我国环境下，目前存在 GB2312-1980^[25]、GB13000-1993^[26]和 GB18030-2000^[27]标准，其中 GB18030 在 GB2312 基础上进行扩充，在技术上是 GBK (《汉字扩展规范 GBK1.0》) 的超集，已经是国家的强制性标准。GB13000-1993 是 ISO 10646-1 的等同标准，GB18030-2000 与它在字汇上兼容，通过代码映射表可以进行自由转换。

(2) 特殊信息编码，涉及数学符号和公式、化学符号、矢量信息、地理坐标等的编码，例如基于 XML 的开放标记语言，如 SVG (Scalable Vector Graphics)^[28]、SMIL (Synchronized Multimedia Integration Language)^[29]、MathML (Mathematical Markup Language)^[30]、GML (Geography Markup Language)^[31]、CML (Chemical Markup Language)^[32]等。

(3) 数字文献结构编码，涉及如何定义文献结构，普遍要求采用 XML DTD^[33]/XML Schema^[34]来定义文献结构，而且相关的文献模式定义应经过 XML 语法验证 (validated)。

2.2 数据格式

数据格式涉及文本、图像、音频、视频、多媒体等数据内容，需要解决的问题包括格式体系和格式标准。

(1) 格式体系指数字内容创建中需要多种承担不同责任的数据格式，通常包括：
保存格式 (Preservation/Archiving Formats)，作为长期保存格式 (有时又称原版格式)，要求保存原始数据形式 (例如图像、录音、录象等) 的内容及其表现，采取非压缩格式；
浏览格式 (Access/Viewing/Service/Reference Formats)，作为正常存储和显示的格式，要求保证视觉质量、又降低传输成本，可采用压缩格式，可从保存格式中派生；
预览格式 (Previewing Formats/Sampling Formats)，作为预览信息，提供粗略内容表现，可采用大压缩比的格式，可从保存格式或服务格式中派生。

上述格式体系主要针对数字图像而言，但其根据不同用途来建立多种相互关联格式的思想和实践对音频、视频内容等都有实际意义。

(2) 文本数据的格式标准涉及两种类型，作为文本文件、或作为图像文件。
作为文本文件时，描述体系要求采用 HTML、XHTML、XML (早期还包括 SGML 格式)，其中 XML 格式的定义须是经过验证的 XML DTD 或 XML Schema，用 XML 标记的文本数据在交换时应可用 HTML/XHTML 格式表现。在不能有效处理 HTML/XML 环境下，应采用纯 ASCII 格式或 CSV (Comma Separable Variable Format) 格式 (例如 DCMI DCSV^[35])。如果文本资源本身是专门格式文本 (例如 doc、rtf、ps 等)，在保证应用软件可获得性的同时，应提供将这些格式文本转换纯文本文件或 HTML/XML 格式文本的公开方法，形成可靠的数据迁移机制 (Data Migration) 以保证未来能把专用格式文本转换为开放格式文本。当然，有些领域规定 (或共同采用) 某种专门的文本格式，形成该领域的事实格式标准，例如数学和工程计算领域的 TeX/LaTeX 格式^[36]。

作为图像形式的文本数据可以采用 TIFF (Tagged Image File Format) 格式^[37]、JPEG (Joint Picture Expert Group) 格式^[38]或 PDF (Portable Document Format) ^[39]格式，但由于 PDF 并不是开放格式，有些描述体系 (例如 NOF) 规定如采用 PDF 时要建立开放数据迁移以保证可将 PDF 数据转换为开放格式数据。对纯黑白文本，也可以使用 GIF (Graphics Interchange Format) ^[40]格式扫描形成文本图像。

(3) 图像数据的格式标准涉及格式类型和分辨率，根据保存、浏览或预览格式而有不同要求，例如对保存格式，多数描述体系都要求用非压缩的 TIFF 格式，分辨率往往要求 600 dpi，但 CCOP 允许使用 PNG (Portable Network Graphics) ^[41]；对浏览格式，可采用 JPEG 或 SPIFF (Still Picture Interchange File Format) 格式^[42]；对预览格式，可采用 GIF 格式；对线图图像 (Line-drawings)，可采用 PNG 或 GIF。

(4) 视频数据的格式标准一般首选 MPEG 格式^[43] (但 CCOP 专门指出不应使用 MPEG-1 格式)，另外也可使用 Apple Quicktime^[44]、MS Real Video^[45]等专用格式。由于视频格式都存在压缩，因此数字视频数据的“保存格式”往往采用数字录象格式，例如 DV、DVCam、DVCPro、digiBeta 等格式。

(5) 音频数据的格式标准比较复杂，除了常推荐的 MP3 外，还有 WAV (Waveform Audio File Format) ^[46]、Apple Quicktime、MS Real Audio^[45]等。与视频情况类似，音频数据的“保存格式”采用数字录音格式，例如 CD-Audio (44KHz @ 16Bits)、DAT (44KHz @ 16Bits 或更高)、AIFF (Audio Interchange File Format) ^[47]等。

(6) 矢量数据的格式标准主要是 SVG，也有建议 VML^[48]的，另外业界的 Micromedia Flash^[49]也可能是可接受的格式，但类似于 PDF 格式，NOF 规定如采用它的话应建立数据迁移机制来保证将数据转换为今后出现的开放格式。

2.3 内容标识

内容标识方面的标准与规范主要涉及对数字对象唯一标识符，而“数字对象”可能是不同层级的内容对象，例如数字图像（扫描或原生的），由多个数字图像组合而成的数据文件（例如多页图书），由多个文本、图像、音频、视频等数据对象组成的多媒体数据文件（例如课件），这些数字对象的元数据记录，由多个数字对象组成的资源集合，等等。一般地，描述体系没有规定具体的标识符结构，而是对数字对象标识的原则予以规定：

（1）数字对象必须按照规范的命名体系用一个唯一标识符予以命名^[50-51]，标识符本身应是逻辑的、不与物理地址捆绑的、而且可以通过标识符解析系统转换为相应的物理地址。

（2）数字对象命名所采用的命名体系的规则应是公开和明确界定的，命名体系应遵从 IETF/URI 体系，应尽量采取标准或通用的标识符命名体系，例如 DOI^[52]、SICI^[53]/BICI^[54]、或 PURL^[55]等。如果自己建立命名体系，应保证命名体系名称（作为 NID）本身的可解析性和命名体系解析机制的正常运转。

（3）提供数字对象的资源系统应该能接受以唯一标识符形式提供的指令，并将唯一标识符准确地解析为自己的内部标识。

（4）如果资源系统因技术或其它原因不能加入或建立公共命名体系及其解析机制，应建立内部的数字对象标识规则（或文件命名规则），使其它系统能够利用这些规则来标识相关的数字对象，也支持参考文献链接等功能。

（5）作为大范围的数字信息服务系统，需要考虑多个唯一标识符系统的互操作。

（6）许多数字对象可能由多个数字对象组成、甚至是动态组成的，它们的链接与复用往往需要通过标识机制来支持，可借鉴 CDL/METS 标准^[56]和 ADL/SCORM 标准^[57]。

除了上述数据编码、数据格式和数字对象标识外，多数描述体系要求数字对象必须建立相应的元数据，并可通过数字对象唯一标识符将两者链接起来，有些描述体系（例如 IMSL 和 e-GIF）还建议数字对象有一定的验证机制，例如数字签名或数字水印。

3. 关于数字对象描述（元数据）的标准规范

元数据作为描述数字对象的数据，是所有数字信息资源建设项目的重要基础，因此也是描述体系的重点。一般来说，第 1 类描述体系往往规定对数字对象进行描述的原则和基本方法，而第 2 类描述体系则往往在一个具体范围内规定实际应用的元数据标准与规范。

3.1 元数据应用原则

描述体系都专门论述了元数据的应用原则，并在以下各点上形成共识：

（1）任何希望提供公共、长期和可靠服务的数字信息资源系统都应该编制关于数字对象的元数据；如果因为特殊原因没有或暂时没有编制数字对象的元数据，也应该提供关于资源集合的元数据。

（2）采用标准的或业界通用的元数据格式；有些描述体系（例如 CCOP）专门规定没有充分的合理理由，数字资源系统不要创建自己的元数据格式。

（3）所选择应用的元数据格式应适用于具体的资源类型和应用要求；尤其在美国，由于各个领域都存在各自的元数据格式，例如 TEI^[58]、GILS^[59]、FGDC/CSDGM^[60]、EAD^[61]、VRA^[62]、IEEE LOM^[63]等，甚至关于同类对象也有不同元数据格式，例如 MARC^[64]与 ONIX^[65]，这些格式往往针对不同的应用需要。因此，IMLS 鼓励各数字资源建设单位选择适合自己资源类型和应用任务的标准的或通用的元数据格式。与美国不同，欧洲和加拿大在承认各个建设单位应选择适用的元数据格式的同时，往往建议或要求采用某一元数据格式作为核心集。

（3）元数据应包括技术元数据，即关于数字对象创建、使用等的技术条件的数据，从而支持所描述的数字对象的长期保存及可能的仿真或迁移处理；元数据也应包括管理元数据，即关于数字对象使用过程中的存取权限、知识产权、保存控制等的的数据，从而支持对数

字对象的有效管理。当然，描述性、技术性和管理性元数据也许应通过开放链接方式组织在一起，以适应元数据交换、复用和动态定制等方面的要求。

(4) 元数据内容描述应使用标准的内容编码体系，包括主题或分类词表、资源类型、语种、国别或地区、日期或时期等，从而保障内容描述方式的标准化和描述内容的可交换。

(5) 元数据格式应支持互操作，这一方面体现在形成由格式定义、语义定义、概念集 (Ontology^[66]) 定义、标记语言定义、内容编码体系定义、应用规范 (Application Profiles) 定义等组成的定义链，一方面往往意味着所有定义应该是公开的和基于开放标准和开放语言的，再一方面往往要求元数据格式提供与其它通用格式的规范转换机制，尤其当所选用的格式不是标准格式时更应如此。

(6) 元数据本身也是数字对象，因此也可唯一标识和长期保存，也有它自己的管理数据，也应该提供相应的验证机制。

3.2 关于元数据标准的选择

描述体系的一个重要任务是规定或推荐具体的元数据标准，一些描述体系会根据不同资源类型分别规定不同的格式，另一些会按照统一的检索和交换需要来规定统一的核心格式及其扩展方式，还有一些则只是制定元数据格式选择原则，并不具体规定元数据格式。

(1) 部分描述体系允许使用多种元数据格式，根据不同的资源类型推荐多个格式，例如 OhioLINK 对它的 Digital Media Center 的资源格式规定：一般科学与技术资料，采用 Dublin Core (DC) 格式^[67]；人文科学/档案资料/音乐资料，采用 DC 格式；生命科学和医学资源，采用基于 DC 的扩展格式；地理信息资源，采用 FGDC/CSDGM 格式；艺术与建筑资源，采用 VRA Core 格式。

(2) 许多描述体系或系统推荐使用一种元数据格式作为核心格式，允许在核心格式基础上按规范方式进行扩展。例如 CCOP 规定所有项目或者直接使用 DC 格式、或者提供所使用元数据格式与 DC 之间的规范转换；CDP 等“地方性描述体系”也规定所有项目必须提供 DC 格式的元数据。又如，NSDL 规定将 DC 作为核心元数据格式，并通过复用 IEEE LOM 元数据格式中的若干元素对 DC 进行扩展，构成 NSDL 教育元数据，规定所有 NSDL 项目必须使用 DC 或扩展后的 DC，另外还可接受其它 8 种可利用现有转换模块与 DC 转换的元数据格式；另外在政府信息领域，英国 e-GMF (E-Government Metadata Framework)、加拿大 TBITS 39.1 (Government Online Metadata Standard)^[22]、澳大利亚 AGIL (Australia Government Information Locator)^[68]、欧盟 MIREG (Managing Information Resources for e-Government)^[69]等都规定在 DC 基础上构建政府信息元数据格式。

(3) 有的描述体系在不具体规定元数据格式、或在推荐一种核心元数据格式时，也可能对具体领域或资源类型的元数据提出不同要求，例如 CCOP 规定教育资源应能使用 IMS^[70]、CanCore^[71]元数据、并且建议各资源建设单位充分考虑与自己应用范围相关的元数据标准，NOF 建议在描述数字图像时考虑 NISO TMI^[72]等格式。

3.3 关于内容主题描述语言的选择

描述数字对象的元数据中都有内容主题描述元素，描述体系都要求使用规范主题词表来标引主题，以保证主题描述的规范性和一致性。

(1) 一般地，覆盖范围大的描述体系没有具体规定必须采用的标引词表，只是要求在描述数字对象时采用对应学科领域的标准词表，例如 CCOP、IMLS、NOF 等，但它们也可能对特定类别的数字资源主题描述提出应采用的词表，例如 CCOP 要求所有联邦政府项目在主题描述时应采用 Treasury Board 指南规定的词表^[73]。

(2) 部分描述体系根据不同主题领域或不同资源类型推荐或规定了多种词表，例如 OhioLINK 对其 Digital Media Center 资源规定：一般科学与技术资料，采用本学科的标准词表；人文科学/档案资料/音乐资料，采用 LC Subject Headings (一般主题标引)、Thesaurus for

Graphic Materials^[74] (图像元素标引)、Thesaurus of Geographic Names^[75] (地名标引); 生命科学和医学资源, 采用 Medical Subject Headings^[76] (一般主题标引)、Taxonomic Resources and Index to Organic Names^[77] (生物体标引)、Geographic Names Information Systems^[78] (地理名称标引); 艺术与建筑资源, 采用 Art and Architecture Thesaurus^[79] (主题标引)、Thesaurus for Graphic Materials (图像元素标引)、Union List of Artists Names^[80] (人名标引)。

(3) 部分描述体系规定在自己覆盖范围内采用统一的主题词表, 例如多数电子政务元数据描述体系。英国 e-GMF 规定将建立一个英国跨部门词表 (UK Pan-Government Thesaurus), 澳大利亚 AGLS 也要求采用统一政府词表。

(4) 值得注意的是 INDECS 元数据体系^[20], 系统分析了电子商务中知识产权保护所涉及的实体及相互关系, 并在此基础上建立了元数据词典, 明确定义了每个实体名称及其语义、实体间各种关系名称及其语义。这些名称可以用在描述知识产权交易对象、知识产品、交易文件、交易过程等的各种元数据格式中, 但它们都应该遵循由该词典定义的名称和语义, 从而促进相关元数据的互操作。从一定意义上, 这个关系体系已经建立了一种概念集基础。

实际上, 在选择元数据格式时我们面对的是一个非常复杂的环境, 创建和应用元数据的目多元化致使多种元数据格式存在并继续涌现, 而且也很难有任何一种格式能够满足所有需要。为了在充分满足各方面需要、充分利用现有元数据格式建设和元数据记录的情况下保障多种元数据格式的互操作, 许多领域已开始探索建设开放元数据体系, 通过规范的元数据继承、复用、扩展和转换机制来充分利用已有的各种元数据、同时保障不同元数据间的准确和高效转换, 我们应对此有足够重视。

4. 关于资源组织描述的标准规范

数字对象可能按照一定的主题、资源类型、用户范围、生成过程、使用管理范围等因素被组织在一起, 形成实际使用的资源集合 (Collections)。这些资源集合的形成是数字资源组织的过程, 对这个过程所组织的集合的描述、以及对这个过程本身的描述, 对于数字信息的检索和利用具有重要意义, 各个描述体系都在不同层次上规定了对资源组织的描述要求及相应的标准规范。

4.1 资源组织描述的发展与要求

由于前期数字图书馆建设的分散状态和图书情报领域对具体文献描述的传统关注, 资源组织过程及资源集合的规范描述被认为是一个本地化问题而没有得到重视, 直到众多数字图书馆建设项目不断涌现、一些大范围数字图书馆体系开始建立, 人们才开始提出资源组织本身的规范化和资源集合元数据的标准化, 并将其作为整个资源建设的一个重要任务和元数据体系的一个有机部分来考虑。资源集合描述可以有多个层次, 例如:

(1) 可对资源集合本身进行描述、形成一个关于资源集合本身的元数据记录, 往往涉及资源内容、资源建设与管理者、资源使用与管理条件、与其它资源集合的关系等方面的数据内容。这个层次的元数据主要支持对资源集合的发现。

(2) 进一步地, 可对资源集合的组织机制进行描述, 这些机制可以是简单的类别组织、频道划分、模块集合、或者是复杂的知识组织系统 (Knowledge Organization Systems, 包括分类法、主题词表、Site Maps 等)^[81]。这个层次的描述也是元数据, 可支持对资源集合的整合处理、集成与定制。描述结果 (元数据) 可以是文本、结构化文本、规范格式、甚至计算机可读形式。

(3) 再进一步地, 可对资源集合的管理机制进行描述, 例如对资源选择标准、资源使用政策、知识产权管理政策、隐私保护政策、资源长期保存政策等及其实施机制的描述, 这些描述形成管理机制元数据, 能够支持用户和其它系统有效的发现、选择和利用相应的资源集合。与组织体系描述数据类似, 对管理机制的描述结果可以是多种形式, 发展趋势是构建

规范的、结构化的和计算机可读的管理机制元数据。

(4) 再进一步地，可以对资源组织建设的过程、原则、方法及相应的标准规范进行描述，形成资源建设规范，指导资源建设。这一层次的描述虽然难以被归纳到传统的元数据中，也可能难以用标准语言来统一描述，但它对资源建设的重要性则不容置疑。

4.2 对资源组织的描述数据的要求

目前规范化工作较为成熟的是资源集合本身的描述，建立规范的资源集合描述元数据是大范围资源建设体系的一个基本要求，并往往提供一定机制来存储和检索这些元数据。例如，NSDL 规定，任何一个参加 NSDL 的资源项目应采用 DC 来描述自己的集合，并将该 DC 记录提交 NSDL 的 Metadata Repository 供公共检索；CDL 要求自己范围内的各个资源集合采用 EAD 来进行描述，并提供了一个 EAD 描述模板登记机制；CCOP 规定采用 RSLP CDS^[82] 来描述资源集合，而且 RSLP 记录采用 RDF 描述语言。其实，数字环境下的资源集合还包括网站、数据库、网络资源目录等形式，有关领域也开发了相应的标准规范来描述这些资源集合，我们也应给予必要关注。

对于数字图书馆建设来说，关于资源集合的组织机制和管理机制的规范描述（除了分类标引标准外）是一个新的领域，正在借鉴 W3、电子商务和其它领域的经验，开始考虑和试验相应的标准，例如知识组织系统方面的 VocML^[83]、XTM^[84]、Zthes^[85]、ISO TMF^[86]等和管理机制方面的 P3P^[87]、XACL^[88]、ODRL^[89]、PICS^[90]等。

关于资源组织过程的指导性规范已得到越来越多数字图书馆建设项目的重视，各种形式的指南已经存在，但在早期这些指南更多地是关于数字化过程及其技术标准、设备规格、工作流程、质量控制、人员培训等，例如 RLG/DLF Guides to Quality in Visual Resource Imaging^[91] 等，而现在逐步扩大到资源建设的整个生命周期，包括资源选择、描述、组织、服务、知识产权保护、资源长期保护等技术、政策、流程和管理问题。本文引用的 IMLS、CCOP、NOF、CDP、以及 DNER 资源建设指南系列等都属于这类指南，DESIRE 手册^[92]和 CLIR 报告^[93] 等则更为详细和具体地对主题信息网关建设中的任务、程序和规范进行了描述，而加拿大 CLFSG^[94] 则对信息网站的建设和形态作出了详尽的规定。随着元数据体系的进一步成熟，这些指南本身可能通过 UML (Unified Modeling Language)^[95] 方式实现，其中具体内容将逐步用规范元数据形式表示，形成一个可链接、交互和扩展的信息集合，可用于配置和评价数字资源建设。

5. 关于数字资源系统服务的标准规范

任何数字资源的价值都体现在它对用户的服务，但是与资源组织的规范描述类似，服务也长期被视为本地化问题而没有成为标准规范的目标。随着网络化的发展，信息服务本身已打破本地化局限，因此它的技术因素和管理机制成为制约其实际开展和被有效利用的关键因素之一，所以人们开始利用标准规范来约束数字资源系统的服务机制，以保障系统服务在网络空间的可使用性和系统之间的互操作性。

5.1 系统服务的标准规范层次

数字信息系统服务涉及多个层次，粗略地可分为：

- (1) 接入条件，即用户要接入系统所必须具备的技术条件；
- (2) 数据传输条件，即用户要与系统交换数据内容所必须具备的技术条件；
- (3) 数据检索条件，即用户要对系统数据内容进行检索所必须具备的条件；
- (4) 数据应用条件，即用户要利用系统提供的数据内容所必须具备的技术与管理条件。

需要说明，这里的“用户”指用户系统，也包括第三方系统。而且，对于更加复杂的系统，还可能涉及其他的技术与管理条件，例如 HL7^[96]、IEEE 1073^[97]、ebXML 等体系机制。当然，系统服务的标准规范主要是关心系统间的互操作，并不排斥、甚至允许任何系统

在本地服务中采用自己的特殊方法与机制（从而支持自主系统），关键在于信息系统在与外界交互时采用标准的服务机制。

5.2 关于接入条件和传输条件的标准规范

(1) 用户服务接入条件的基本规范属于 W3C Web Accessibility Initiative^[98]的范围，WAI 提出和提供了一系列的建议和参考规范，例如 Content Accessibility Guidelines^[99]和 User Agent Accessibility Requirements^[100]等，以保障用户能方便地获取系统服务。根据 WAI 的建议，许多描述体系提出了接入条件的具体标准，例如 NOF 要求，所有资源都应通过支持 HTTP 协议和 HTML 语言的通用 Web 浏览器来读取，而且应能采用 WAI 建议的方式来保障残疾人的使用（例如提供纯文本版）。如果系统服务需要使用其它通信协议，系统应提供 Web 浏览器（实际上是 HTTP 协议）与这些协议的接口。如果系统服务要用到额外的插件，系统应保证没有这些插件的用户仍然能使用相应的服务（作为补救措施，系统可提供获取相应插件的链接或登记服务系统）。

(2) 数据传输条件主要涉及：所传输的数据内容是否能够用标准语言和格式封装、封装后的数据文件是否能通过标准网络协议传输、所传输的数据文件是否能被通用浏览器解读。对这些问题，描述体系多要求文本数据内容采取 HTML、XHTML、XML 方式封装，其它内容形式数据采用标准格式（例如 TIFF、JPEG、MPEG、WAV 等），封装后的数据文件采用 HTTP 或 FTP 等标准协议传递。实际上，图书馆界也在开发基于 XML 和 HTTP 协议的元数据交换机制，例如 LC 的 METS（Metadata Encoding and Transmission Schema）。

5.3 关于检索条件的标准规范

检索是数字图书馆服务的基本形式，也是制约数字图书馆系统互操作的主要因素。目前，多数描述体系除了要求提供基于 HTTP/HTML 的检索机制外，没有进一步规定更为详细的检索机制。但是，HTTP/HTML 检索机制的功能，尤其在支持异构系统的丰富检索功能和分布系统的集成检索方面受到较大制约，所以多种分布环境下异构系统检索机制不断被提出来，有些甚至在相当大范围内得到应用。

Z39.50^[101]是面向图书馆著录数据检索的公共标准，长期以来在图书馆自动化建设中发挥了重要作用，某些数字信息服务系统（例如美国 GILS）也使用 Z39.50。但由于 Z39.50 协议的复杂性，多数系统在具体应用它时都选择采用了其中某些功能、检索式格式、检索参数和语义定义等，从而使采用不同 Z39.50 功能和参数的系统仍然不能互操作。为避免这种情况，一些图书馆联合起来建立 Z39.50 应用协议（Application Profiles^[102]），具体规定这些图书馆在使用 Z39.50 协议时必须遵守的具体功能、格式、参数和语义定义，例如 Bath Profile^[103]、One Profile^[104]和 The Z Texas Profile^[105]等。另一方面，由于 Z39.50 属于专用的 M2M（Machine To Machine）协议，不能方便地嵌入 Web 环境，尤其是用户 Web 浏览器，所以在数字图书馆建设中并没有成为主流。考虑到这种限制，ZIG（Z39.50 Implementors Group）开始探索适应开放环境的 Z39.50 检索技术，包括基于 XML 的 Z39.50 编码方式 XER^[106]和基于 HTTP 的 ZNG 机制^[107]。许多分布检索体系还采用或实验了其它机制，例如 X500/LDAP^[108]、WHOIS++^[109]、以及 SDLIP^[110]和 STARTS^[111]等，例如 CrossROADS^[112]、IMESH^[113]等跨网关检索系统。这些机制本身虽然有一定的标准或试验协议支持，但其开放性都有局限。

从 2000 年起，OAI（Open Archives Initiative）^[114]作为一种开放检索机制开始得到广泛重视和应用。它的渊源可追溯到 NCSTRAL^[115]及其 Dienst 协议和 Handle 命名体系，并通过 UPS 模式逐步得到优化，最后以 Open Archive Metadata Harvesting Protocol^[116]来具体实现。OAMHP 要求数字资源系统能够用 DC 元数据描述数字对象（或将自己的本地元数据转换为 DC 元数据），并提供这些元数据的开放搜寻。目前 NSDL 通过 OAMHP 来建立它的核心集成系统（Core Integration Systems），通过由此生成的元数据库（Metadata Repository）来支持

对多个数字资源系统 (Digital Repositories) 的检索和对数字对象的调用。另外, 欧洲各国也开始研究和推动 OAI 机制的应用^[117]。

5.4 关于数据应用条件的标准规范

数据应用条件主要涉及用户系统能否方便有效地使用所检索的数据内容, 这可以通过采用标准数据格式在一定程度上解决 (我们在前面已论及), 但是由于许多数据内容 (例如 GIS 数据、计算数据、统计数据、虚拟现实数据等) 由于产生方式、内容构成、用途和管理要求等方面的原因, 往往要求有必要的软件模块 (可表现为浏览器插件) 来进行处理。为了支持通用用户系统 (例如通用浏览器) 对这类数据内容的方便处理, 有关系统正探索多种方式, 包括建立共享插件登记系统和在元数据中描述所需系统软件及其链接信息, 使得用户可以在调用使用数据对象时可调用相应的处理软件。不过, 作为数字图书馆领域整体, 目前对此还没有成熟的解决方案。反之, W3C 等机构正探索用 XML 开放标记语言来描述这些复杂的数据内容, 例如 SVG、SMIL、SSML^[118]、VRML/X3D^[119]等, 从而支持任何基于 XML 的用户系统对各种复杂数据内容的处理。

5.5 分布数字对象机制的标准规范

面对开放和分布的数字信息服务环境, 数字图书馆界一直在探索基于分布对象机制的数字图书馆体系, 将各种数字资源系统或服务系统视为一个数字对象, 建立标准的界面定义机制, 对它们的界面、功能、数据流、传输协议等进行规范描述, 然后通过开放的搜寻和调用机制来实现对分布、异构和变化的数字信息系统的发现、调用和配置。最初的努力倾向于建立在 CORBA^[120]、J2EE^[121]、DCOM^[122]等方式上, 但现在的趋势正走向 Web Services^[123]方式, 利用 XML 对数字信息系统进行规范描述, 利用登记系统实现这些描述信息的公共登记和开放搜寻, 通过开放协议支持基于规范描述的信息系统调用、配置和利用。正在建立的这方面的标准规范包括 WSDL^[124]、WSFL^[125]、UDDI^[126]等。数字图书馆界已经提出“开放数字图书馆”的概念^[127-128], 可以通过 Web Services 机制来更灵活地实现各种数字信息系统的方便和智能的互操作, 保障各种系统在整个网络空间的可使用性。

6. 关于数字资源长期保护的标准规范

由于数字信息依赖相应的技术、方法、系统才能被人们所利用, 因此数字信息长期保护涉及保存数字比特流、信息格式、信息处理环境、信息内容验证管理机制、信息组织机制等相关内容和机制等一系列任务, 需要信息格式标准、政策法律机制、基础技术体系、保存责任体系、知识产权措施、经济支撑方法等^[129]。对此, 图书馆界及档案、博物等领域已开始作出积极的努力, 重要成果包括:

(1) 美国 RLG 提出了数字资源长期保护的问题框架^[130], 比较全面地对存在问题、研究方向、可能技术和管理措施等进行了描述, 并建立了长期保存责任框架。

(2) 美国空间数据系统咨询委员会提出了开放档案信息系统参考模型 (Open Archive Information System, OAIS)^[131], 已被普遍接受为数字信息长期保存系统基本构架, 并已作为 ISO 标准草案。该模型提供了一个功能框架和一个信息框架, 前者包括摄取模块、长期存储模块、数据管理模块、检索传递模块和系统管理模块, 后者包括通过摄取模块获得的存交信息单元 (SIP)、经过处理后用以存储的存储信息单元 (AIP)、检索时提交的传递信息单元 (DIP)。该模型已在欧洲国家图书馆联合项目、大英图书馆、美国国会图书馆、OCLC、美国研究图书馆协会、美国国家档案馆、哈佛大学等的数字信息保存项目中得到应用。

(3) 美国 RLG/OCLC 联合提出了可信赖数字存储库的属性要求^[132], 界定了符合 OAIS 要求的数字信息长期保存系统应该具备的基本条件和责任体系

(4) 许多研究或试验项目提出了专门支持数字信息长期保护的元数据格式, 例如 CEDARS^[133]、PANDORA/NLA^[134]、NEDLIB^[135]格式, RLG/OCLC 也根据 OAIS 模型和这些

格式提出了由内容信息、保护描述信息和封装信息组成的长期保护元数据结构，并已提出了自己的内容信息元数据的建议^[136]。

由于数字信息长期保护问题本身正处于一个逐步探索和试验的阶段，许多问题及相关的规范也正在分析和试验之中，包括数字信息保护政策规范、保护技术（例如数据更新、技术仿真、数据迁移）标准、数字保存系统管理规范等。

7. 建立标准规范描述体系的目标与任务

各国数字图书馆建设、尤其是大范围合作项目，都在项目启动初期致力于建立数字信息资源建设的标准规范描述体系，指导、协调和约束参与项目建设的各个单位对标准规范的选择和采用。我们也应参照这一成功经验，在对我国的实际标准规范应用环境和制定程序进行分析的基础上，建立适应国家科学数字图书馆（CSDL）建设需要的标准规范描述体系^[137]。这里，对这个描述体系的目标和任务进行初步界定。

7.1 描述体系的目标

描述体系的目标是通过选择、推荐或制定数字信息资源或服务系统建设中的一系列标准与规范，保障这些资源与服务的可获得性、可互操作性和可持续性。可获得性（Accessibility）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可以为用户方便地使用；互操作性（Interoperability），即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（Sustainability/Preservation），即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。

7.2 描述体系的任务

描述体系并不是简单地罗列有关标准或规范，而是对如何选择、应用标准规范提出指导原则、体系结构和应用战略，并在此基础上选择和描述具体的标准规范及应用要求。

（1）确定标准与规范发展战略，其中包括：

a. 描述标准规范选择原则，主要涉及国际化程度、是否基于开放标准、是否基于开放结构和开放语言、是否基于合作发展机制等。例如，NOF 规定应该使用开放标准，即使因特殊原因不能使用开放标准，也应建议数据迁移策略（Migration strategy）来保障用非开放标准创建的数据内容能被转换为开放标准下的数据；又如，e-GIF 规定所有系统互联标准应基于 Internet 和 WWW、所有数据格式与整合标准应基于 XML；CCOP 规定使用开放格式、在没有开放格式时采用成熟的被广泛接受的“事实标准”格式、尽量减少格式种类。

b. 建立标准规范体系范围与架构，例如根据数字信息生命周期、数字信息资源集合（Digital Collections）的建设与管理、或根据数字信息系统的功能结构层次来划分标准规范的层次，并描述这些层次间的相互关系；而且，描述体系可能还规定，标准规范的形式包括正式标准（Standards）、规范（Frameworks、Guidelines、Handbooks）、应用协议（Application Profiles）、最佳应用实践（Best Practices）等多种形式，各自可能具有不同的效力。

c. 确定标准规范的发展机制，例如落实标准规范发展的责任体系，建立标准规范选择、制定、审查、推广应用的协作机制，实施基于注册系统（Registry Mechanism）的标准规范公共登记机制，建立标准规范的版本管理规范。

（2）选择推荐标准规范

根据所确定的标准规范发展原则，针对所建立的标准规范体系范围和层次，分别规定或推荐相应的标准和规范及其应用要求，实际指导数字信息资源和服务系统建设。

（3）建立标准规范应用战略，其中包括：

a. 确定描述体系的效力，例如描述体系可能具有明确的法律或行政效力，相关建设单位或投标单位必须遵守；或者，描述体系是建议性的应用指南，供相关建设单位作为最切实

践参考。无论什么情况，描述体系一般都会对各个建设单位如何应用所描述的标准规范作出规定和解释，例如某些标准是执行标准（Must use）、某些是推荐标准（Should use）、而其它是选择标准（May use），对后两种情况还可能规定不采用时的责任机制。例如，如果不采用规定的标准元数据格式，则必须提供本地格式与标准格式的转换方式。

b. 描述标准规范的应用支持机制，例如针对相关标准，描述体系一般提供定义文本、相关 URI、标准维护机构等，而且还提供配套的应用协议（Application profiles/handbooks），如果没有现成的应用协议，描述体系还将提出编制应用协议的规范、程序及这些应用协议的发布与注册要求。

7.3 描述体系的开放性

描述体系不是一个封闭体系、而是一种开放机制。这体现在以下方面：

（1）建立描述体系的目的是促进数字信息资源和服务系统在整个网络空间的可获得性、可互操作性和可持续性，因此无论建立这个体系的具体地域、行业或单位，其描述体系本身都是面向网络空间、面向分布体系、面向未来发展。

（2）为此，描述体系不应该是自定义一套标准或规范，而是充分吸纳和选择那些在国际范围广泛接受和应用、得到有效标准化过程支持和市场支持、具有良好可持续发展前景的标准和规范，减少标准规范制定和应用成本。

（3）为保障所选择的标准规范的可互操作性、可扩展和可持续发展，描述体系积极选择开放标准，即这些标准的定义是公开的和可被公共搜寻的，并有可靠的公共登记、持续维护、开放转换（即转换模块本身是开放的）和开放扩展机制，最好是基于开放描述语言的和可自动解析的。

另外，为保障上述三方面要求的实现，描述体系制定过程及其更新过程应该采取开放和合作机制，通过公开、协作和持续的过程收集信息、征集意见、形成共识、推动应用。

国家科学数字图书馆（CSDL）是基于开放体系结构的数字信息系统，通过开放建设和集成各种分布、异构和多样化的数字信息资源和服务系统，来形成个性化的虚拟信息服务机制。为保证这些资源和服务的可获得性、可互操作性和可持续性，CSDL 应该按照国际化、开放性、技术成熟性、低应用成本等原则，明确建立自己的标准规范描述体系，约束和协调自己系统内的资源建设，并提供与其他系统互操作的开放规范。

8. 参考文献：

- [1] nof-digitise Technical Standards and Guidelines. Revised Nov. 2000
<http://www.peoplesnetwork.gov.uk/nof/technicalstandards/index.html>
- [2] Working with the Distributed National Electronic Resources. Feb. 2001
<http://www.jisc.ac.uk/dner/programmes/guidance/DNERStandards.html>
- [3] Standards and Guidelines for Digitization Projects for Canadian Culture Online Program, Dec.3, 2001.
http://www.pch.gc.ca/ccop-pcce/pubs/ccop-pceguide_e.pdf
- [4] Institute of Museum and Library Services. A Framework of Guidance for Building Good Digital Collections. November 6, 2001 <http://www.ims.gov/pubs/forumframework.htm>
- [5] RLG Cultural Materials Initiative Recommendations for Digitizing for RLG Cultural Materials 25 Jan 2002 (Draft) <http://www.rlg.org/culturalres/prospective.html>
- [6] LC Digital Formats for Content Reproductions. <http://memory.loc.gov/ammem/formats.html>
- [7] UK Cabinet Office. E-Government Interoperability Framework, V.3, Oct. 2001
http://www.govtalk.gov.uk/documents/e-GIF_version_3_approved.pdf
- [8] ebXML Technical Architecture Specification, 16 Feb, 2001. <http://www.ebxml.org/specs/ebTA.pdf>
- [9] Guidance on Interoperability with the Distributed National Electronic Resource (DNER) for product suppliers. Oct. 25, 2001. http://www.jisc.ac.uk/dner/collections/dner_interoperability.htm
- [10] Arizona Digital Project Guidelines. Arizona State Library, Archives and Public Records, Digital Imaging Task Force, March 2000, Version 1.3 <http://www.lib.az.us/digital/>
- [11] Colorado Digitization Project General Guidelines for Descriptive Metadata Creation and Entry.

- <http://coloradodigital.coalition.org/glines.html>
- [12] LC Access Aids and Interoperability. <http://memory.loc.gov/ammem/award/docs/interop.html>
- [13] RLG Cultural Materials Alliance Description Guidelines. V1.2.0, 11 January 2002
<http://www.rlg.org/culturalres/descguide.html>
- [14] OhioLINK Multimedia Center Standards. <http://www.ohiolink.edu/media/dmcinfo/metadata.html>
- [15] California Digital Library Digital Image Format Standards. July 9 2001
<http://www.cdlib.org/about/publications/CDLImageStd-2001.pdf>
- [16] California Digital Library Digital Object Standard: Metadata, Content and Encoding, May 2001
<http://www.cdlib.org/about/publications/CDLObjectStd-2001.pdf>
- [17] NSDL Metadata Primer. <http://metamanagement.comm.nslib.org/outline.html>
- [18] Metadata Framework for Resource Discovery of Agricultural Information. FAO/UN. 2001
<http://www.fao.org/agris/MagazineArchive/MetaData/OAICConfRevised.doc>.
- [19] CEN/ISSS (Information Society Standardization System/European Committee for Standardization) Metadata Framework. <http://www.cenorm.be/iss/Workshop/delivered-ws/mmi/metadata/web/main.htm>
- [20] The INDECS Metadata Framework <http://www.indecs.org/pdf/schema.pdf>
- [21] UK Cabinet Office, Office of the e-Envoy. E-Government Metadata Framework, May 2001
<http://www.govtalk.gov.uk/documents/UK%20Metadata%20Framework%20v1%202001-05.pdf>
- [22] TBITS 39: Treasury Board Information Management Standard, Part 1: Government On-Line Metadata Standard. http://www.cio-dpi.gc.ca/its-nit/standards/tbits39/crit391_e.asp
- [23] ISO/IEC 10646-1:2000 Information technology -- Universal Multiple-Octet Coded Character Set (UCS) -- Part 1: Architecture and Basic Multilingual Plane
- [24] The Unicode Standard, Version 3.0 <http://www.unicode.org/unicode/uni2book/u2.html>
- [25] GB2312-1980. 《信息交换用汉字编码字符集基本集》
- [26] GB18030-2000. 《信息交换用汉字编码字符集的扩充》
- [27] GB13000-1993. 《通用多八位编码字符集：1. 体系结构与基本多文种平面》
- [28] Scalable Vector Graphics (SVG) 1.0 Specification, W3C Recommendation 04 September 2001
<http://www.w3.org/TR/SVG/>
- [29] Synchronized Multimedia Integration Language (SMIL) 1.0 Specification, W3C Recommendation 15-June-1998 <http://www.w3.org/TR/REC-smil>
- [30] Mathematical Markup Language (MathML™) 1.01 Specification, W3C Recommendation, revision of 7 July 1999. <http://www.w3.org/TR/REC-MathML/>
- [31] Geography Markup Language (GML) 2.0, OpenGIS® Implementation Specification, 20 February 2001
<http://opengis.net/gml/01-029/GML2.html>
- [32] Chemical Markup Language. <http://www.xml-cml.org/>
- [33] Extensible Markup Language (XML) 1.0 (Second Edition) W3C Recommendation 6 October 2000
<http://www.w3.org/TR/REC-xml>.
- [34] XML Schema Part 0: Primer W3C Recommendation, 2 May 2001. <http://www.w3.org/TR/xmlschema-0/>
- [35] DCMI DCSV: A syntax for writing a list of labelled values in a text string.
<http://dublincore.org/documents/2000/07/28/dcmi-dcsv/>
- [36] Getting Started with TeX and LaTeX. <http://www.tug.org/begin.html>
- [37] TIFF Revision 6.0, 1992. <http://partners.adobe.com/asn/developer/pdfs/tn/TIFF6.pdf>
- [38] JPEG Specification. <http://www.w3.org/Graphics/JPEG/itu-t81.pdf>
- [39] PDF Reference, Third Edition, Version 1.4. Addison-Wesley, 2001 (ISBN 0-201-75839-3)
- [40] Graphics Interchange Format, Version 89a. <http://astronomy.swin.edu.au/~pbourke/dataformats/gif/>
- [41] Portable Network Graphics. <http://www.w3.org/TR/REC-png.html>
- [42] SPIFF - Still Picture Interchange File Format.
<http://www.faqs.org/faqs/graphics/fileformats-faq/part3/section-143.html>
- [43] MPEG. <http://www.cselt.it/mpeg/>
- [44] Apple Quicktime Developer Resources <http://developer.apple.com/quicktime/>
- [45] Microsoft Windows Media Technologies. <http://www.microsoft.com/windows/windowsmedia/default.asp>
- [46] Waveform Audio File Format. <http://www.borg.com/~jglatt/tech/wave.htm>
- [47] Audio Interchange File Format. <http://www.borg.com/~jglatt/tech/aiff.htm>
- [48] Vector Markup Language (VML), World Wide Web Consortium Note 13-May-1998.
<http://www.w3.org/TR/NOTE-VML>
- [49] Micromedia Flash. <http://www.macromedia.com/software/flash/>
- [50] Naming and Addressing: URIs, URLs, ... <http://www.w3.org/Addressing/>
- [51] 张晓林. 数字对象的唯一标识符技术. 现代图书情报技术, 2000.5
- [52] The Digital Object Identifier System <http://www.doi.org/>

- [53] Serial Item and Contribution Identifier Standard. ANSI/NISO Z39.56-1996 Version 2
<http://sunsite.berkeley.edu/SICI/>
- [54] Book Item and Component Identifier. Draft Standard of ANSI/NISO. www.niso.org/pdfs/BICI-DS.pdf
- [55] Persistent Uniform Resource Locator <http://www.purl.org/>
- [56] Metadata Encoding and Transmission Protocol <http://www.loc.gov/standards/mets/>
- [57] Sharable Content Object Reference Model. <http://www.adlnet.org/scorm/scorm.cfm>
- [58] Text Encoding Initiative. <http://www.tei-c.org/>
- [59] Global Information Locator Service <http://www.gils.net/>
- [60] Content Standard for Digital Geospatial Metadata (CSDGM). <http://www.fgdc.gov/metadata/constan.html>
- [61] Encoded Archival Description. <http://www.loc.gov/ead/>
- [62] VRA Core Categories, Version 3.0. <http://www.vraweb.org/vracore3.htm>
- [63] IEEE Learning Object Metadata. <http://ltsc.ieee.org/wg12/>
- [64] MARC Standards <http://www.loc.gov/marc/>
- [65] ONIX Product Information Standards. <http://www.editeur.org/onix.html>
- [66] Web-Ontology (WebOnt) Working Group. <http://www.w3.org/2001/sw/WebOnt/>
- [67] Dublin Core Metadata Initiative. <http://www.dublincore.org/>
- [68] Australia Government Information Locator.
http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html
- [69] MIREG Metadata Framework - Element Set, Draft - 2001-08-28.
<http://dublincore.org/groups/government/mireg-metadata-20010828.shtml>
- [70] IMS Learning Resource Meta-data Specification. <http://www.imsproject.org/metadata/>
- [71] Canadian Core Learning Resource Metadata Application Profile. <http://www.cancore.ca/>
- [72] NISO Technical Metadata for Digital Still Images. http://www.niso.org/committees/committee_au.html
- [73] TBITS 39: Treasury Board Information Management Standard, Part 2: Controlled Vocabulary Standard.
http://www.cio-dpi.gc.ca/its-nit/standards/tbits39/crit392_e.asp
- [74] Thesaurus for Graphic Materials. <http://www.loc.gov/rr/print/tgm1/toc.html>
- [75] Thesaurus of Geographic Names. <http://www.getty.edu/research/tools/vocabulary/tgn/>
- [76] Medical Subject Headings <http://www.nlm.nih.gov/mesh/>
- [77] Taxonomic Resources and Index to Organic Names <http://www.york.biosis.org/triton/>
- [78] Geographic Names Information Systems. <http://mapping.usgs.gov/www/gnis/>
- [79] Art and Architecture Thesaurus <http://www.gii.getty.edu/vocabulary/aat.html>
- [80] Union List of Artists Names. <http://www.gii.getty.edu/vocabulary/ulan.html>
- [81] Networked Knowledge Organization Systems/Services. <http://nkos.slis.kent.edu/>
- [82] RSLP Collection Level Description. <http://www.ukoln.ac.uk/metadata/rslp>
- [83] Vocabulary Markup Language. <http://xml.coverpages.org/VOCML-DTD10.txt>
- [84] XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification. <http://www.topicmaps.org/xtm/1.0/>
- [85] Zthes: a Z39.50 Profile for Thesaurus Navigation. <http://www.loc.gov/z3950/agency/profiles/zthes-03.html>
- [86] ISO/DIS 16642. Computer applications in terminology -- Terminological markup framework (TMF)
- [87] Platform for Privacy Preferences (P3P) Project. <http://www.w3.org/P3P/>
- [88] XML Access Control Language: Provisional Authorization for XML Documents. October 16, 2000
- [89] The Open Digital Rights Language Initiative. <http://odrl.net/>
- [90] Platform for Internet Content Selection (PICS). <http://www.w3.org/PICS/>
- [91] Guides to Quality in Visual Resource Imaging, July 2000 <http://www.rlg.org/visguides/>
- [92] DESIRE Information Gateway Handbook. <http://www.desire.org/handbook/>
- [93] Louis A. Pitschmann, Building Sustainable Collections of Free Third-Party Web Resources, June 2001.
<http://www.clir.org/pubs/reports/pub98/contents.html>
- [94] Common Look and Feel Standards and Guidelines. http://www.cio-dpi.gc.ca/clf-upe/apr_e.asp
- [95] Unified Modeling Language. <http://www.uml.org/>
- [96] Health Level 7. <http://www.hl7.org/>
- [97] IEEE 1073 Point of Care Medical Device Communication Standards.
<http://www.ieee1073.org/overview/overview.html>
- [98] W3C Web Accessibility Initiative. <http://www.w3.org/WAI/>
- [99] Web Content Accessibility Guidelines 1.0. W3C Recommendation 5-May-1999
<http://www.w3.org/TR/WAI-WEBCONTENT>
- [100] User Agent Accessibility Guidelines 1.0. 12 September 2001. <http://www.w3.org/TR/UAAG10/>
- [101] Z39.50 Standards. <http://www.loc.gov/z3950/agency/>
- [102] About Z39.50 Profiles. <http://lcweb.loc.gov/z3950/agency/profiles/about.html>
- [103] The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery

- <http://www.nlc-bnc.ca/bath/bp-current.htm>
- [104] ONE-2 Profile. <http://www.one-2.org/>
- [105] Z Texas Profile A Z39.50 Specification for Library Systems Applications in Texas
<http://www.tsl.state.tx.us/ld/projects/z3950/tzigprofilerelease20.html>
- [106] ez39.50 - XER over SOAP. <http://www.lib.ox.ac.uk/jafer/ez3950/ez3950.html>
- [107] The ZNG Initiative. <http://www.loc.gov/z3950/agency/zng.html>
- [108] Lightweight Directory Access Protocol (Version 3) Specifications
<http://www.innosoft.com/ldapworld/ldapv3.html>
- [109] Architecture of the WHOIS++ service. <http://www.ietf.org/rfc/rfc1835.txt>
- [110] The Simple Digital Library Interoperability Protocol (SDLIP-Core).
<http://www.diglib.stanford.edu/~testbed/doc2/SDLIP/>
- [111] Stanford Protocol Proposal for Internet Retrieval and Search.
<http://www-db.stanford.edu/~gravano/starts.html>
- [112] CrossROADS. <http://www.ukoln.ac.uk/metadata/roads/crossroads/>
- [113] IMesh: International Collaboration on Internet Subject Gateways. <http://www.imesh.org/>
- [114] Open Archives Initiative. <http://www.openarchives.org/>
- [115] Networked Computer Science Technical Reference Library. <http://www.ncstrl.org/>
- [116] The Open Archives Initiative Protocol for Metadata Harvesting, July 2, 2001
http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html
- [117] Open Archives Forum. <http://www.oaforum.org/>
- [118] Speech Synthesis Markup Language Specification. W3C Working Draft 5 April 2002
<http://www.w3.org/TR/speech-synthesis/>
- [119] The Virtual Reality Modeling Language (VRML). ISO/IEC 14772:1997.
<http://www.web3d.org/Specifications/VRML97/>, X3D. (VRML97 in XML: Extensible 3D).
<http://www.web3d.org/x3d.html>
- [120] Common Object Request Broker Architecture. <http://www.corba.org/>
- [121] Java 2 Platform Enterprise Edition <http://java.sun.com/j2ee/>
- [122] Distributed Component Object Model. <http://www.microsoft.com/com/tech/DCOM.asp>
- [123] Web Services Activity. <http://www.w3.org/2002/ws/>
- [124] Web Services Description Language (WSDL) 1.1, W3C Note 15 March 2001. <http://www.w3.org/TR/wsdl>
- [125] Web Service Flow Language (WSFL) 1.0, May 2001.
<http://www-4.ibm.com/software/solutions/webservices/pdf/WSFL.pdf>
- [126] Universal Description, Discovery and Integration. <http://www.uddi.org/specification.html>
- [127] Suleman, H. and Fox, E. A Framework for Building Open Digital Libraries, D-Lib Magazine, December 2001. <http://www.dlib.org/dlib/december01/suleman/12suleman.html>
- [128] 张晓林。开放数字信息服务机制：概念、结构与技术。中国图书馆学报，2002.3
- [129] 张晓林。数字信息的长期保存问题。图书馆，2001.5
- [130] Preserving Digital Information: Final Report and Recommendations. <http://www.rlg.org/ArchTF/index.html>
- [131] Reference Model for an Open Archival Information System (OAIS) May 1999.
<http://www.ccsds.org/RP9905/RP9905.html>
- [132] RLG/OCLC Digital Archive Attributes Working Group: Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources. <http://www.rlg.org/longterm/attributes01.pdf>
- [133] Metadata for Digital Preservation : the Cedars Project Outline Specification.
<http://www.leeds.ac.uk/cedars/colman/metadata/metadata-spec.html>
- [134] National Library of Australia Preservation Metadata for Digital Collections. Exposure Draft
<http://www.nla.gov.au/preserve/pmeta.html>
- [135] Metadata for long term-preservation, July 2000. <http://www.kb.nl/coop/nedlib/results/D4.2/D4.2.htm>
- [136] OCLC/RLG Preservation Metadata Working Group. A Recommendation for Content Metadata.
<http://www.oclc.org/research/pmwg/contentinformation.pdf>
- [137] Gill, T. and Miller, P. Re-inventing the Wheel? Standards, Interoperability and Digital Cultural Content. D-Lib Magazine, Jan. 2002. <http://www.dlib.org/dlib/january02/gill/01gill.html>

9. 附录：英文缩略语

- AAT: Art and Architecture Thesaurus
ADL: Advanced Distance Learning
AGIL: Australia Government Information Locator
AIFF: Audio Interchange File Format

AIP: Archiving Information Package
BICI: Book Item and Content Identification
CanCore: Canadian Core Learning Resource Metadata Application Profile
CCOP: Canadian Culture Online Program
CDL: California Digital Library
CDP: Colorado Digitization Project
CEDARS: CURL Exemplar in Digital ARchivalS
CEN/ISSS: European Committee for Standardization/Information Society Standardization System
CLFSG: Common Look and Feel Standards and Guidelines
CLIR: Council for Library and Information Resources
CML: Chemical Markup Language
CORBA: Common Object Request Broker Architecture
CSDGM: Content Standard for Digital Geospatial Metadata
CSV: Comma Separable Variable Format
DC: Dublin Core
DCMI: Dublin Core Metadata Initiative
DCOM: Distributed Component Object Model
DCSV: Dublin Core Structured Values
DIP: Dissemination Information Package
DLF: Digital Library Federation
DNER: Distributed National Electronic Resources
DOI: Digital Object Identifier
EAD: Encoded Archival Description
ebXML: Electronic Business XML
e-GIF: e-Government Interoperability Framework
e-GMF: e-Government Metadata Framework
FGDC: Federal Geographic Data Committee
GIF: Graphics Interchange Format
GILS: Government Information Locator Service
GML: Geography Markup Language
GNIS: Geographic Names Information Systems
HTML: HyperText Markup Language
IETF: Internet Engineering Task Force
IEEE LOM: IEEE Learning Object Metadata
IMS: Instruction Management System
IMSL: Institute of Museum and Library Services
INDECS: Interoperability of Data in E-Commerce Systems
ISO TMF: ISO Terminological Markup Framework
J2EE: Java 2 Platform Enterprise
JPEG: Joint Picture Expert Group
KOS: Knowledge Organization Systems
LCSH: LC Subject Headings
LDAP: Lightweight Directory Access Protocol
MARC: Machine-Readable Cataloging
MathML: Mathematical Markup Language
MESH: Medical Subject Headings
METS: Metadata Encoding and Transmission Schema
MIReG: Managing Information Resources for e-Government
NCSTRAL: Networked Computer Science Technical Reference Library
NEDLIB: Networked European Deposit Library
NISO TMI: National Information Standard Organization Technical Metadata for Image
NKOS: Networked Knowledge Organization Systems
NOF: New Opportunity Fund
NSDL: National SMETE (Science, Medicine, Engineering, Technology Education) Digital Library.
OAI: Open Archives Initiative
OAIS: Open Archive Information System
OAMHP: Open Archive Metadata Harvesting Protocol
ODRL: Open Digital Rights Language

ONIX: Online Information eXchange
P3P: Platform for Privacy Preferences
PANDORA: Preserving and Accessing Networked Documentary Resources of Australia.
PDF: Portable Document Format
PICS: Platform for Internet Content Selection
PNG: Portable Network Graphics
PURL: Persistent Uniform Resource Locator
RDF: Resource Description Framework
RDFS: Resource Description Framework Schema
RLG/CMI: Research Library Group Cultural Material Initiative
RSLP CDS: Research Support Libraries Programme Collection Description Standard
SCORM: Sharable Content Object Reference Model
SDLIP: Simple Digital Library Interoperability Protocol
SICI: Serial Item and Content Identification
SIP: Submission Information Package
SMIL: Synchronized Multimedia Integration Language
SPIFF: Still Picture Interchange File Format
SSML: Speech Synthesis Markup Language
STARTS: Stanford Protocol Proposal for Internet Retrieval and Search
SVG: Scalable Vector Graphics
TEI: Text Encoding Initiative
TIFF: Tagged Image File Format
TGM: Thesaurus for Graphic Materials
TGN: Thesaurus of Geographic Names
TRION: Taxonomic Resources and Index to Organic Names
UDDI: Universal Description, Discovery and Identification
ULAN: Union List of Artists Names
UML: Unified Modeling Language
URI: Uniform Resource Identifier
VocML: Vocabulary Markup Language
VRA: Visual Resource Association
VRML/X3D: Virtual Reality Modeling Language/Extensible 3D
WAI: Web Accessibility Initiative
WAV: Waveform Audio File Format
WSDL: Web Service Description Language
WSFL: Web Service Flow Language
XACL: Extensible Access Control Language
XHTML: Extensible HyperText Markup Language
XML: Extensible Markup Language
XML DTD: XML Document Type Definition
XTM: XML Topic Maps
ZNG: Z39.50 Next Generation
Zthes: Z39.50 Profile for Thesaurus Navigation