二十一世纪圖書館学丛書

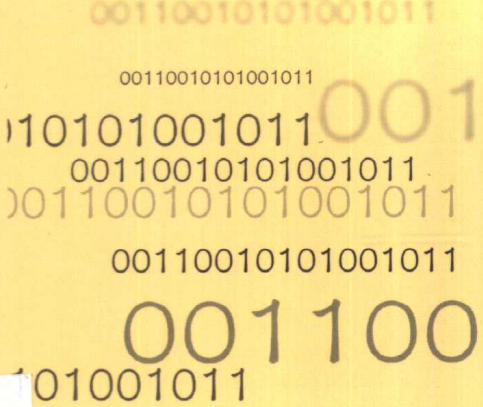
任继愈题

SERIES OF LIBRARY SCIENCE IN THE 21 CENTURY

丛书主编 丘东江

元数据导论

刘嘉 著



华艺出版社

需要全本请访问 www.ertongbook.com

6

G202/
L72

元数据导论

Introduction to Metadata

刘嘉 著

华艺出版社

图书在版编目（CIP）数据

元数据导论 / 刘嘉 著. — 北京：华艺出版社, 2002.1

ISBN 7-80142-370-4

Ⅰ. 元… Ⅱ.刘… Ⅲ.元数据-研究 Ⅳ. G201

中国版本图书馆 CIP 数据核字（2001）第 063946 号

元数据导论

华艺出版社

北京市海淀区北四环中路 229 号 10 层

邮编：100083 电话：82885151 转

北京市后沙峪印刷厂

850×1168 毫米 1/32 12.7 印张 242 千字

2002 年 1 月第 1 版 2002 年 1 月北京第 1 次印刷

定价：32.00 元

ISBN 7-80142-370-4 / Z • 188

《21世纪图书馆学丛书》

丛书题名

任继愈　国家图书馆馆长

顾　问

（以姓氏笔画为序）

马远良　上海图书馆馆长、研究员

马费成　武汉大学信息管理学院院长、教授、博士生导师

卢子博　南京图书馆原副馆长、研究馆员

卢克斯（Claudia Lux）博士　国际图联管理委员会委员、德国中央暨州立图书馆馆长

孙承鉴　国家图书馆原副馆长、研究员

孙蓓欣　国家图书馆副馆长、研究馆员

吴建中博士　国际图联管理委员会委员、上海图书馆副馆长、研究馆员

总　序

21 世纪是全球信息化的世纪，是知识经济占主导地位的世纪，是一个充满希望的世纪。它将对图书馆学情报学研究提出新的更高的要求。

丘东江同志主编的《21 世纪图书馆学丛书》旨在顺应时代潮流，把握知识经济和网络时代图书馆学情报学发展的脉络，强调精品意识和求实创新的学术风格，全力推出理论与实践相结合的最新研究著述。

《21 世纪图书馆学丛书》由国内外图书馆学专家学者合力而成，并特邀著名人士担任顾问；国家图书馆馆长任继愈为该丛书题写书名；国际图联执委、OCLC 总裁分别为部分著作题词、作序；国际图联秘书长和 UBCIM 计划主席和国外著名专家授予翻译权。

《21 世纪图书馆学丛书》以探讨图书馆学情报学最新理论与实践为主要内容，其中包括：《知识管理的理论与实践》、《国际图联（IFLA）与中国图书馆事业》（上、下）、《元数据导论》、《未来网络的基磐技术——XML 的理论与应用》、《Z39.50 的原理、应用与技术》、《OCLC：全球在线计算机图书馆中心》、《虚拟图书馆的昨天、今天和明天》、《国际标准书目著录（ISBD）》（上、下）、和《国外图书馆学重要著作选译》。

我希望这套丛书的问世，对我国图书馆事业的发展，提高图书馆学研究水平有所裨益。

吴慰慈

2001 年 7 月 30 日

需要全本请访问 www.ertongbook.com

序

如同过去的情形一样，在新技术刚开始出现时，人们往往惊诧于它能够解决一些以往的技术所无法解决的问题之能力，但随着新技术应用的不断推广，由此而产生的新问题甚至负面影响却会愈加明显、层出不穷，需有更新的技术和手段来解决。因特网自然亦属此类。一方面，网络在信息传递的速度、广度、数量等等方面，非过去任何传播媒体可望其项背，其信息自由发布的空间无远弗届，信息种类包罗万象，信息数量庞大无比且飞速增长；而另一方面，由于网络信息资源的组织与整理工具和手段跟不上网络信息快速增长的速度，因而无法使所有的信息为其真正的需要者所用。更何况网上的信息良莠并存，对网络信息资源的甄辨问题同样未能得到很好的解决。人类有别于动物的本质性差异便在于人类能够有意识地去发现问题，并积极地解决问题，从而推动人类社会持续地向前发展。就目前网络资源组织与整理方面所出现的问题，社会各界中凡有能力者，皆积极应对，提出各自的方案，而这自然也是体现自身价值的良机；同时，他们还在此基础上协同工作，以期建立优质的有序的全球网络信息环境。从早期的 Archie 和 Gopher 系统，到近期的搜索引擎和主题指南，都是这方面努力的成果。

搜索引擎和主题指南的运作方式基本上属于全文检索。这种运作方式固然可满足部分检索需求，但很明显地存在其他问题，

其中，以效率低下和无法有效过滤资源最为突出。因此，制定与发展一套标准的资料描述格式，用以准确、清晰地描述网络信息资源，便成为一个重要的课题。这正是元数据日渐受到重视的原因，使用元数据正成为组织网络信息资源的重要手段。

有鉴于此，本书通过采用历史回溯、文献探讨、个案分析、实证研究等研究方法，根据历史发展的脉络，在简述网络信息资源组织的发展历程之后，说明应用元数据成为网络信息资源组织的重要手段之必然性，进一步阐明元数据的定义、类型、功能和应用，并着重介绍两种典型的元数据格式。元数据也已被应用于中文网络信息资源的组织，笔者亦力求对中国大陆和台湾在此方面的情况做一较为完整的描摹。

在本书中，力求在以下方面有所突破：

- 在中国大陆，首次就元数据进行较为系统而全面的理论和实践方面的研究，介绍元数据的定义、类型、功能及其在整个世界范围内的广泛应用，并做深入探讨；
- 说明图书馆环境下的元数据格式--机读目录格式--仍然可以用于网络信息资源的组织，并进一步首次完整地介绍国际上已行之有效的使用机读目录格式为网络资源编目的条例和规则；
- 全面而系统地反映都柏林核心发展和应用的情况，阐明都柏林核心作为网络信息资源描述基本框架的重要地位及其与其它元数据格式的映射；
- 根据充分利用都柏林核心数据的体认，结合中国机读目录格式的特性，首次提出都柏林核心到中国机读目录格式的映射；

- 全面反映中国大陆和台湾元数据应用的现况以及业已取得的成果。

上述说是突破，事实上，更应该说是自己的一种努力的方向。然，以一己道行之浅薄，书中必多纰漏之处，还请各位专家和同道斧正。

本书主体之原形乃本人的博士学位论文。初稿的最后完成和修改是在医院里进行的，这使得原本不易的写作过程更添艰辛。而尽管如此，手捧沉甸甸的纸稿，心中的喜悦和释然自不待言。想去年留系后的苦况之下，竟还能成就如此的一点点果实，更是感慨万端。

首先，笔者对国家图书馆丘东江老师的不弃和提携致以深挚的谢意。若无丘老师的帮助，此书的付梓出版便尚属未知之数，而对困苦中人的相助，更令笔者十分感激。此外，深深感谢导师周文骏教授和师母多年来对笔者的孜孜教诲，师恩绵绵，泽惠终生。笔者还想借此向远在故乡上海的父母和弟弟以及陆廷鹤老师对笔者始终如一的支持，聊表感谢之意。

向学之路颇多坎坷和艰辛，吾自上下而求索，此心矢志不改。

2001 年 3 月 8 日于
北大畅春园

目 录

序……1

第一章 网络信息资源的组织……1

1.1 网络信息资源的界定……1

1.2 网络信息资源的类型……3

1.3 网络信息资源的特殊性……9

1.4 网络信息资源的组织模式及其评价……22

1.5 小结……36

第二章 元数据概述……41

2.1 何为元数据……41

2.2 元数据是全新的概念吗……48

2.3 元数据的类型与特征……51

2.4 元数据的作用……61

2.5 元数据的应用……64

2.6 利用元数据进行知识管理……70

2.7 与元数据相关的其它问题……75

2.8 小结……80

第三章　资源描述框架……85

3.1 何为资源描述框架……85
3.2 资源描述框架的数据模型……92
3.3 资源描述框架的语法……97
3.4 小结……101

第四章　元数据的编码标准……104

4.1 为何与如何为记录编码……105
4.2 常用的记录编码标准……106
4.3 小结……128

第五章　元数据与描述……129

5.1 国际标准书目著录……131
5.2 《英美编目条例》第 2 版 1988 年修订版……134
5.3 档案、个人论文和手稿记录……138
5.4 TEI 标题……138
5.5 政府信息定位器服务记录……140
5.6 数字地理空间元数据内容标准……143
5.7 VRA（可视资源协会）的可视资源核心种类……144
5.8 索引记录……145
5.9 使用过程中的记录……146
5.10 博物馆存取记录……149
5.11 小结……150

第六章　元数据的检索与检索控制……152
6.1 检索点的选择……153
6.2 检索（规范）控制……168
6.3 小结……182
第七章　图书馆界的元数据格式——机读目录格式……183
7.1 图书馆以 MARC 格式为网络信息资源编目的原因…184
7.2 网络信息资源编目项目……194
7.3 用于网络信息资源编目的美国机读目录格式……206
7.4 小结……245
第八章　都柏林核心元素集……250
8.1 都柏林核心的发展历程……251
8.2 都柏林核心元素集的元素和限定词……266
8.3 都柏林核心的特色和设计原则……280
8.4 都柏林核心的应用……285
8.5 都柏林核心与机读目录格式之间的关联……289
8.6 都柏林核心与机读目录数据的映射……292
8.7 小结……307
第九章　元数据在中国大陆与台湾的应用……313
9.1 都柏林核心到中国机读目录格式的映射……314
9.2 元数据在中国大陆的开发与利用……322
9.3 元数据在台湾的开发与利用……339
9.4 有关中文元数据开发与利用的思考与建议……346

需要全本请访问 www.ertongbook.com

9.5 小结……347

附录 1 三种元数据格式的数据元素和数据结构映射……352
附录 2 网络信息资源编目实例……359
附录 3 USMARC 记录、DC HTML 和 DC RDF 记录对照实例…367
附录 4 论文中所使用的重要缩写索引……376
附录 5 中国大陆与台湾网络资源组织方面用语对照表…378

参考与引用文献……381

第一章　网络信息资源的组织

近年来，因特网的迅速兴起对信息传播的方式产生了巨大的冲击；同时，因特网与万维网的结合使信息传播的障碍大大降低。由此而产生的效应之一，就是网络上的信息爆炸。面对因特网所提供的海量信息，人们起初无比惊喜；可在对因特网信息的存取和利用过程中，却并非一帆风顺，在某些情况下甚至感到步履维艰。尽管人们经常把万维网视作为“数字图书馆”，但事实上，它却缺乏人们对于一个图书馆所期许的有组织的特性。它发展得如此之快，以至于人们在对万维网中的网站进行编目或控制检索时已有力不从心之感[1]。人们不禁像 John Naisbitt 那样发出“我们正被淹没在信息中，却乏于知识”的慨叹。

为此，各界相关专家和学者在网络信息资源的组织中付出大量的努力，以期通过改善网络信息资源组织的状况，提高对此类资源的存取和利用。元数据是在计算机网络成为普遍的操作平台、人们为需要更有效地组织网络信息资源的情况下而提出的。有鉴于此，在探讨元数据之前，专设本章，就网络信息资源的定义、类型、特殊性及其组织做一较全面的阐述。

1.1 网络信息资源的界定

近年来因特网的迅速普及促进了网络信息资源数量的飞速增长。就这样一种日益增长的信息资源，出现了不同的概念用以表

达它的内涵。有的从“虚拟现实”这一用语引申出来，相对于现有的图书馆文献资源而言，把计算机网络中存在的信息资源称为“**虚拟资源**”（Virtual resource）。也有的使用“**联机资源**”（Online resource）一词。尽管这个词容易和以往的商用联机检索服务混淆，但是，由于现在的许多联机检索系统实际上已经通过因特网来提供各种信息服务，所以，可以认为两者之间的区别已经变得不那么重要了，完全可以用以往的联机网络提供的信息资源的延伸和扩展这个意义去理解它。而日本的一些图书馆学情报学研究人员则比较倾向于使用“**网络信息资源**”这一用语。在欧美，网络信息资源（Networked information resource）或网络资源（Networked resource）的使用也相当频繁。

事实上，“网络信息资源”和“虚拟信息资源”或“联机信息资源”在本质上并没有根本的区别。但是，“网络信息资源”要比另外两个概念更直观、更准确地表达了在本书中所讨论的对象。近年来，“网络经济”、“网络社会”等一些概念已经逐渐被大家接受，这些概念反映了 90 年代的社会和经济发展的新特征。“网络信息资源”这一概念则能够更直接地表达在这样一种新的社会和经济环境中信息资源的时代特征。

作为一个新的专业用语，尽管“网络信息资源”已经频繁出现，但是，它还没有成为一个专业术语，对它，并没有一个权威性的统一的定义。从字面上讲，网络信息资源一般可以被理解为“通过计算机网络可以利用的各种信息资源的总和”[2]。由于这个概念实际上是随着因特网的普及而导致的网络信息资源的开发和

利用的社会需要这样一个特定的背景下产生的，因此，黄纯元认为它实际上是指“通过因特网可以利用的各种信息资源”。[3]

而笔者以为，由于除了因特网以外，还有许多其它类型的网络，而这些网络上的信息资源，应当也属于网络信息资源组织的范畴之内，故将网络信息资源界定为“通过计算机网络可以利用的各种信息资源的总和”，最能完整地说明此概念的外延。诚然，在实际研究中，还是以对因特网上的信息资源的组织为重点，这是因为，因特网上的信息资源才是最能够被广泛存取与利用的资源。

网络信息资源和图书馆原有馆藏的最大不同，除了数字和非数字、拥有(Ownership)和存取(Access)之外，便是交互型资源的加入。图书馆过去常要为不同的用户群体编制“图书馆资源使用指引”（Library guide）或“信息地图”（Pathfinder），这些传统的“指引”或“地图”的范围通常只包括馆藏中既有的字辞典、百科全书、书目、索引等。网络上有各图书馆的联机公共查询目录、各种书目数据库可供检索，也有全文数据的电子书、浏览和检索的功能都在加强之中，网络时代的“信息地图”更加入了许多相关的论坛(Listerv)、新闻组（Newsgroup）等超链接，指引读者去参加各种进行中的研讨，获得更广泛的信息。网络资源已扩展、改写图书馆原有馆藏的概念。[1]

1.2 网络信息资源的类型

网络的信息包容量极大，几乎包含所有类型的信息资源。

OCLC 因特网资源计划小组曾将因特网资源加以分析，根据其分析结果显示，因特网的资源内容五花八门，从电子图书、电子期刊到非正式的个人信件都有，除传统资源以外，还包括：电子论坛、引文、公告、文献、抒情诗、自述文件、文章、草稿、手册、推荐信、会议录、报告、议案、社论、月报、评论、通讯、标准、札记、答辩书、事实文件、诗、摘要、小册子、政策、调查、公报、宪章、剖面图、见证、会议论文、幽默小品、提案、指南、气象资料、学报、宣传、研讨会等数十种不同类型的资源。[5] 因特网上的信息资源丰富多彩，正是其流行的主要原因之一。

从不同的角度出发，人们对网络信息资源进行了不同的类型划分。

1.2.1 从图书馆学情报学的角度出发

一些研究人员从图书馆学情报学的角度对网络信息资源进行类型化和体系化的研究。例如，日本的户田慎一把网络信息资源分成7类：

- ✧ 电子期刊、电子通讯期刊、图书的文本；
- ✧ 论文的抽印本、技术报告；
- ✧ 法律文件、判例、政府出版物；
- ✧ 数值数据、统计资料、实验数据；
- ✧ 软件；
- ✧ 图像数据、声音数据；
- ✧ 数据库。[6]

这种分类方式实际上是沿用了图书馆学情报学对非网络环境中的文献资料类型分类的方法，只是把它延伸到了网络环境中去，但是他并没有充分揭示网络信息资源分类的特点。网络信息资源区别于非网络信息资源的一个重要的特点，是网络信息资源的存取和利用方法的多样性。在非网络环境中，文献资料的存取和利用则相对比较单一。而且，网络环境下存取信息资源的方式和这些方式所能获取的信息资源的类型和范围有着密切的关系。如，通过远程登录方式和通过网络新闻小组所能获取的信息资源的类型和范围是有很大区别的。因此，在讨论信息资源的分类时，不能脱离网络的存取手段。

1.2.2 从网络信息资源的存取方式出发

海野敏认为网络中的信息存取方式尽管使用了比较复杂的技术，但是，它所要达到的功能却是简单地模拟和再生人类传播方式。所以，他使用了模拟非网络环境中信息存取方式的方法对网络信息资源进行归类，提出了以下 6 种网络信息资源的类型。

- ✧ 邮件型　邮件型的信息存取方式是以电子邮件和电子邮件群体服务（Mailing list）为代表的。
- ✧ 电话型　它是指以特定的个人或群体为对象即时传播信息的方式。代表性的手段有会话（Talk）和交互网中继对话（IRC—Internet Relay Chat）。
- ✧ 揭示板型　它是以不特定的大多数网络利用者为对象的非即时的信息传播方式。比较具有代表性的是网络新闻

需要全本请访问 www.ertongbook.com