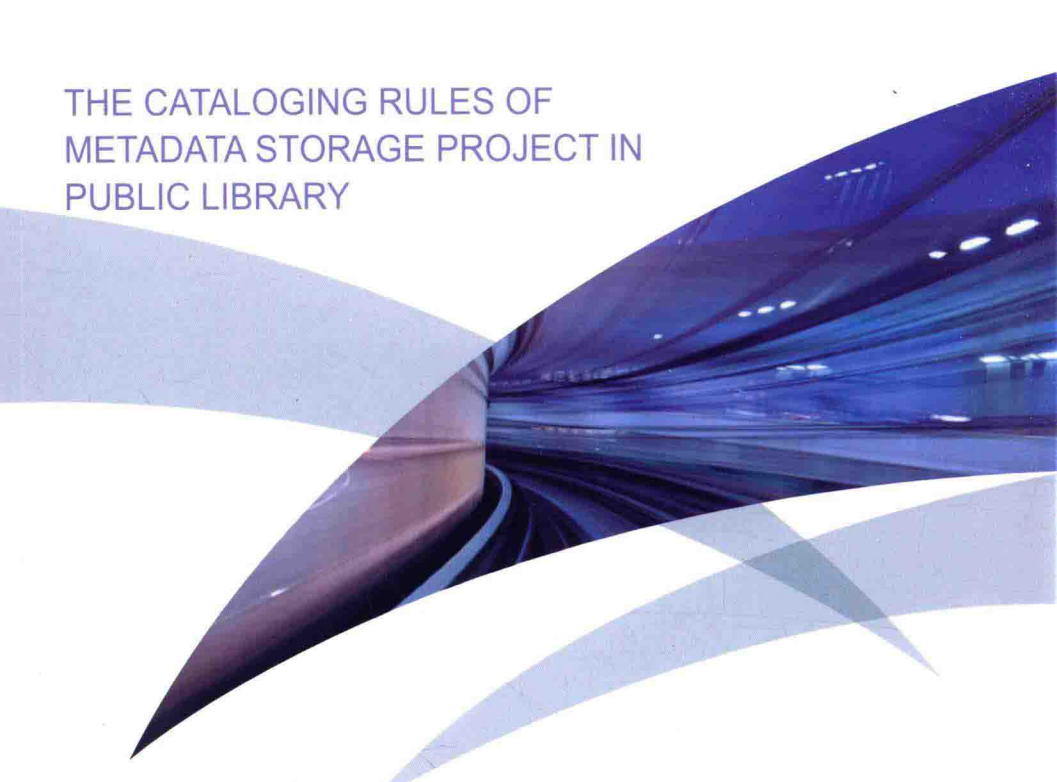




THE CATALOGING RULES OF
METADATA STORAGE PROJECT IN
PUBLIC LIBRARY

公共图书馆元数据 仓储编目规则

萨 蕾 韩 萌 孟 然 编著



國家圖書館出版社
National Library of China Publishing House

公共图书馆元数据 仓储编目规则

萨 蕾 韩 萌 孟 然 编著



國家圖書館出版社
National Library of China Publishing House

图书在版编目(CIP)数据

公共图书馆元数据仓储编目规则/萨蕾,韩萌,孟然编著.--北京:
国家图书馆出版社,2015.12

ISBN 978-7-5013-5742-0

I. ①公… II. ①萨… ②韩… ③孟… III. ①公共图书馆—
数字资源—编目规则 IV. ①G254.3

中国版本图书馆 CIP 数据核字(2015)第 292039 号

书 名 公共图书馆元数据仓储编目规则
著 者 萨 蕾 韩 萌 孟 然 编著
责任编辑 金丽萍 唐 澈

出 版 国家图书馆出版社(100034 北京市西城区文津街 7 号)
(原书目文献出版社 北京图书馆出版社)

发 行 010-66114536 66126153 66151313 66175620
66121706(传真),66126156(门市部)

E-mail nlcpress@nlc.cn(邮购)

Website www.nlcpress.com →投稿中心

经 销 新华书店

印 装 北京科信印刷有限公司

版 次 2015 年 12 月第 1 版 2015 年 12 月第 1 次印刷

开 本 880×1230(毫米) 1/32

印 张 4.25

字 数 95 千字

书 号 ISBN 978-7-5013-5742-0

定 价 30.00 元

前 言

随着计算机、网络技术的发展和国家信息基础设施环境的完善,数字图书馆的建设得到稳步推进,数字资源不断增长,但由于不同资源及系统间存在异构性,传统的整合模式与检索服务已难以满足对图书馆资源的统一管理 with 揭示,这不仅降低了资源的利用率,更给用户带来极大不便。为提升服务水平,节省用户查询时间,提高资源利用率,亟须建立一种新型的资源管理与检索模式。

元数据仓储是将异构数据库系统中的元数据复制并提取,通过分析、综合、转换、清洗和装载,使分散的、异构的数据转换为集成的、同构的数据,从而提供一个集中的、统一的“一站式”检索服务。元数据仓储相当于用户对本地单个数据库进行直接检索,可提升检索响应速度和整合效率,有利于数据挖掘和知识发现,可实现各类数字资源的一站式检索服务。元数据仓储的建设将为数字资源的有序组织、统一管理和高效检索提供基础,为数字资源的深层次揭示提供保证,为读者提供方便、快捷、功能强大的数字资源服务,并进一步推动数字图书馆建设的可持续发展。

2013 年,国家向各省下达专项资金用于省级图书馆开展数字图书馆推广工程数字资源联合建设工作。这一项目开启了全国范围内的数字资源合作建设体系的建设,具有划时代的意义。项目实施两年来,取得了显著的成果,共建馆涵盖了全国所有省、市级公共图书馆,建设了大量的数字资源,充分证明了联合建设是全国公共图书馆建设数字资源行之有效的方式。元数据仓储项目是联建工作最早开展的项目之一,有利于加强全国公共图书馆自建数字馆藏的统一展示,实现资源的统一管理 with 利用,对联建工作及我国数字图书馆事业都起到重要的支持作用。项目经过连续两年的建设,成果斐然。但是,图书馆自建数字馆藏类型多样、元数据格式各异以及联建馆数量的持续增

加,对项目的可持续发展带来了一定影响,因此,项目建设标准化、规范化的要求日益凸显。为更好地建设元数据仓储项目,特制定本规则,主要为公共图书馆共同进行自建数字馆藏元数据的制作提供著录细则,同时提供编目实例,两者相辅相成,帮助编目人员快速准确地理解本规则。

感谢各省级、副省级公共图书馆同仁的支持和协助,感谢国家图书馆各位领导的支持与帮助,感谢国家图书馆推广工程培训与资源建设组梁蕙玮、郭炯、汪静、李云龙、韩萌、苏宾、孟然的辛勤付出,感谢国家图书馆出版社编辑的细致审核,正是在各方的支持、指导、帮助和参与下,本书才得以顺利出版。

由于元数据仓储的建设涉及面广,本书撰写时,很难涵盖所有的资源类型或囊括资源的所有属性,对于疏漏之处,敬请各位专家学者及图书馆同人不吝指正。

萨 蕾

2015 年 5 月

目 录

第一部分 综述	(1)
1 元数据研究	(3)
2 DC 元数据研究	(8)
第二部分 总则	(15)
1 范围	(17)
2 规范性引用文件	(17)
3 主要名词术语	(18)
4 著录对象	(18)
5 著录单位	(18)
6 著录信息源	(19)
7 著录用文字	(20)
8 著录规则的内容结构	(20)
9 术语集说明	(21)
第三部分 著录项说明	(25)
1 正题名	(27)
2 其他题名	(31)
3 责任者名称	(32)
4 责任方式	(34)
5 分类法	(36)
6 主题词或关键词	(40)
7 摘要	(42)
8 目录	(43)

9	图像	(44)
10	附注	(45)
11	出版者名称	(46)
12	出版地	(48)
13	开始日期	(49)
14	终止日期	(50)
15	出版日期	(51)
16	创建日期	(52)
17	发布日期	(53)
18	更新频率	(53)
19	内容形式	(54)
20	媒体类型	(56)
21	学科类型	(56)
22	原始标识符	(57)
23	文件大小	(58)
24	时长	(58)
25	文件格式	(59)
26	文件数量	(60)
27	颜色	(61)
28	分辨率	(61)
29	声音	(62)
30	比特率	(62)
31	运行环境	(63)
32	页码	(64)
33	文字语种	(65)
34	字幕语种	(66)
35	声道语种	(66)
36	丛编	(67)
37	所属数据库	(68)

38	出自	(70)
39	相关资源	(71)
40	时间范围	(72)
41	空间范围	(73)
42	版本	(74)
43	访问方式	(75)
44	访问范围	(75)
45	发布地址	(76)
46	适用对象	(76)
47	出版频率	(77)
48	馆藏范围	(78)
49	标识符	(79)
50	其他字段	(84)
附录 1	多层次著录	(85)
附录 2	从 CNMARC 格式元数据段到本规范的映射	(87)
附录 3	四部分类法	(91)
附录 4	内容形式	(103)
附录 5	媒体类型	(105)
附录 6	数据库学科类型表	(106)
第四部分	术语表(按类型)	(107)
1	数据库	(109)
2	图书	(109)
3	图像	(110)
4	音频	(111)
5	视频	(112)
6	连续性资源	(113)
7	文章	(114)
8	网页	(115)

第五部分 实例	(117)
1 数据库	(119)
2 图书	(120)
3 图像	(121)
4 音频	(122)
5 视频	(123)
6 连续性资源	(125)
7 文章	(126)
8 网页	(126)
参考文献	(128)

第一部分 综述

1 元数据研究

1.1 定义

元数据最通俗的定义是“关于数据的数据”。对元数据的研究往往是从元数据的定义入手,因此产生了多种更明确更细化的定义。包括以下几种:

- data that defines and describes other data. [ISO/IEC 11179 - 3: 2003(E)]

- 元数据是关于数据的数据。此术语指任何用于帮助网络电子资源的识别、描述和定位的数据。(IFLA)

- 用于提供某种资源的有关信息的结构数据(Structured data),或者说是描述其他数据的数据(Data about other data)^[1]。

- 元数据是用于描述数据内容(what)、覆盖范围(where when)、质量、管理方式、所有者(who)、提供方式(how)的数据,是数据与数据用户之间的桥梁^[2]。

- 元数据是与对象相关的数据,此数据使其潜在的用户不必预先具备对这些对象的存在或特征的完整认识。它支持各种操作。用户可能是程序,也可能是人^[3]。

- 元数据是对信息包(Information package)的编码描述,其目的在于提供一个中间级别的描述,使得人们据此就可以做出选择,确定其想要浏览或检索的信息包,而无须检索大量不相关的全文文本^[4]。

- 关于信息资源或数据的一种结构化的数据^[5]。

- 是面向某种特定应用的用于描述资源属性的机器可理解的信息。通过规范语法结构和语义结构,使得机器能够无二义性地表现和获取信息^[6]。

.....

对上述定义进行分析,可以看到,这些定义对元数据的特征进行

了概括,其特征包括以下几点。

1.1.1 元数据的描述对象

元数据的描述对象包括数据、信息或知识。数据是对客观事物、事件的记录、描述,是可由人工或自动化手段加以处理的数字、文字、图形、图像、声音等符号的集合。信息是客观世界中各种事物的状态和特征的反映,是与问题相关的数据,可以采用文本、图形、图像、音频、视频等形式记录下来,能通过媒介进行传输。知识是人们从实践经验中总结出来且被新的实践所证实的规律及经验的总结,是可以用于推理的规则^[7]。

不对描述对象进行限制更能体现元数据的作用,也更符合元数据的本质特征。基于此认识,元数据所描述的对象既包括实体文献,也包括数字资源;既包括单个的独立资源,也包括集成性资源(如期刊、网站等),以及多媒体资源等;既包括物理实体,也包括虚拟实体;既包括知识组织工具(如词表)、服务、信息系统等,也包括元数据本身。而从数据的类型上看,包括了书目型数据、文献型数据、数字型数据、数值型数据等。

1.1.2 元数据涵盖的范围

随着因特网的发展,元数据的概念出现在图书情报领域,一般认为,元数据这一词汇最早出现在1988年美国国家航空航天局(NASA)的《目录交换格式》(*Digital Information Formats*,简称DIF)手册中。在这一词汇出现之前,图书馆已经进行了多年的编目工作,制作了大量的书目数据、索引数据等。最为通用的IFLA的定义也将元数据的描述对象局限于网络电子资源。但是,也有学者认为书目数据是元数据,TEI标题也是,或其他形式的描述^[8]。尽管描述对象发生了很大的变化,使得描述网络资源的元数据与传统的书目数据在元数据标准上有着一定的差异,但是,从元数据的描述对象来看,传统的书目数据、目次数据等也应是元数据的一部分,描述网络资源的元数据与书目数据仍然是同一性质,同时,将元数据局限于对网络电子资源的描述不利于元数据的发展。因此,对元数据的定义应是较为宽泛的,从

元数据所起的作用出发进行定义更为科学合理。

1.1.3 元数据的属性

元数据最基本的属性包括两点:

其一,元数据具有明确的语义和结构。元数据具有明确的语义和结构,这一性质使得元数据互操作具有了可行性。元数据的互操作性体现在对异构系统间互操作能力的支持。基于互操作实现的信息资源的共享与融合可以使元数据发挥更大的使用价值,因此,互操作性在元数据使用与发展的过程中具有重要意义。然而,不同的元数据都具有一定的个性化特点,因此,在互操作时,应以损失最小化为原则。

其二,元数据是机器可读、可理解的数据,即元数据的使用者可能是人,也可能是机器。在数字时代,这一点是元数据存在的基础,也是更好发挥元数据功能的重要保障,也确保了元数据在生成、组织、管理、利用、保存等过程中与机器的语义互通。因此,这一点是至关重要的。

1.1.4 元数据的特点

元数据最显著的特点为描述性及模块化两点。

(1) 描述性

描述是元数据最重要最基本的功能,不论是描述元数据还是管理元数据,都是在对描述对象的某一方面的属性进行描述。

描述性体现在元数据的形成过程中,其重要性在于只有加强描述性才能实现元数据管理、使用的功能,因此,在制作元数据时,要遵循客观性、准确性、完整性等描述原则。

(2) 模块化

模块化(Modularity),指按照所描述的信息系统内容,将元数据划分为针对不同层次、功能或应用的逻辑模块,每个元数据格式只是一个这样的模块,分别对信息系统的不同内容进行描述,分别满足不同的逻辑功能和应用需要^[9]。

模块化主要以两种方式体现:①不同的元数据类型形成了不同的模块。如:MARC元数据标准家族包括了书目数据、馆藏数据、规范数

据。②不同的功能形成了不同的模块。如国家图书馆元数据规范中就要求,元数据应包括描述信息、技术信息、管理信息。

模块化的优点在于:模块化使得元数据标准的建设具有动态性,一个模块可以独立使用,也可以实现按照具体需求与其他模块的多次重组;对于实际应用,往往需要多个模块的组合才能满足需求,模块化与资源描述与使用的客观需要相吻合;模块化有利于对元数据标准的复用,在复用时,可从其他元数据中选择部分模块直接复用,或按照自身的需求进行一定的扩展。

1.1.5 元数据的结构

元数据包括内容结构、语法结构、语义结构。内容结构定义了元数据的构成要素,如描述性元素、管理性元素、元素选取使用规则等。语法结构定义了元数据的格式结构及其描述方式,包括元数据的结构、元素结构、元素复用方式、与描述对象的捆绑方式等。语义结构定义了元素定义、元素内容编码规则定义等。语义通过属性元素表达,结构是语义的抽象载体,可以提供人类和机器的双重理解;语法(句法)是置标方案,用以传达语义和结构。

1.2 类型

研究者对元数据的类型也有多种认识。主要有以下几种:

(1)杨超认为数字图书馆领域的元数据可以分为以下三类^[10]:

①描述性元数据(Descriptive Metadata):描述对象知识内容的信息,例如 MARC 编目记录,检索工具或者类似的模式、框架。

②管理性元数据(Administrative Metadata):允许仓储管理对象的必需的信息,包括扫描信息、储存格式等(通常这类元数据也被称作技术元数据)、版权和许可信息、数字对象长期保存所需要的各种信息(保存元数据)。

③结构性元数据(Structural Metadata):将对象互相连接以形成逻辑单元的信息,例如一书中每页内容的镜像互相关联从而形成这本书自身。

(2) Anne J. Gilliland-Swetland 将元数据分为以下五种类型^[11]：

- ①管理型：在管理信息资源中利用的元数据。
- ②描述型：用来描述或者识别信息资源的元数据。
- ③保存型：与信息资源的保存管理相关的信息。
- ④技术型：与系统如何行使职责或元数据如何发挥作用相关的元数据。
- ⑤使用型：与信息资源利用的等级和类型相关的元数据。

(3) 根据元数据在组织信息资源的功能上划分，元数据可分为以下几种^[12]：

①知识描述型元数据 (Intellectual Metadata)，用来描述、发现和鉴别数字化信息对象，如 MARC、DC，它主要描述信息资源的主题、内容特征，体现在所形成的记录上，每条记录都是对数据值和内容的表达。

②结构型元数据 (Structural Metadata)，描述数字化信息资源的内部结构，如层次、类属、目录、章节、段落的特征。

③存取控制型元数据 (Access Control Metadata)，用来描述数字化信息资源能够被利用的基本条件和知识产权特征，以及这些资源的期限和使用权限。

④评价型元数据 (Critical Metadata)，描述和管理数据在信息评价体系中的位置。

(4) 从元数据的层次体系角度可以分为以下元数据类型：

①赵军认为：广义元数据的层次体系由第一层至第六层依次为，信息内容格式元数据、内容对象元数据 (狭义元数据)、资源集合元数据、管理与服务机制元数据、过程与系统元数据、宏元数据^[13]。

②肖珑、赵亮认为：根据数字资源从产生到服务的生命周期、元数据描述和管理内容的不同以及元数据作用的不同，分为多种类型。内容元数据：描述数字对象内容及结构的元数据。专门元数据：描述单一数字对象 (如学位论文、古籍、网络资源、期刊论文等) 的内容、属性及外在特征的元数据。资源集合元数据：按照学科、主题、资源类型、用户范围、生成过程、使用管理范围等形成的信息资源集合 (如数据

库、知识组织系统等)的描述。管理元数据:数字对象的加工、存档、结构、技术处理、存取控制、版权管理以及相关系统等方面信息的描述。服务元数据:数字资源服务的揭示与表现、服务过程、服务系统等方面的相关信息描述。元元数据:对元数据的标记语言、格式语言、标识符、扩展机制、转换机制等信息的描述。上述体系中,信息内容元数据、专门数字对象元数据和资源集合元数据更多的是发挥对资源本身的内容、属性、外在特征的描述作用,称为“描述元数据”,另外几种则比较多地发挥了对资源和元数据的管理作用,称为广义上的“管理元数据”^[14]。

仔细考察上述的各种观点,可以看到,元数据的类型以描述型元数据与管理型元数据为主。对于描述型元数据基本上都有共识,而对于管理型元数据的认识,则有狭义与广义之分,狭义的理解是将管理功能与保存功能、技术特征和使用功能区分开,而广义的理解是将服务型元数据、技术型元数据、保存型元数据、存取控制型元数据等都作为管理元数据的一种。细化元数据的分类有利于按照不同的功能需求设置元数据的结构,而将管理元数据理解为广义概念,有利于对元数据进行管理和交换。

2 DC 元数据研究

2.1 概述

DC 元数据(都柏林核心元数据,Dublin Core Metadata),于1995年3月由OCLC与国家超级计算应用中心(NASA)联合发起,因创始地在美国俄亥俄府都柏林而得名。其目的是建立一套描述数字资源的标准。DC元数据是图书馆界在著录数字资源时使用的最为通用的元数据格式。

从诞生之日起,DC元数据经历了一个不断发展完善的过程。下面简要介绍一下DC元数据的发展历程。