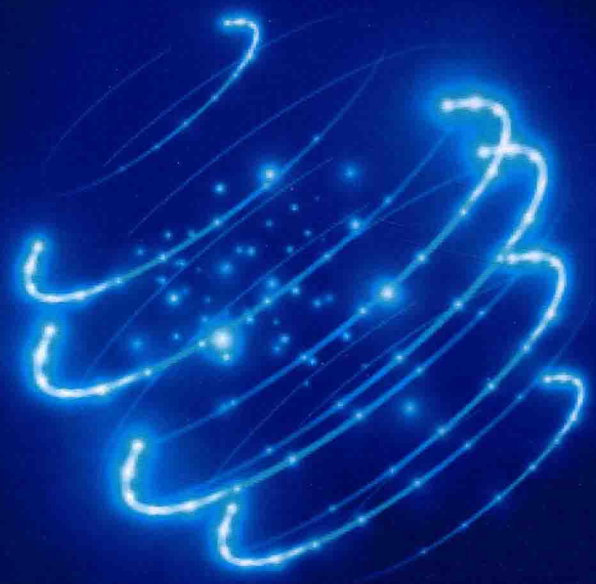
情报学视域下的数据研究：理论、原理与方法

Data Research from the Perspective of Information Science : Theory, Principles and Methods

曹祺 著



WUHAN UNIVERSITY PRESS
武汉大学出版社

情报学视域下的数据研究：理论、原理与方法

Data Research from the Perspective of Information Science : Theory, Principles and Methods

曹祺 著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

情报学视域下的数据研究:理论、原理与方法/曹祺著.—武汉:武汉大学出版社,2018.10

ISBN 978-7-307-20597-0

Ⅰ.情… Ⅱ.曹… Ⅲ.①情报学—数据管理—研究 ②情报学—数据处理—研究 Ⅳ.G250.2

中国版本图书馆 CIP 数据核字(2018)第 238236 号

责任编辑:陈 帆 责任校对:汪欣怡 整体设计:马 佳

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)

印刷:北京虎彩文化传播有限公司

开本:787×1092 1/16 印张:14 字数:329 千字 插页:1

版次:2018 年 10 月第 1 版 2018 年 10 月第 1 次印刷

ISBN 978-7-307-20597-0 定价:58.00 元

目　录

第1章 引 言

情报学的概念源于欧美国家，情报学是研究情报的产生、传递、利用规律和用现代化信息技术与手段使情报流通过程、情报系统保持最佳效能状态的一门科学。

在情报学领域，“1997年，美国管理科学家托马斯·达文波特首次提出信息生态学的概念，将生态理念引入信息管理中，从而开辟了信息管理的新领域。信息生态科学作为一门新兴的生态学研究前沿领域，已经得到了国内外科技界的广泛关注”。①

信息生态学面临的主要问题是信息生态的失调(Information Ecological Imbalance)，陈曙将信息失调分成信息超载(Information Overload)、信息垄断(Information Monopoly)、信息侵犯(Information Encroachment)、信息污染(Information Pollution)四类，并提出“信息生态系统自身不能不是对立的统一。对于任何图书情报机构而言，这种对立统一始终制约着以文献为载体的信息的生产和消费、信息的储存和传递、信息的民主和法制以及信息的污染和净化等”②。

信息超载指的是每天增加大量的新数据。王云梅指出：“目前，国际互联网几乎覆盖了全球所有的国家和地区，用户高达3亿以上。全世界每年出版70万种期刊，6万多种新书，新增期刊近万种，发表科技文章500多万篇，编写学术报告或学术论文25万多份，登记专利40多万项。仅在美国每年就有5万本图书出版，全球出版的各级各类报纸，估计数目在40万左右。互联网上的主页已达到1.3亿页，并以每天10万页、2000万单词的速度递增。”③

本书的研究对象为元数据(含标识符数据)和全文数据。

情报学关注的数据主要是论文和专利，因此本书的数据以专利数据为主。元数据的研究内容包含歧义消除问题、溯源管理问题。全文数据的研究内容包含数据存储、数据渲染和数据分析。同时结合最新的区块链技术，基于区块链技术背景，对数据管理进行探索研究。

---

① 杜欣明：《信息生态学的学科建设与发展问题初探》，《现代情报》2006年第7期。

② 陈曙：《信息生态的失调与平衡》，《情报资料工作》1995年第4期。

③ 王云梅：《信息生态系统及其有效机制的构建》，《图书馆工作与研究》2010年第2期。

第 2 章　元数据管理研究

2.1　Handle 系统

2.1.1　Handle 系统背景及应用

资源唯一标识符是一类统一资源标识符(URI)的统称，其目的是通过对数字对象的公认标准管理来永久性地标识某一个数字化对象。Handle 系统是众多资源唯一标识符中应用最广的一种。Handle 系统出现于 1995 年，由 TCP/IP 协议的联合发明人之一，被称为互联网之父的 Robert Kahn 在美国国防部高级研究计划署(DARPA)资助下由美国 CNRI(Corporation for National Research Initiatives)提出并实现的，是一种基于互联网的分布式数字对象命名与标识系统。CNRI 是一个非营利组织，承接促进公众利益方面的项目，对基于网络信息技术的战略发展进行研究，并提供资金支持。2005 年，美国国家创新研究所(CNRI)公开了 Handle 6.1 系统，并进行了源代码开源，允许商业机构应用，同时简化了 Handle 前缀申请。①

Handle 编码是“一套集命名、注册与解析功能模块的完备标志解析系统，是一种智能化的标识码”②。Handle 系统采用两级的服务模式，包括全球 Handle 命名服务 GHR(Global Handle Registry)和区域 Handle 命名服务 LHS(Local Handle Services)。其中，全球服务负责管理 Handle 的命名空间，系统提供唯一的 Handle 注册机制，任何一种现有的本地命名空间都可以通过这种注册机制加入到 Handle 系统中，成为全球性的命名空间，通过这种命名机制还可以将诸如解析和管理之类的服务授权给本地。本地服务主要由命名机构提供，如规定地方命名空间，优化 Handle 系统整体效果，对相关命名授权下所有的 Handle 进行管理并提供本地名称的解析服务，等等。

Handle 系统最大的特点是网站和服务器的数量没有限制，系统定义了一种层次型的服务模型，任何一个本地命名空间的服务既可由本地服务来提供，也可由全球服务来提供，或由两者共同提供，每个命名机构提供的服务不必相同，这种分布式的服务体系降低了系统的风险。Handle 系统通过自定义的协议保证数据的完整性和保密性，可在网络上进行安全的解析和管理服务。每个 Handle 都可以定义自己的管理者，管理者可通过以自

---

① 郭晓峰、孙洵：《Handle 系统的发展及应用》，《数字图书馆论坛》2013 年第 8 期。

② 邹慧、马迪、王伟、刘阳、毛伟、邵晴：《Handle 系统与域名系统互联互通机制：一种基于标记语言描述协议数据单元的实现》，《计算机应用研究》2019 年第 1 期。

己角度定义的 Handle 所有权对网络上任意位置的 Handle 进行安全高效的管理。

Handle 系统目前的应用包括 DOI System, Entertainment Identifier Registry (EIDR), CORDRA/ADL, Global Environment for Network Innovations (GENI), DSpace-Digital Repository System, Handle 系统-Globus Toolkit Integration Project, The National Digital Library Program(NDLP), 对象内容涉及电影和电视剧、内容对象仓储、虚拟实验室关于未来互联网的可视化和新服务的应用、数字仓储中资源长期保存、国会数字图书馆的馆藏标识等。目前, 全世界运行着成千上万的 Handle, 该服务广泛应用于各大团体机构, 如用户联盟、国家图书馆、国家实验室、大学、计算中心、政府机构、公司和研究团体等。通过 Handle 标识的对象有期刊论文、技术报告、图书、学位论文、政府文件、元数据、分布式学习内容和数据集, 并将应用于数字水印、GRID、仓储等更多的领域。"美国洛斯阿拉莫斯国家实验室(Los Alamos National Laboratory)使用 Handle 系统管理了超过 6 亿件非公开内部资料。澳大利亚国家数据服务(Australian National Data Service, ANDS)基于 Handle 系统管理科研数据。欧洲持久标识符联盟(European Persistent Identifier Consortium, ePIC)基于 Handle 构建了面向欧洲研究社区, 为成员机构提供注册服务。中国国家物联网标识公共服务'国家物联网标识管理公共服务平台'由中科院计算机网络信息中心牵头, 联合工信部电子科学技术情报研究所(ETIRI)、工信部电信研究院、中国物品编码中心基于 Handle 建立物联网标识统一管理和公共服务系统。"①

2.1.2 Handle 系统的命名空间定义

作为一个标识符系统, Handle 系统把标识符统称为 Handle。Handle 系统名称空间(Namespace)定义了 Handle 的构成法则。Handle 是由不同字符构成的字符串, 系统中的每个 Handle 由两部分组成: Handle 的命名授权部分(Naming Authority, 或视为前缀)和跟随其后的在该命名授权下唯一的本地名称(Local Name, 或视为后缀)。命名授权(简称 NA)和本地名称间通过 ASCII 字符"/"(0x2F)来分开。前缀部分用于命名授权, 即后缀命名的授权和规范命名; 后缀为本地名称, 用于特定形式资源的规范化命名。从 Handle 系统的特点来看, 其关键在于命名权的统一管理, 强调了信息身份的标准化识别和管理。图 2-1 给出了以 ABNF 表示法②来定义的 Handle 语法构成。

举例来说: 10.45/abc 是一个符合 Handle 语法的标识符, 它可以在互联网范围内唯一标识某个资源对象。这个标识符的命名授权(前缀)分为两级: 10 是顶级命名授权, 45 是 10 的下一级, 这类似于域名的分级制度。整个命名授权 10.45 通常分配给某个机构, 而本地名称 abc 可以是该机构内部生成的一个用来标识自有资源的号码。

基于 Handle 结构的标识符最基本的命名原则是唯一性。在 Handle 系统中, 唯一标识符 Handle 是由前缀和后缀两部分构成的, Handle 的唯一性因此也由前后缀一起来达成。在不同的前缀下, 后缀可以相同, 但在同一个前缀下, 所有的后缀必须是互不相同的。后

① 罗鹏程、崔海媛、聂华、朱玲、韦成府:《高校图书馆持久标识符应用研究》,《大学图书馆学报》2017 年第 5 期。

② Crocker D, Augmented BNF for Syntax Specifications: ABNF, 1997.

```
<Handle>=<Naming Authority>"/"<Local Name>
<Naming Authority>=* (<Naming Authority>".") <NAsegment>
<NAsegment>=1* (%x00-2D/%x30-3F/%x41-FF
        ; 任何可以映射成UTF-8编码的
        ; Unicode 2.0字符的8位字节
        ;除了0x2E和0x2F (对应于ASCII字符'.' 和'/')
<Local Name>=* (%x00-FF)
        ;任何可以映射成UTF-8编码的
        ; Unicode2.0字符的8位字节
```

图 2-1　Handle 语法构成

缀的唯一性首先由唯一标识符的注册者在生成标识符后缀时来确认，在最终注册时会由 Handle 系统来保障整个标识符的唯一性。从 Handle 系统名称空间的定义上来看，其结构简单，易于实施，对于前后缀构成几乎没有约束，因此可以很好地继承现有的一些标识符命名规则。此外，它规定使用 Unicode2.0 字符集及 UTF-8 编码，可以实现国际化支持。唯一标识符生成后，需要在 Handle 系统中注册才能生效。Handle 系统负责对唯一标识符及相关信息进行管理，并提供一套机制对唯一标识符进行解析。

2.1.3　Handle 系统解析原理

Handle 系统采用一种分布式、可伸缩、可扩展的结构，通过 Handle 协议将系统的各个服务组件联系起来，其整体架构如图 2-2 所示。

Handle 系统定义了一套分层的服务模型，其整体模型是由许多 Handle 服务来构成的，处于 Handle 系统顶层的服务称为全球 Handle 注册中心。Handle 系统中所有的命名授权均由 GHR 来管理，只有在 GHR 中注册后，命名授权才能够生效。

在系统中，命名授权是作为 Handle 来管理的，这样的 Handle 称为 NA Handle。NA Handle 为客户端访问和利用 Handle 服务组件提供必要的信息。在 GHR 下分布了很多其他的 Handle 服务，通常被称为区域 Handle 服务，每个 LHS 服务管理着 Handle 系统下的一个子名称空间(Sub-Namespace)，不同 LHS 服务下的名称空间互不重叠，子名称空间通常由一些命名授权下的 Handle 集组成，负责这些命名授权的 Handle 服务称为主服务(Home Service)，并且是唯一的为这些命名授权下的 Handle 提供解析和管理服务的。

由于命名授权存在分级关系，所以对命名授权负责的 LHS 也存在多层的上下级关系。LHS 实际上是一个逻辑上的服务概念，一个 LHS 由一到多个服务站点(Service Site)构成，不同的站点可以分布在不同的地域，而同一个 LHS 服务下的各服务站点的功能是一样的，可以将它们视为镜像关系。同一个 LHS 下服务站点数量的多少可以根据实际需要来配置，

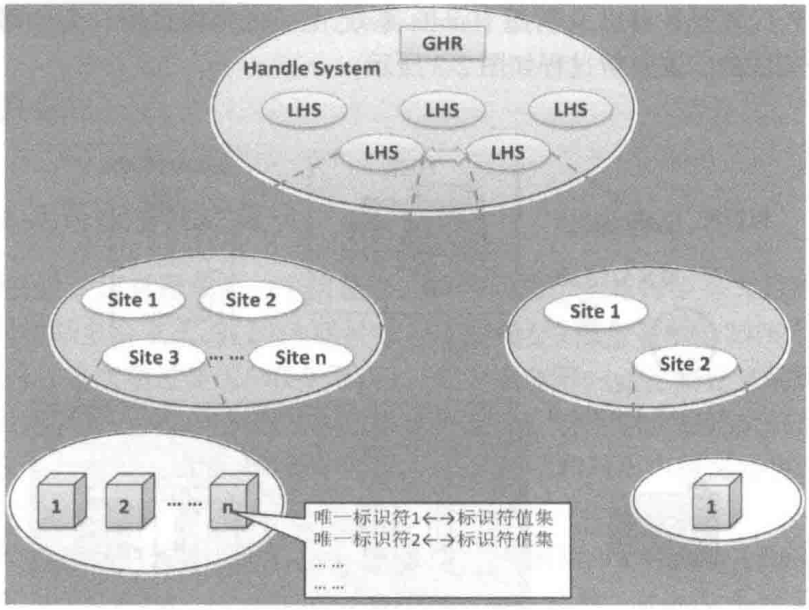


图 2-2 Handle 系统整体架构示意图

每个服务站点最终由若干台 Handle 服务器来构成，发送至服务站点的 Handle 请求最终被分发给这些 Handle 服务器。一个 LHS 所负责的子名称空间下所有的唯一标识符及与标识对象相关的信息(这里称为标识符值集)存储在这些 Handle 服务器上，由其响应用户请求并返回相应结果。

Handle 系统可以由任意数量的 Handle 服务组成，而构成 Handle 服务的服务站点的数量从设计上来说并没有限制，同样对构成服务站点的服务器的数量亦无限定。服务站点间的复制并不要求每个站点包含同样数量的服务器，换句话说，只要每个服务站点拥有相同的复制了的 Handle 集，则每个站点可以将这些 Handle 分布在不同数量的 Handle 服务器上。这种分布式方法为系统适应任意数量级的操作提供了可伸缩性，并且可以减轻或避免系统单点出错造成的危害。

2.1.4 Handle 系统解析实例及特点

基于 Handle 的唯一标识符由命名授权和本地名称构成。一个机构如果需要为自己的资源注册唯一标识符，首先需要向注册代理机构申请某个级别的命名授权，类似于域名的申请；一旦获得命名授权，机构便可以将资源的内部标识(需要具有内部唯一性)作为本地名称与命名授权结合成一个 Handle，并在 Handle 系统中注册。

在 Handle 系统中，不仅维护着众多的唯一标识符，更重要的是维护唯一标识符所标识对象相关的一些信息，这些信息在解析时可以返回给用户，以便告知被标识对象是什么样的资源，如何获取该资源及相关的服务，等等。

唯一标识符及相关信息的注册可以使用管理工具批量进行。数据注册完成后，唯一标识符可以在互联网上发布，用户便可以利用各种解析途径来对唯一标识符进行解析。

目前支持的解析途径包括：使用专门的客户端管理工具，使用装有解析插件的浏览器，使用 HTTP 代理服务器以及利用 Handle 系统 API 来编程解析。无论哪种解析方式，其基本原理是相同的，其解析过程如图 2-3 所示：

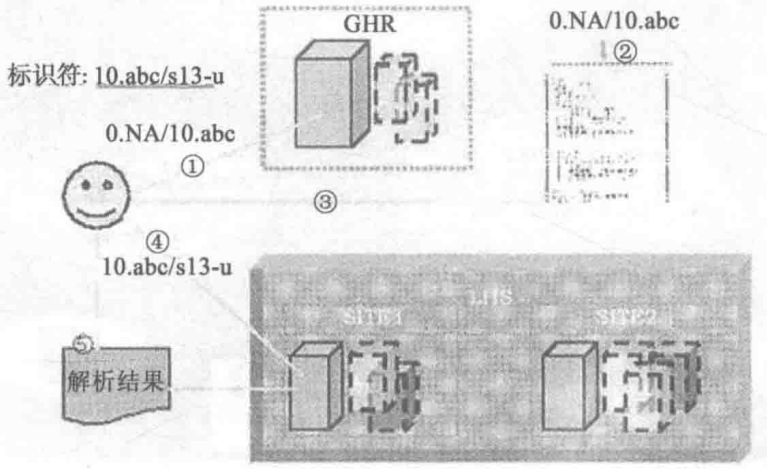


图 2-3　Handle 系统唯一标识符解析过程实例图

假设用户在浏览互联网时遇到一个唯一标识符为 10. abc/s13-u，它标识着某篇文献，用户希望解析该标识符来得到与文献相关的信息，或者下载全文，需经过如下过程：

(1)用户的浏览器首先将该唯一标识符的 NA Handle：0. NA/10. abc 发送给 GHR 系统；

(2)GHR 查询其 Handle 数据库，找到负责命名授权 10. abc 的 LHS 服务信息，该服务信息描述了 LHS 的构成，包括该 LHS 由哪几个站点构成，每个站点包含几台服务器，各服务器的服务地址和端口是什么，该命名授权下的 Handle 如何在这些服务器上分布，等等。

(3)客户端浏览器得到该服务信息后，据此判断 10. abc/s13-u 这个标识符应该发往服务站点中的哪台服务器去解析。

(4)浏览器向对应的 Handle 服务器发出解析请求，在此期间，可能会根据需要对通讯双方进行认证。

(5)Handle 服务器根据客户端的请求从数据库中检索出与唯一标识符对应的信息，将解析结果返回给用户。

与其他的解析系统或机制相比，Handle 系统的优势主要在于：

(1)命名系统灵活，与 URN 兼容，可保持标识符的唯一性及持久性。

(2)基于 Handle 的命名机制可以包容现有的标识符方案。

(3)内建一套完善的 Handle 协议来支持对 Handle 的解析。

(4)对单个 Handle 可实现多重解析。

(5)Handle 命名和 Handle Protocol 均实现国际化支持。

(6)分布式的服务和管理模式。

(7)安全而高效的解析和管理机制。

Handle 系统的这些特点使其明显优于其他解析方案。

2.2 DOI 系统

2.2.1 DOI 系统的背景及应用

“面对科研行业，信息超载是一个信息生态的大问题，全球每年有大量的学术论文发表，美国国家科学基金会发布的《2018 年科学与工程指标》报告显示，2016 年中国发表学术论文 42.6 万份，首次超过美国(40.9 万份)，成为全球第一。”这些论文是科研工作者在科研过程中不可或缺的研究资料，为了对论文进行定位、访问以及对论文的元数据进行管理，在这种背景下，由国际出版商协会，国际科学、技术和医学出版商协会，美国出版商协会共同倡议并创建了 DOI 系统。

DOI 系统于 1997 年发布于法兰克福书展，由国际 DOI 基金(International DOI Foundation，IDF)对 DOI 系统进行管理和功能完善。IDF 认为，Handle 系统具有迄今为止最完善的数字对象管理架构，因此 DOI 选择基于 Handle 系统来进行研发。但对于管理知识内容、促进电子商务建设的 DOI 系统，还需要在 Handle 基础上增加新功能以完善其框架。IDF 作为一个国际组织，成立董事会对其不同类型的会员机构进行管理，董事会成员由会员机构选举产生。会员机构包括创始会员、一般会员、注册机构会员和附属会员，每种类型会员的职责和权限各不相同。其中注册机构会员负责 DOI 系统的维护、DOI 号码的分配和保存、相关政策的制定以及数据库的使用和维护，注册机构只对 IDF 会员提供服务，其成员有资格参选成为董事会成员。DOI 系统最初只服务于文字出版类资源，作为数字环境下进行版权管理和保护的工具。公众认为，DOI 系统是一个能够胜任管理和识别数字网络内容、标识整合数字资源和多媒体应用的通用框架。DOI 系统建立后，IDF 选择 CNRI 作为其技术合作伙伴，且从 1998 年开始参与 INDECS 项目，INDECS 框架支持 DOI 数据模型。IDF 持有词汇映射框架 VMF(Vocabulary Mapping Framework)站点且参与其管理，IDF 的数据字典是 VMF 的一个命名空间。

2000 年，DOI 语法通过了国际标准化组织(International Organization for Standardization，ISO)标准化，2012 年 5 月 10 日通过了国际 DOI 基金会的《信息文档数字标识符系统标准》(Information and documentation—Digital object identifier system)，即 ISO26324 标准。该标准规定了数字对象标识符系统的语法、描述方式和解析功能组件以及 DOI 名称的创建、注册和管理的一般规则。

从 2012 年 5 月至 2018 年 7 月的 6 年中，DOI 标识符得到了广泛的应用。在国外，据 DOAJ 统计，开源期刊中采用了 DOI 标准的有全球 128 个国家的 11843 本期刊，共标识了 3216223 篇文章。在国内，根据中国科学技术信息研究所统计，国内期刊采用 DOI 标准的文章共计 28864962 篇。

DOI 标准之所以得到广泛的推广是基于 DOI 标准建立的学术信息生态更高效，能对大量的信息进行精确的标注：

(1)对于作者而言，优势在于通过 DOI 编码高效管理参考文献。

(2)对于读者而言，优势在于通过 DOI 编码能快速搜索到需要查阅的论文。

(3)对于图书馆而言，可以作为馆藏数据持久化和数据分析的一个字段。

从构成要素上来说，DOI 包括 4 个组成要素，即标识符、元数据、解析系统和规则。通过这些要素，DOI 能够提供数字对象与其元数据，提供数字对象与数字对象(逻辑上相关)之间具体物理位置的链接。

2.2.2　DOI 系统的命名空间定义

根据美国标准 ANSI/NISOZ39. 84-2000DOI 的编码方案规定，DOI 是由前缀和后缀两部分组成，其结构式为：<DOI>=<DIR>. <REG>/<DSS>。

由于编码规则对前缀与后缀的字符长度没有任何限制，因此，理论上 DOI 编码体系的容量是无限的。① DOI 前缀由两部分组成，一个是目录代码(Directory code，DIR)，为 DOI 的特定代码，其值为 10，所有 DOI 代码都以“10.”开头，用以将 DOI 与 Handle 系统技术的系统区别开。另一个是登记机构代码(Registrant's code，REG)，是 DOI 注册代理机构的代码，由 DOI 的管理机构——国际 DOI 基金负责分配，由 4 位阿拉伯数字组成。DOI 后缀(DOI suffix string，DSS)由 DOI 注册代理机构(registration agency，RA)自行给出，是一个在特定前缀下唯一的后缀，其编码方案完全由登记机构自己来规定，规则不限，只要在该出版商的所有产品中具有唯一性即可，是对数字对象定义的本地标识符。后缀可以是一个机器码，或者是一个已有的规范码，如国际标准书号(International standard book number，ISBN)或国际标准连续出版物编号(International standard serial number，ISSN)。DOI 的命名结构使每个数字资源在全球具有唯一的标识。

以《中国科技资源导刊》2017 年第 49 卷第 4 期第 4 篇文章为例，其 DOI 标识为：10. 3772/j. issn. 1674-1544. 2017. 04. 004。其中“10.”为 DOI 的特定代码，3772 为《中国科技资源导刊》DOI 注册代理机构的代码，J 为杂志(Journal)缩写，issn. 1674-1544 为《中国科技资源导刊》的 ISSN 号，2017. 04. 004 为 2017 年第 4 期第 4 篇。

DOI 标识符系统解析对应的元数据是 DOI 系统的组成要素，是促进 DOI 系统服务多样化的必需要素，是有效管理数字权益的基础。一个完整的标识系统，不仅要标识其在网络上的入口位置，还要有该位置上对象的具体信息，如所描述资源的题目、载体、作者等相关信息。所有注册的 DOI 都要求具有最低限度的核心元数据的声明，并且公开发布，允许任何用户访问。这种公开是单向的，也就意味着任何用户都可以免费查询其所对应的元数据，但是如果需要从相关的元数据，如题名、作者等，反向查询对应的 DOI，目前尚无法实现。②

① 张光威：《提高论文引用率行之有效的工具——数字对象标识符(DOI)》，《海洋地质与第四纪地质》2008 年第 4 期。

② 宋丹辉、徐宽：《数字对象唯一标识 DOI 的发展与应用研究》，《图书馆学研究》2006 年第 8 期。

2.2.3 DOI 系统解析原理

DOI 以两种技术为基础：Handle 系统和 Indecs 元数据。Handle 系统是用于互联网信息的命名、解析和管理的技术平台；Indecs 元数据是用于在电子商务环境下实现数据互操作的元数据框架。Handle 系统技术包含了多重解析(multiple resolution)的功能，即一个 DOI 不仅指向一个统一资源定位符(uniform resource locator，URL)，还可以指向多个 URL，以及 URL 以外的其他各种类型的元数据。实际上，DOI 可以认为是一种统一资源标识符(universal resource identifier，URI)或统一资源名称(universal resource name，URN)，是信息的数字标签和身份证。DOI 的多重解析功能，使得在解析出多个 URL 时，可以选择离用户最近的一个镜像站点下载数据，同时，还能链接到该资源的许多相关信息。多重解析不仅确保了对资源的访问，而且有利于资源的深度利用。①

DOI 系统的解析机制基于 Handle 系统，因此 DOI 编码兼容 Handle 编码。但是 DOI 系统主要是应用于出版行业，DOI 编码系统继承了 Handle 编码系统的特性，但是又有进一步的发展。

从出版流程而言，DOI 编码分为期刊注册、文章注册。例如，对于 DOI 的顶级注册商台湾华艺公司会先根据出版社提交期刊的 ISSN 号申请注册 REG 代码，完成期刊注册。出版社在每出版一本期刊时，对每一篇文章注册 DSS 代码并发布完整的 DOI 号码，然后华艺公司会同步传给国际 DOI 基金会(IDF)。DOI 系统和 Handle 系统不同，肖红②等相关学者对此作了比较，如表 2-2 所示：

表 2-2　　国内外唯一标识符系统在管理方面的对比[2]

项目	Handle 系统	DOI	公共图书馆唯一标识符系统
发起机构	美国 CNRI(非营利组织，美国国防部高级研究计划署资助)	国际出版商协会，国际科学、技术和医学出版商协会，美国出版商协会	文化部
管理者	美国 CNRI(非营利组织)	IDF 基金会(非营利组织)	国家中心(挂靠国家图书馆)
国家参与	否	否	国家中心代国家管理
系统服务对象	更广泛的团体和机构	个人或机构	国内公共图书馆
范围	全球	全球	中国
服务模式	两层服务(GHR 和 LHR)	对不同类别会员提供不同服务	为公共图书馆提供唯一标识符注册服务

① 莫琳芳、李喆、林永丽、王映红、张阵阵、甘辉亮：《数字对象唯一标识符的应用与发展现状》，《海军医学杂志》2016 年第 4 期。

② 肖红：《国内外数字资源唯一标识符系统对比研究》，《图书情报导刊》2016 年第 6 期。

续表

项目	Handle 系统	DOI	公共图书馆唯一标识符系统
系统权威性	业界著名、功能完善	业界著名、功能完善	运行初期
行业标准	应用广泛，无标准	成为 ISO 标准	国内无标准
资金来源	资助、服务费和注册费	注册费	国家专项费用
具体服务单位	全球或本地的注册机构	签约的注册代理	组织结构人员
成员分类	两个大的层级	分多种会员，每种权限不同	只服务于公共图书馆
会员收费	收	收	否
资源注册费	收	收	否

此外，从编码内容而言，DOI 编码系统本身由国外的 IDF 研制完成，并不完全符合中文文献。因此，2012 年，文化部在兼容 ISO26324 标准的基础上，提出了数字对象唯一标识符规范标准，即《WH/T 48-2012 标准》，编码也叫做 CDOI 编码。

根据《WH/T 48-2012 标准》的分类，对于古迹、拓片，DOI 编码并不涉及，如表 2-3 所示：

表 2-3　　信息资源名称规范列表①

资源名称	Collection Name
古籍	rarebook
舆图	atlas
拓片	rubbing
家谱	genealogy
地方志	chorography
期刊论文	JNArt
会议论文	ConPaper
学位论文	ETD
电子图书	ebook
音频资源	audio
网络资源	websit

① 欧阳宁、王莹、谢丽佳：《数字对象唯一标识符 CDOI 探析》，《图书馆理论与实践》2018 年第 3 期。

国内学者在 DOI 的基础上提出了 CDOI，CDOI 的系统架构如图 2-4 所示：

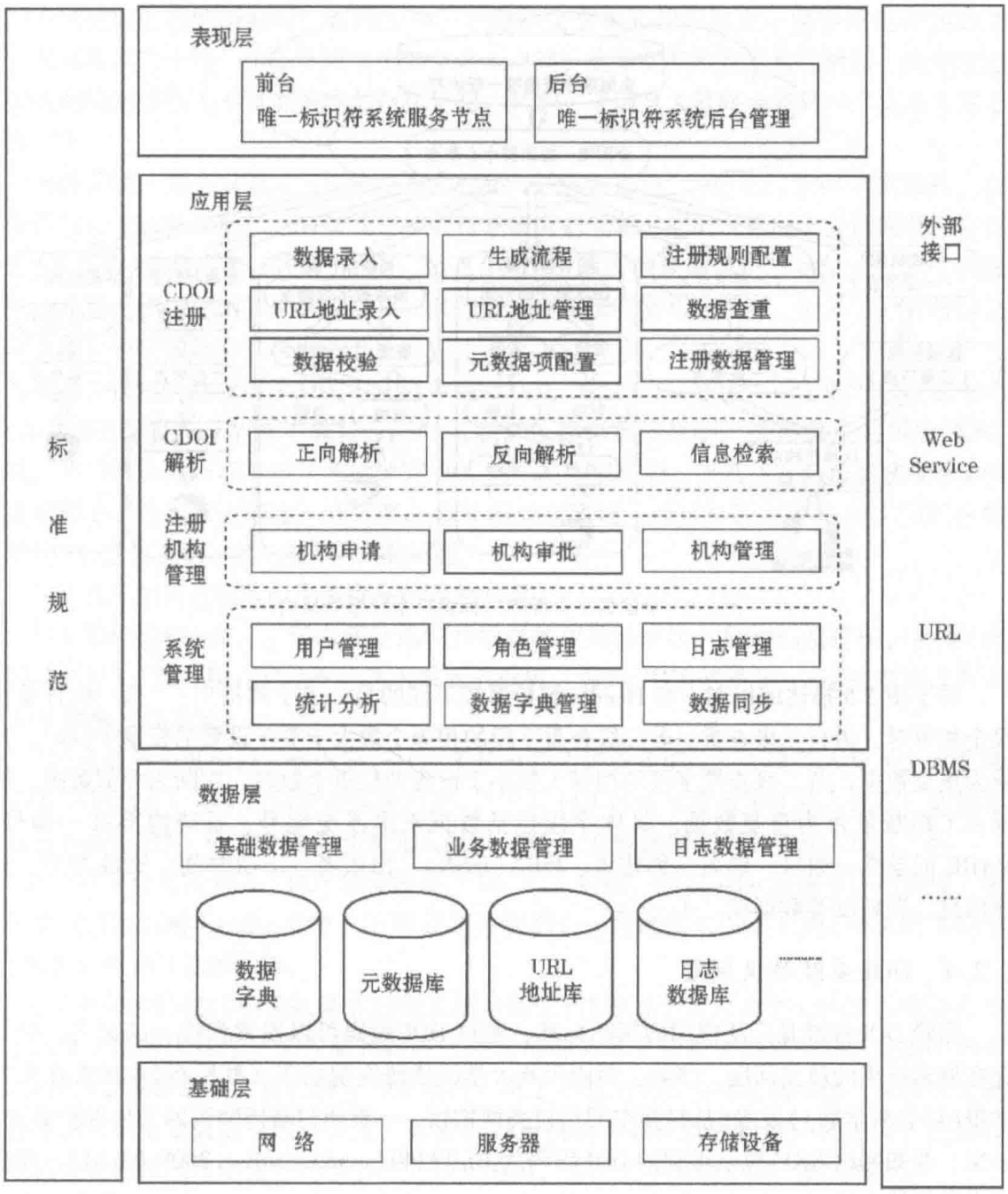


图 2-4 CDOI 系统架构图①

① 童忠勇、李志尧、孙秀萍：《国家数字图书馆数字资源唯一标识符系统的设计与实现》，《图书馆学研究》2013 年第 21 期。