

元数据标准规范 建设及应用

国家图书馆 萨蕾

➤ 元数据概述

➤ 描述性元数据

➤ 元数据应用

元数据概述

定义

- ◆ 关于数据的数据
- ◆ 关于信息资源或数据的一种结构化的数据。
- ◆ 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)。
- ◆ 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。

元数据概述

- ◆ 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。
- ◆ 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，确定孰为其想要浏览或检索的信息包，而无需检索大量不相关的全文文本。

元数据概述

- ◆ 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。

元数据概述

类型

描述性元数据 (descriptive metadata)

管理性元数据 (administrative metadata)

结构性元数据 (structural metadata)

元数据概述

管理型元数据

描述型元数据

保存型元数据

技术型元数据

使用型元数据

元数据概述

知识描述性元数据
(Intellectual Metadata)

结构型元数据
(Structural Metadata)

存取控制型元数据
(Access Control Metadata)

评价型元数据
(Critical Metadata)

元数据概述

功能

- ◆ 查找（**find**）符合用户检索要求的实体（即利用实体的属性或关系在一个文档或数据库中找到一个或一组实体）；
- ◆ 识别（**identify**）一个实体（即确认所描述的实体对应于所查找的实体，或者区分具有相似特征的两个或多个实体）；
- ◆ 选择（**select**）适合用户需要的一个实体（即选取一个在内容、物理形式等方面能满足用户要求的实体，或放弃一个不适合用户需求的实体）；
- ◆ 获取（**acquire**）或存取（**obtain access**）所描述的实体（即通过购买、借阅等方式获取一个实体，或者以电子方式通过联机连接远程计算机来检索一个实体）

元数据概述

特点

- ◆ 模块化
- ◆ 可扩展性
- ◆ 开放性

描述性元数据

- 一、MARC元数据
- 二、DC元数据
- 三、MODS元数据
- 四、元数据互操作

MARC元数据

MARC (Machine-Readable Cataloging, 计算机可读目录)

是一种为描述、储存、交换、处理及检索信息资源而设计
的标准，采用编目条例作为数据建模的规则，以机读
目录格式作为编码方案，以ISO2709作为交换格式。

MARC元数据

MARC优点

- ◆采用独立的元数据标识系统
- ◆采用纯文本文件
- ◆多年来得到持续的发展与完善
- ◆在世界各图书馆得到普遍应用，积累了大量的元数据

MARC元数据

发展

- ◆1969年，美国国会图书馆发布MARC磁带
- ◆1973年，MARC成为国际标准（ISO2709）
- ◆1977年，UNIMARC（通用MARC格式）出版
- ◆1996年，CNMARC作为行标(WH/T0503-96)开始实施
- ◆1999年，MARC21出版

MARC元数据

◆在IFLA的UNIMARC(1977)、美国的USMARC (1982)、英国的UKMARC(1969)三大格式基础上，各国均纷纷开发出本国的MARC格式

USMARC:

加拿大、西班牙、韩国、拉丁美洲、印尼、匈牙利、挪威.....

UKMARC:

澳大利亚、法国、芬兰、泰国、意大利、丹麦、新加坡、印度、瑞典.....

UNIMARC:

捷克、克罗地亚、日本、南非、中国、俄罗斯.....

MARC元数据

◆各种类型MARC的融合

1990年代之后，越来越多的国家开始放弃原有书目格式，转而使用USMARC以及后来的MARC21，如：澳大利亚(1991)、泰国(1992)、新加坡(1996)、南非(1997)、加拿大(1999)、瑞典(2000)、英国(2002)、印度(2003)、德国(2004)、奥地利(2004)、克罗地亚(2006)和芬兰(2008)

从1994年到2002年，经过多年的努力，英、美、加三国完成了从其国内格式向MARC21统一的进程，2002年，英国签署协议，完全使用MARC21格式。

MARC元数据

◆与其他数据格式的融合

2002年，LC推出MARCXML（适用于MARC21）；
丹麦国家图书馆开发MarcXchange（适用于各种MARC），2008年，成为国际标准ISO255577。

MARC元数据

CNMARC格式结构

- ◆记录头标：由24位字符构成。含有GB/T2901定义的关于记录结构的数据和为GB/T2901的特定形式而定义的几项数据元素。
- ◆地址目次区：含有一个或多个目次项，每个目次项描述一个字段。
- ◆数据字段区：由若干定长和变长字段构成，每个字段之间由字段分隔符隔开。
- ◆记录结束符：用以标识该记录结束和与下一个记录的区分。

MARC元数据

CNMARC主要内容

头标区：包含根据GB/T2901的规定所提供的对记录进行处理时所需的通用信息。

0—标识块：包含用以标识记录或标识在编文献的号码。

1—编码信息块：包含编码数据元素。

2—著录信息块：包含除“附注项”和“标准号与获得方式项”以外的其他ISBD规定的著录项目。

3—附注块：包含以自由行文方式对著录项目或检索点做进一步陈述的附注信息，可涉及文献或其内容的物理组成的各个方面。

MARC元数据

- 4—款目连接块：两种连接技术：嵌入字段技术、标准子字段技术。
- 5—相关题名块：为编目需要，以特定题名标识那些拥有多个题名的同一作品。
- 6—主题分析块：包含按照词语或符号的不同体系构成的主题数据。
- 7—知识责任块：包含需要建立检索点的知识责任。
- 8—国际使用块：包含国际上一致约定的不适合在0-7功能块处理的字段。

MARC元数据

必备字段：

记录头标 001记录标识号 100通用处理数据

200题名与责任说明 801记录来源

特殊文献类型必备字段：

文字类文献记录：101文献语种

测绘制图资料记录：120字段 123字段 206字段

电子资源记录：230字段 304字段

乐谱等类文献记录：125字段

拓片资料记录：191字段

MARC元数据

实体文献著录之一：一般资料标识

一般资料类型标识(General Material Designations, GMD)

1977年版ISBD(G)中明确指出：“一般资料标识的目的在于用概括性的术语在著录的开始说明文献所属的资料类别，GMD在题名后用方括号表示，属于可选项。”

MARC元数据

中国国家标准：

1983年中国文献标准化委员会制订出GB3469-83《文献类型与文献载体代码》的国家标准。

该标准在参考ISBD和AACR2相关规定的基础上，结合中国文献的实际情况，构造了文献类型和文献载体两套列表。基本内容如下表：

MARC元数据

文献类型			文献载体
1 专 著	10 产品样本	19 缩微胶卷	1 印 刷 品
2 报 纸	11 中 译 本	20 缩微平片	2 缩微制品
3 期 刊	12 手 稿	21 录 音 带	3 录音制品
4 会 议 录	13 参考工具	22 唱 片	4 录像制品
5 汇 编	14 检索工具	23 录 像 带	5 机读磁性制品
6 学位论文	15 档 案	24 电 影 片	6 其 它
7 科技报告	16 图 表	25 幻 灯 片	
8 技术标准	17 古 籍	26 其 它	
9 专利文献	18 乐 谱		

MARC元数据

2014年4月中国国家标准GB/T 3469-2013《信息资源的内容形式和媒体类型标识》颁布

增加了一个在国际标准中没有的限定词表“资料类别说明”。描述对象主要针对传统印刷品及其衍生品（如缩微品和电子图书）。但要求严格控制使用，只有当内容形式为“文本”或“图像”时，才能作为特殊限定词出现。

MARC元数据

内容形式	内容限定				媒体类型
数据集	测绘型	动态	二维	听觉	音频
图像	记谱型	静态	三维	味觉	电子
运动	表演型			嗅觉	缩微
音乐				触觉	显微
实物				视觉	投影
程序					立体
声音					视频
话语					多媒体
文本					其他媒体
多种内容形式					
其他内容形式					

MARC元数据

扩充内容限定词（资料类别说明）

会议文献	专利文献
学位论文	期刊
手稿	拓片
专著	标准文献
报纸	技术报告
古籍	

MARC元数据

ISBD统一版：

IFLA发布的 ISBD（《国际标准书目著录》）统一版中设置了第0项“内容形式和载体类型项”，三部分构成：

- ◆ 内容形式（Content Form）：表示资源内容的一种或多种基本形式。
- ◆ 内容限定（Content Qualification）：对内容形式的限定说明，即指出受编资源的类型、感觉性质、维数和/或活动的有无。
- ◆ 载体类型（Media Type）：用以承载资源内容的一种或多种载体的类型。

MARC元数据

内容形式	内容限定	载体类型
数据集	地图型 运动 2维 听觉	音频
图像	记谱型 静止 3维 味觉	电子
运动	表演型	缩微
音乐		显微
实物		投影
程序		立体
声音		多载体
言谈		其他载体
文字资料		
多内容形式		
其他内容形式		

MARC元数据

MARC21:

在2009年版的MARC21中新增336和337字段分别记录内容类型、媒体类型。

336 字段第1、第2 指示符均未定义。它由6 个子字段组成，分别是 \$a（内容类型术语）、\$b（内容类型代码）、\$2（来源）、\$3（指定材料）、\$6（连接）、\$8（字段连接及编号）。

337 字段第1、第2 指示符均未定义。同样由6 个子字段组成，分别是 \$a（媒体类型术语）、\$b（媒体类型代码）、\$2（来源）、\$3（指定材料）、\$6（连接）、\$8（字段连接及编号）。

MARC元数据

UNIMARC:

UNIMARC2008年正式发布的第三版以及其后的更新内容中都还没有出现过内容类型和媒体类型的字段。

UNIMARC委员会2012年6月发布的修订征集提议中，提到了两种实现方案。一是用203字段统一容纳内容和媒体类型，二是用281和282字段分别容纳。

MARC元数据

BIBFRAME:

形成一个由14个类型组成的词表：文字资料、地图、数据集、乐谱、动作谱、音频资料、图画、动画、三维物体、软件/多媒体、混合资料、资源集合、手稿和触觉资料

MARC元数据

实体文献著录之二：综合著录与分析著录

无总题名文献：

基本著录：

例1：

2001#@a 罗密欧与朱丽叶 @b 专著 @d Romeo and Juliet
@a 哈姆莱特 @d Hamlet @f (英)莎士比亚著 @g 朱生豪译
@z eng

例2：

2001#@a 千家诗 @b 专著 @f 谢枋得选编 @g 郭明志译注
@c 古诗源 @f 沈德潜编 @g 孟庆祥，孟繁红，孟繁翠译注

MARC元数据

分析著录：

基本著录——

从地球到月球[专著]；烽火岛 / (法)儒勒·凡尔纳著；叁壹
编译

分析著录——

烽火岛[专著] / (法)儒勒·凡尔纳著；叁壹编译。

MARC元数据

丛编与多卷书：

综合著录：

以共同题名为正题名，各单独部分的信息著录在内容附注。

分析著录：

编制分析款目，以部分题名著录作为正题名，共同题名著录在丛编项。

MARC元数据

实体文献著录之三：多载体配套文献

有主次之分：

著录主要部分，其余载体作为附件著录于载体形态项，或在附注项说明

无主次之分：

一般文献类型标识：多载体

载体形态项：可著录在一个215字段中，也可重复215字段分别著录不同载体。

MARC元数据

MARC元数据的未来

关联数据（Linked Data）：

Tim Berners-Lee在2006年提出关联数据的概念。其目的是构建一个计算机能理解的、具有结构化语义的数据网络，这一网络同时也是人类可读的。

定义了一种URI规范，使人们可以通过HTTP/URI机制，直接获得数字资源。

定义了一种Web上的富链接机制，将超文本链接变为数据链接。

MARC元数据

关联数据四原则：

- (1)使用URI作为任何事物的标识名称,不仅是标识文档;
- (2)使用HTTP URI,使任何人都可以参引(dereference)这一全局唯一的名称;
- (3)当有人访问名称时,以RDF形式提供有用的信息;
- (4)尽可能提供链接,指向其他的URI,以使人们发现更多的相关信息。

MARC元数据

关联数据的发展：

- ◆ 2006年，W3C倡导关联开放数据项目（Linked Open Data Project），多领域的参与者将数据发布为关联数据并相互关联，形成关联数据云（LOD Cloud）。
- ◆ 2011年9月，开放知识基金会网站的统计数据：图书馆关联数据项目达到51个，其中有20个书目数据，13个规范数据，6个术语服务。
- ◆ 2013年12月，Datahub网站的统计数据：图书馆关联数据项目达到115个。

MARC元数据

关联数据应用于图书馆的作用：

- ◆ 加强对知识点的挖掘及知识关联的揭示
- ◆ 拓展资源发现的范围和能力
- ◆ 实现元数据的共享与重用
- ◆ 采用RDF等网络通用的编码格式，加强元数据的兼容性与互操作性
- ◆ 赋予元数据机器可理解的能力，推动元数据与语义网的融合

DC元数据

概述

DC元数据（都柏林核心元数据，Dublin Core Metadata）
1995年3月，由OCLC与国家超级计算应用中心（NCSA）
联合发起，因创始地在美国俄亥俄府都柏林而得名。目
的是建立一套描述数字资源的标准。由DCMI负责维护

DC元数据

描述对象：

“资源（Resource）”。资源是任何可以标识的东西。可以是实体的，也可以是抽象的。常见的例子有电子文档，图像，服务（例如，“洛杉矶今天的天气预报”），还有其他资源的集合。并非所有的资源都是网上可检索的；例如，人，机构，还有图书馆里装订成册的书都可以被认为是资源。

（引自《关于元数据的十万个为什么》刘炜）

DC元数据

发展历史

- ◆ DC-1（1995）：DC元数据产生。确定DC元数据的描述对象、13个元素，确立各种基本原则。
- ◆ DC-2（1996）：讨论DC元数据应用的语法和结构问题，提出Warwick框架作为元数据应用的一般“容器”和概念框架。
- ◆ DC-3（1996）：提出DC元数据在互联网图像资源描述方面的应用方案。增加了描述Description、权限Rights两个元素。

DC元数据

- ◆ DC-4（1997）：讨论DC元数据应用的扩展问题。提出三类“堪培拉限定词”，即对DC元数据元素可以从取值的language、Scheme和Type三个方面进行限定或扩展。提出了DC限定版。
- ◆ DC-5（1997）：“荷兰终结”，指对DC的非限定版DCMES的最终确立。讨论了一对一原则。
- ◆ DC-6（1998）：形成了DCMI的基本组织形式和运行模式。

DC元数据

- ◆ DC-8（2000）：提出应用纲要形式解决元数据的领域应用问题。
- ◆ DC2003：提出抽象模型，对各类应用纲要进行规范。
- ◆ DC2007：提出了被称为“新加坡框架”的元数据应用规范。
- ◆ 至今，DC已发展成为涵盖一系列文档的标准规范体系，其中，除了以DC元数据术语集为主的语义规范外，还发展起包括词表、编码规范、模型等文档的一整套辅助规范。

DC元数据

特点

◆ 简单性原则

定义一个能得到最广泛应用、被全球所理解和接受的最小元素集。

◆ 内在本质原则(Intrinsicality)

集中于描述对象内在的属性。如：知识内容、物理形式。信息来源于描述对象，又不超出描述对象。

◆ 可扩展原则(Extensibility)

为描述独特的资源种类，因此必须有适当的弹性。可以基于DC核心元素集通过各种方式扩展为适应各领域资源描述需要的元数据方案。

DC元数据

◆ 语法独立原则(Syntax-Independence)

只规定元素的基本语义，元素可以以多种方式编码，避免语义的捆绑。

◆ 可选择性原则(Optionality)

所有的元素都是可选择的。

◆ 可重复原则(Repeatability)

所有元素均可重复。在此原则下，不区分作者的排名。

◆ 可修饰原则(Modifiability)

元素可独立使用，也可利用修饰词进行限定。限定不能扩大或改变元素的基本语义。

DC元数据

元素

题名Title、创建者Creator、日期Date、主题Subject、出版者Publisher、类型Type、其他责任者Contributor、格式Format、来源Source、标识符Identifier、语种Language、关联Relation、覆盖范围Coverage
描述Description、权限Rights

2012年发布的术语集（DCMES）共包括：55个术语（含15个元素）、9个词表编码体系、12个句法编码体系、22个类

DC元数据

标签: 题名(Title)

定义: 赋予资源的名称。

标签: 创建者(Creator)

定义: 创建资源内容的主要责任者。

标签: 日期(Date)

定义: 与资源生命周期中的一个事件相关的时间。

标签: 主题和关键词(Subject and Keywords)

定义: 资源内容的主题描述。

标签: 出版者(Publisher)

定义: 使资源成为可获得的负责实体。

DC元数据

标签：资源类型(Resource Type)

定义：资源内容的特征或类型。

标签：其他责任者(Contributor)

定义：对资源的内容做出贡献的其他实体。

标签：格式(Format)

定义：资源的物理或数字表现形式。

标签：来源(Source)

定义：对当前资源来源的参照。

标签：资源标识符(Resource Identifier)

定义：在特定的范围内给予资源的一个明确的标识。

DC元数据

标签: 语种(Language)

定义: 描述资源知识内容的语种。

标签: 关联(Relation)

定义: 对相关资源的参照。

标签: 覆盖范围(Coverage)

定义: 资源内容所涉及的外延或范围。

标签: 描述(Description)

定义: 资源内容的解释。

标签: 权限管理(Rights Management)

定义: 有关资源本身所有的或被赋予的权限信息。

DC元数据

限定 (Refinements)

是对元数据语义的进一步限定和细化，也叫修饰词。包括元素修饰词、编码体系修饰词。

元素修饰词：对元素的语义进行修饰，提高元素的专指性和精确性。

编码体系修饰词：用来帮助解析某个术语值的上下文信息或解析规则。其形式包括受控词表、规范表或解析规则。起到对网络资源进行规范控制的作用。

DC元数据

DC元数据应用纲要 (DCAP)

纲要描述一个标准或一些规范是如何被运用以支持特定的应用、功能、行业需求或特定环境的文档。元数据纲要是一种元数据标准的应用形式，也可以看成是一种规范的元数据方案。

元数据纲要的作用：一、保证了元素基本集的稳定，二、实现了特定应用领域对元数据标准的需求，三、对DC元数据标准的应用进行了一定程度的规范。

DC元数据

2001年《DC图书馆应用纲要》（草案）发布

元素：18个

元素限定：26个

编码体系修饰词：21个

术语属性包括：

术语名称；术语URI；标签；定义来源；初始定义；

DC-Lib定义；初始注释；**DC-Lib注释**；术语类型；限定；被限定；编码体系应用于；可应用的编码体系；约束；出现次数

DC元数据

元素	元素修饰词
题名	交替题名
创建者	
其他责任者	
出版者	
主题	
描述	摘要、目录
日期	创建日期、生效日期、可获得日期、发布日期、修改日期、版权日期、递交日期、接受日期、获取日期
类型	
格式	大小、媒体

DC元数据

元素	元素修饰词
标识符	文献引用
来源	
语种	
关联	版本继承、格式转换于、格式转换为、被替代、替代、部分于、部分为、需求、被参照、参照
覆盖范围	空间、时间
权限	
适用对象	
版本	
馆藏位置	

DC元数据

编码体系修饰词：

Box	ISO8601	SICI
DCMIType	ISSN	TGN
DDC	LCC	UDC
DOI	LCSH	URI
IMT	MESH	W ₃ CDTF
ISBN	Period	
ISO3166	Point	
ISO639-2	RFC3066	

DC元数据

术语名称	题名 (title)
DC-Lib注释	并列题名或音译题名被认为是主要的题名，也就是说，题名元素可重复。 无论题名还是标识符都是必备的。 保留题名起始的冠词并以基于语言的本地算法排序。
术语名称	创建者 (creator)
DC-Lib定义	对资源的知识或艺术内容起主要作用的实体。
DC-Lib注释	不建议以责任方式作为对这个术语的限定。 DCMI不赞成结构化值提供关于创建者的更多信息——它应该属于对另一资源的描述。
术语名称	其他责任者 (contributor)
DC-Lib注释	取自美国国会图书馆Relators列表中的术语集已经被核准为其他责任者的责任方式限定。

DC元数据

术语名称	出版者 (publisher)
DC-Lib注释	取自美国国会图书馆Relators列表中的术语集已经被核准为出版者的责任方式限定。
术语名称	主题 (subject)
DC-Lib注释	强烈推荐在元数据中通过自由文本或受控词表提供主题内容。同时建议受控主题词与特定的编码系统一起使用，如果没有指定编码系统则当作关键词处理。 如果使用限定版DC，默认使用受控词表的编码体系。 如果地理或时间方面的内容已使用覆盖范围元素来表示，若有需要仍可在主题元素中重复。
术语名称	描述 (description)
DC-Lib注释	允许通过URI链接到一个或多个外部的资源描述，但为了便于对描述的内容进行关键词索引，建议仍然包含一个文本描述。

DC元数据

术语名称	日期 (date)
DC-Lib注释	为明确日期的含义，推荐使用元素限定词。 建议使用如下日期编码形式：W ₃ C-DTF、ISO8601（最佳选择） 可以使用诸如“日一月一年”的公认形式，这里日和年用数字表示， 月份使用文字名称或标准的缩写，避免使用诸如“DD/MM/YY” 或“MM/DD/YY”这类可能引起混淆的日期表达。
术语名称	类型 (type)
DC-Lib注释	使用受控的列表并以编码体系标识出值的来源。 推荐从DCMI类型表中选择至少一个值来揭示高层次上的资源类型。
术语名称	格式 (format)
DC-Lib注释	此元素起初用于Internet媒体类型（IMT）。推荐用于电子资源。

DC元数据

术语名称	标识符 (identifier)
DC-Lib注释	<p>关于标准的标识符：如果分配给资源有一个或多个标准的标识符且能被元数据代理所知，至少应从正式编码体系中选择提供一个正式的标识符。</p> <p>在更抽象的层面中使用该元素，图书馆内部用的馆藏标识，如索取号，应许放入DC-Lib的“馆藏位置”元素。</p>
术语名称	来源 (source)
DC-Lib注释	<p>如果缺少资源全球性的唯一标识，可提供一个本地属性的自构或衍生ID，包括给出ID生成机构采用的规范。</p> <p>该元素只用于被描述资源是由非数字化形式转为数字化形式，其他情况下用关联元素。</p>
术语名称	语种 (language)
DC-Lib注释	<p>推荐使用代码而不是文本，建议使用ISO639-2书目编码。本元素有则必备，意指只要有任何口述或书写的内容，都应提供语种。</p>

DC元数据

术语名称	关联 (relation)
DC-Lib注释	如果使用修饰词，则用其适用当中最专指的一个。 建议在下列情况下使用修饰词： 1. 所描述文献是宿主文献的组成部分且在DC标识符元素下没有给出引用信息时； 2. 所描述文献是已出版文献的修订版或重定格式后的发行物，且这方面信息的获得比较容易。
术语名称	覆盖范围 (coverage)
DC-Lib注释	推荐使用带有修饰词的覆盖范围。
术语名称	权限 (rights)
DC-Lib注释	有必要决定在图书馆应用中如何使用这一元素。在不同的应用中，现在就有很多关于权限的元数据正在讨论中。

DC元数据

术语名称	适用对象 (audience)
DC-Lib注释	需要评估哪些限定或编码体系对DC-Lib是适用的。MARC的目标用户代码有可能纳入考虑范围内。
术语名称	版本 (edition)
DC-Lib定义	说明一件作品译本或版本的信息。
DC-Lib注释	指明特定作品的译本或版本信息是至关重要的，这一点对那些经常变化的资源来说尤为重要。此元素不同于反映同一资源不同物理格式的情况。 此元素一般不可重复。
术语名称	馆藏位置 (location)
DC-Lib定义	用来标识拥有资源的机构或者获得资源的途径。
DC-Lib注释	当URI不适用时，此元素提供用户获取资源的物理馆藏位置。 也可以在馆藏位置元素中包含进一步的标识信息，如：索书号、登录号等。

DC元数据

DC元数据抽象模型

(DCMI Abstract Model, DCAM)

在一套概念术语的基础上，提供一个抽象的数据模型。定义了DC元数据描述的各类实体对象及其相互之间的关系，描述DC元数据所使用的描述资源的信息结构。提供了一种独立于任何特定编码方式的信息模型，实现在不同元数据方案之间的共享和重用。

- 2005年，DC元数据抽象模型成为DCMI的推荐规范
- 2007年，发布修订版本。

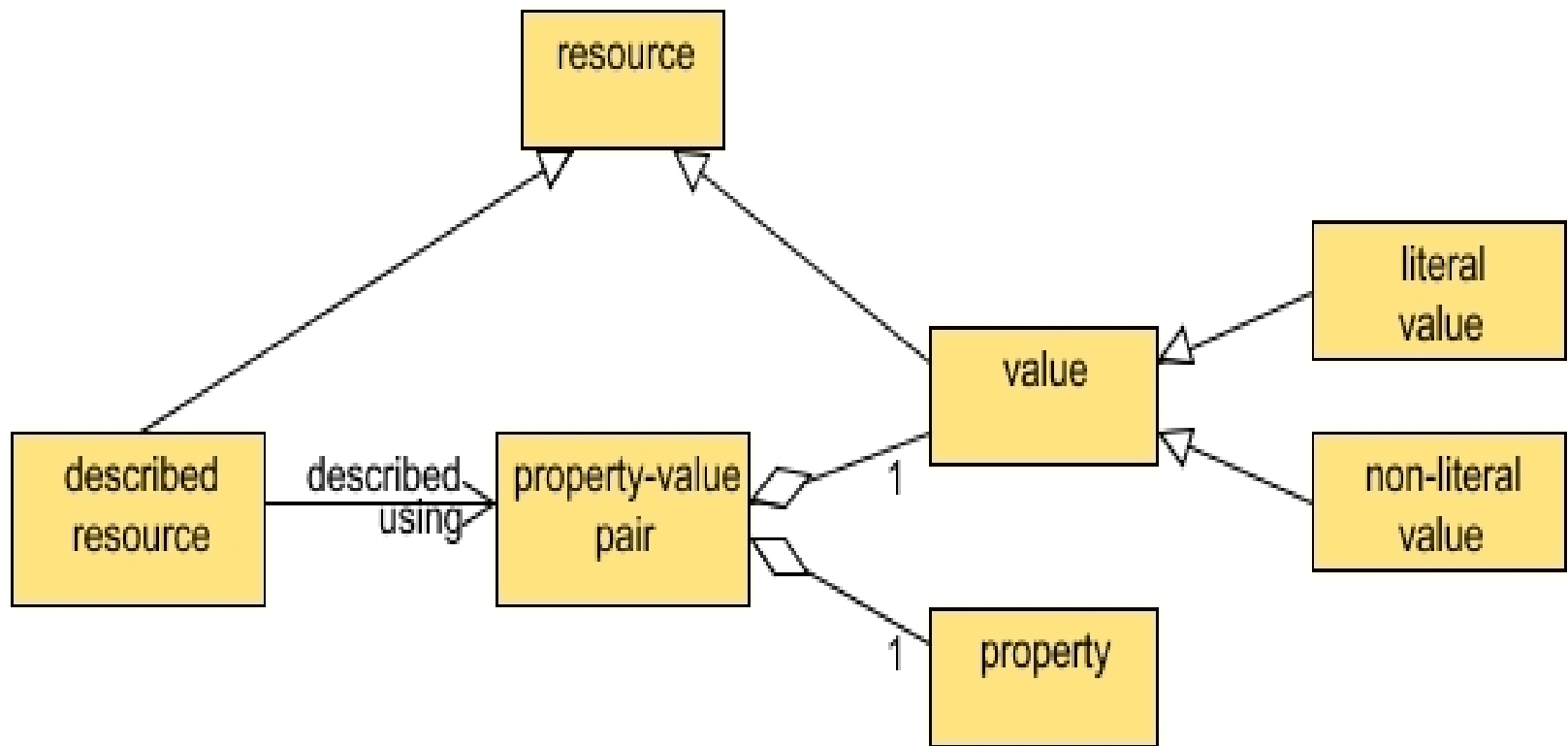


Figure 1 - the DCMI resource model

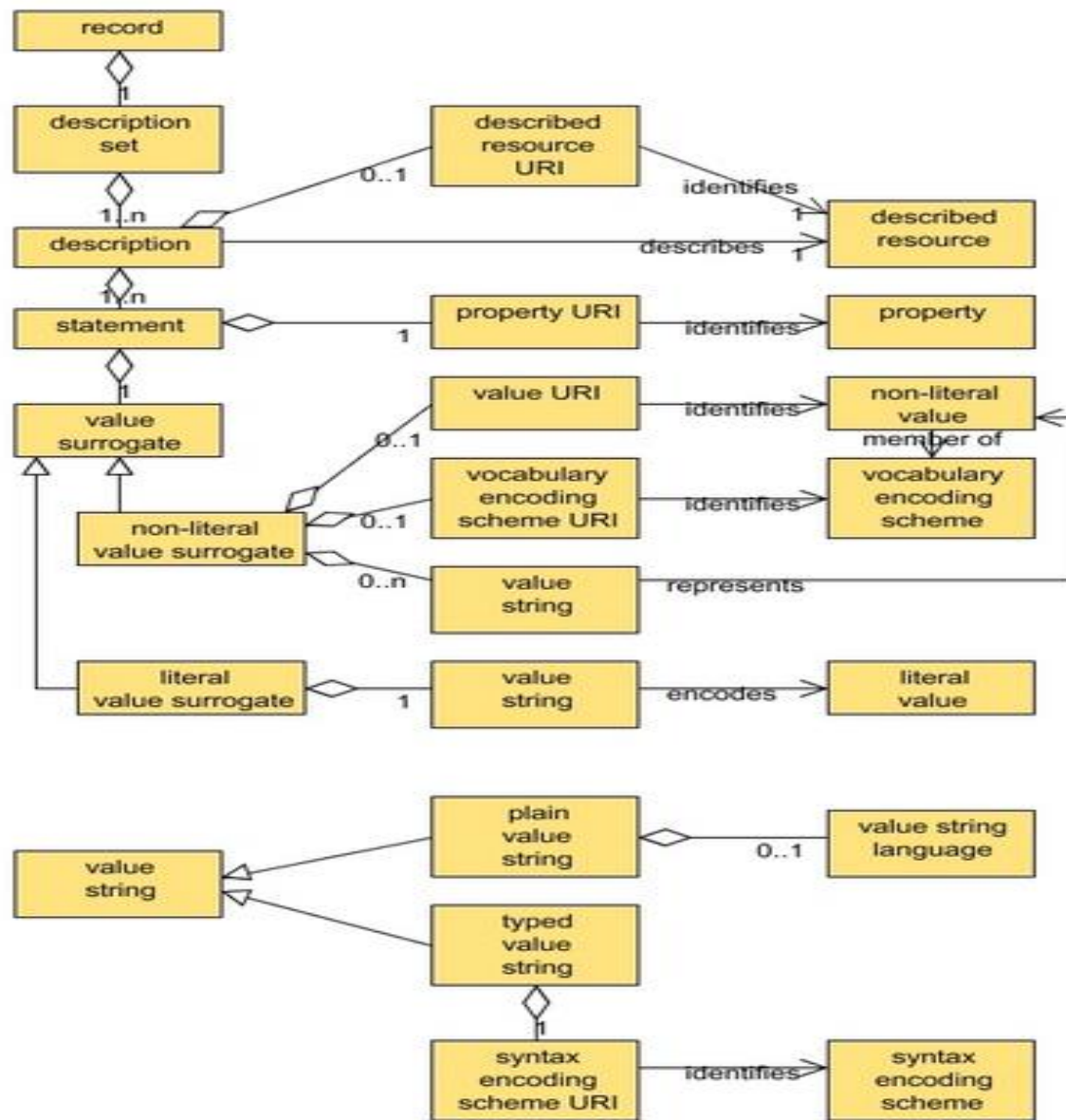


Figure 2 - the DCMI description set model

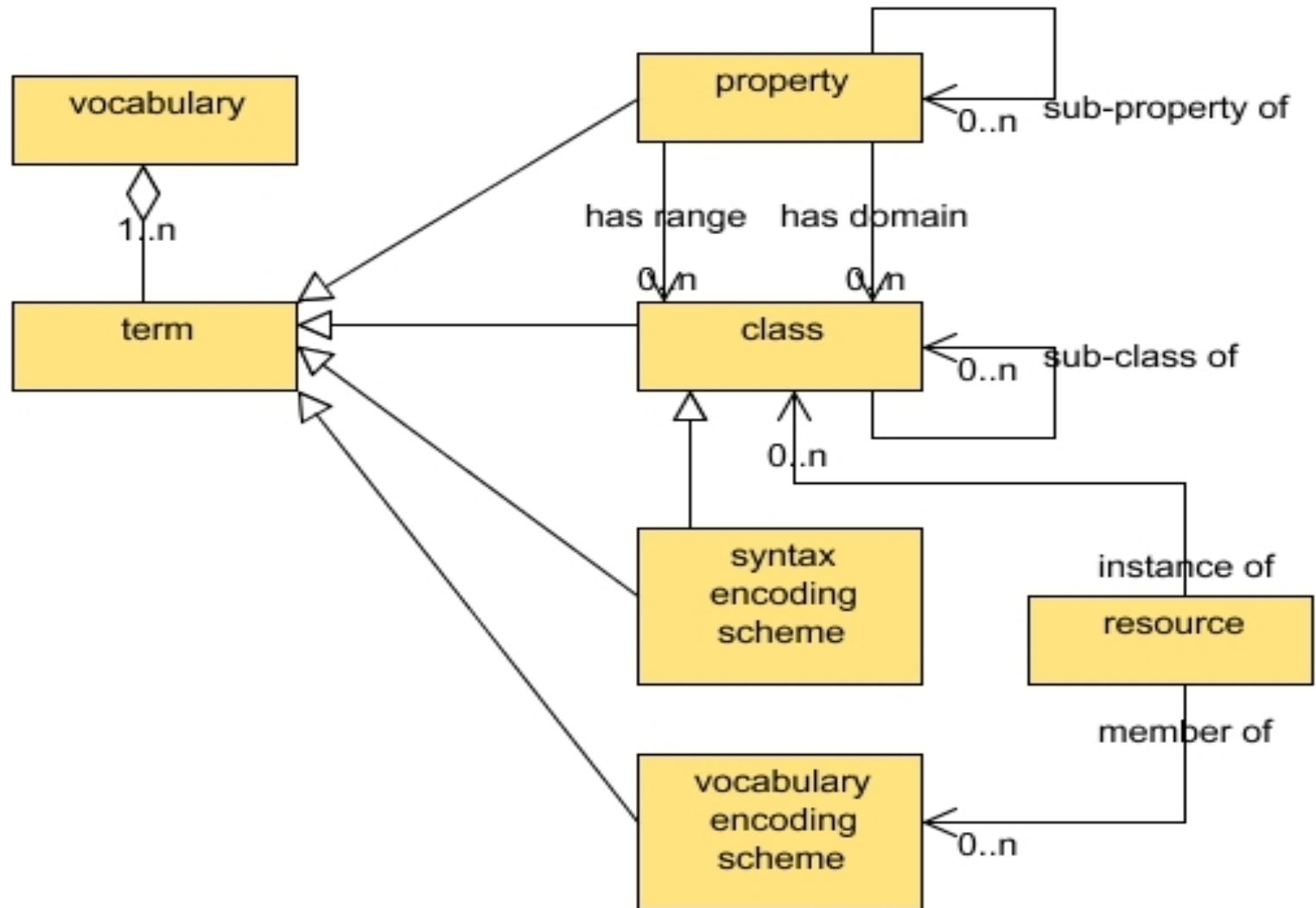


Figure 3 - the DCMI vocabulary model

MODS元数据

概述

MODS（Metadata Object Description Schema，元数据对象描述模式）

美国国会图书馆于2002年6月开发，是继MARC之后的第二种以MARC为基础的文献编目元数据。MODS的元素来自MARC21的字段，由元素、子元素、属性三层结构组成，采用XML作为编码语言。

优点：结合MARC与DC的优点；方便与MARC交换；比MARC更简化；采用语义化标签；采用通用的数据格式。

MODS元数据

元素

20个顶级元素：Title Info（题名信息）、Note（附注）、Name（名称）、Subject（主题）、Type of Resource（资源类型）、Classification（分类）、Genre（体裁形式）、Related Item（相关文献）、Origin Info（来源信息）、Identifier（标识符）、Language（语种）、Location（馆藏位置）、Physical Description（载体形态描述）、Access Condition（检索环境）、Abstract（摘要）、Part（部分）、Table of Contents（目次）、Extension（扩展）、Target Audience（读者对象）、Record Info（记录信息）

MODS元数据

著录规则

- ◆ “题名信息”及“题名”元素是每个MODS记录必备的。
- ◆ 元素及子元素都是可重复的，属性均不可重复。
- ◆ 元素及属性间的次序是不固定的。
- ◆ 某些主元素可以作为子元素用于不同的主元素下，如：Name、Identifier、TitleInfo等。

元数据互操作

定义：元数据互操作的基础是通过元数据标准的互操作，解决元数据标准间的差异与冲突。这一过程是指建立协议规范、语义映射，并在此基础上整合多种元数据标准。

主要实现方法：元数据映射、建立资源描述框架、元数据交换协议等。其中，元数据映射(Metadata Mapping/Metadata Crosswalks),又称元数据转换,是从一种元数据格式的元素、语义和语法到另一种元数据格式元素、语义和语法的映射,是实现元数据互操作较为常用的一种方法。

元数据互操作

CNMARC与MARC21映射问题

一、CNMARC与MARC21:

由于元素取值所依据的标准不一致导致的语义差异。如：在主题标引时，使用的主题词表有中国分类主题词表、美国国会图书馆主题词表LCSH、医学主题词表MESH等，在分类标引时，使用的分类法有中国图书馆分类法、杜威十进制分类法DDC、国会图书馆分类法LCC、国际十进制分类法UDC等。

元数据互操作

二、CNMARC格式之间：

国内对中文文献进行编目多数使用的是CNMARC格式，但是在著录细则上存在着差异，势必导致元数据语义的差异。如：国家图书馆与CALIS两家分别作为公共图书馆与大学图书馆的代表，在编目实践与规范控制等方面存在着不少差异，数年来已有多人撰文论述。这一问题为国内通用元数据映射规则的推行带来困难。

元数据互操作

三、CNMARC的本地化：

如：CNMARC在使用中，基于汉字的特点，增加了一些非通用字段，在映射中无法实现一对一的映射。如：用来记录汉字拼音的字段，用来描述字库中缺失的古字、生僻字等的字段。这些字段与MARC21中的字段没有对应关系，但是由于其体现了国内的文化特点，在元数据的描述及检索中都具有实际的价值，因此，在制定映射规则时应考虑保留其功能。

元数据互操作

MARC与DC映射问题

一、结构匹配差异

◆一对多或多对一

如：DC的主题元素(subject)对应MARC的多个用于标引的字段，包括主题标引、关键词标引、分类标引等。

◆无对应关系

如：DC的日期元素(date)与MARC描述日期的相关字段存在着语义上的差异，无法完全匹配。

元数据互操作

二、应用匹配差异

- ◆ 必备元素与可选元素的匹配差异。
- ◆ 可重复元素与不可重复元素的匹配差异。
- ◆ 元素层次关系的匹配差异。

元数据互操作

三、语义匹配差异

◆ 语义内涵的不一致

如：CNMARC中为实现描述与检索两个功能，以多个字段著录同一内容，一部分字段承担描述功能，一部分字段承担检索与规范的功能，例如题名与统一题名，用于著录责任者的2字段与7字段。而在DC中没有承担规范功能的元素。

◆ 元素取值所依据的标准不一致

DC对知识组织工具进行的映射不再局限于业界所常用的词表，还要涉及到不同领域、不同行业使用的成熟词表或分类体系。

◆ 自由词与规范词转换带来的不一致

元数据应用

- 一、国家图书馆特色资源库建设
- 二、国家图书馆元数据仓储建设及应用

特色资源库建设

资源概况

➤ 资源内容

公共馆：

实体文献：方志、家谱、古籍、地方文献.....

虚拟实体：人物、历史事件、景观、非遗.....

国图：

实体文献：古籍、碑帖、甲骨、年画、家谱、方志.....

虚拟实体：知识点

特色资源库建设

➤ 媒体类型

公共馆：文本、图像、音视频.....

国图：文本、图像、音视频.....

➤ 元数据类型

公共馆：MARC、DC、自定义、无

国图：MARC、DC、自定义

➤ 服务方式

公共馆：文字介绍、电子书浏览、检索与分类浏览

国图：检索、浏览、电子书阅览、知识服务.....

元数据在馆藏特色资源建设中应用原则

特色资源库建设





貴州志稿

责任者 潘文丙

出版地

出版时间 清[1644-1911]

版本 抄本

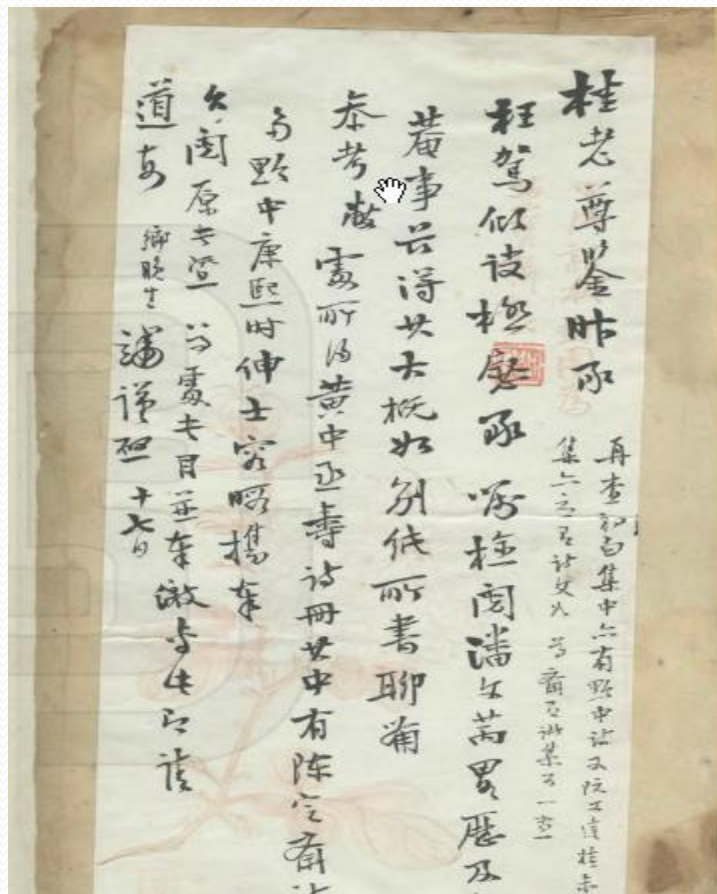
分类号

文种 chi

总册数 1册

描述

索取号 地360\34.04.1



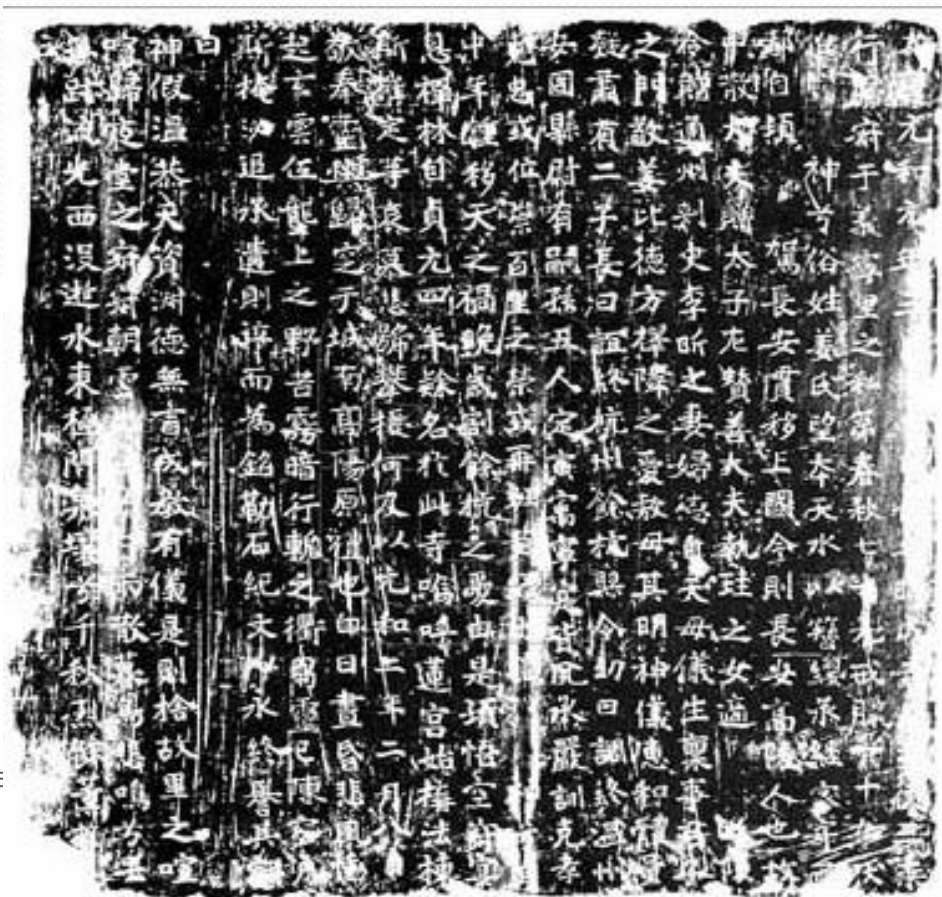


李昕妻姜氏（昭成寺尼）墓誌li xin qi jiang shi mu zhi



[查看大图](#)

责任者	[佚名]撰
年代	唐元和2年（807）2月8日
地点	陕西省西安市出土
拓片原物状况	端方藏石
附刻	
拓片版本	
拓片描述	
书体与行款	正书21行行21字
装裱获得方式	
数目与尺寸	1张36x37cm
题跋印记	鈐：“曾在叢氏耕石齋”“耕石
文种	汉文
书目索引	
关联关系	
拓片录文	
馆藏信息	赧本1174



甲骨世界



[查看大图](#)

北圖2376 bei tu 2376

貞人名字

地點 河南省安陽市殷墟出土

時代 商武丁 (B.C.1250-1192)

背面痕跡

來源方式 購買

尺寸 4.4×2.1cm

數量與尺寸 1塊；面5字，背3字

原物介質 獸骨

書體與形款 篆書

綴合信息

內容主題 天象

地點主題 中國

釋文 (面) 今日不雨？ □ 𠄎；53 gt； (背) 帚豐來。

參考信息 合2798 羅300

對應拓片 [甲骨2376](#)



年画撷英

号响三军令 鼓振万人惊



查看大图

制作时间 [民国年间]

规格 2图片

中图分类 J228.3

索取号 H07382

版本说明 木刻本

责任者

附注 此为喜庆吉祥年画。《号响三军令》画的是状，脚下有一锭元宝，身后是一盆荷花。小子双手各执一根鼓槌正在打架子鼓，身后小子均身着花袄，神情喜庆，一派天真无邪的“万人惊”的字样，但由大胖小子出场担当祥和平安，又令人振奋。作者的笔触虚实有开脸手法娴熟，夸张合理，制作水平还是比



前尘旧影



查看大图





姓氏详情

名稱：何

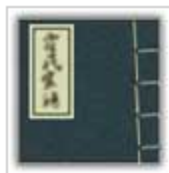
拼音：Hé

郡望地：廬江、東海、陳郡、丹陽、齊郡

分布地：[山東平邑](#) [山西太原](#) [北京](#) [河北尚義](#) [貴州從江](#) [雲南瀘](#)

[東龍口](#) [雲南路南](#) [廣東澄海](#)

分布民族：漢族 回族 土家族
傈僳族 彝族 布依



曹氏宗譜

[暂无家谱全文](#)

題 名：曹氏宗譜

版本年：光緒二年（1876）

譜籍地：江蘇

始 祖：（漢）曹確

收藏地：美國

題 要：寧壽堂刊本

[我要收藏](#) [我要反饋](#)

卷 數：八卷

版 本：活字本

堂 號：寧壽堂

始遷祖：

修撰者：曹盛茂等

相关堂号

历代名

漢
明
清
明
梁
南朝宋
齊
魏
晉
漢

何武

家谱评论

家族名人

相关家谱

相关方志

[江南通志](#)

江苏

清乾隆元年[1736]

[江蘇省通志稿](#)

江苏

1984

[宣統江蘇通志](#)

江苏

清宣統間[1909-1911]

最大家谱深度

最大可视数目

显示历史

自动调整间距

家谱间距

鄭和家譜考釋

曲沃莊簡公李
氏家譜

會稽馬氏分支
宗譜

雲陽包港王氏
族譜八卷

會稽旗收嶺董
氏族譜

仇氏家乘

雲陽涂氏族譜

容氏譜牒

▼ 韓武子

▼ 簡

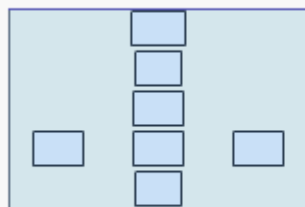
▼ 子與

總子

▼ 宣子

固

頂



第一十代

第一十代

廖興正

第十一代

第十一代

廖忠榮

廖忠華

第十二代

第十二代

廖義認

廖義詎

廖義太

廖義詔

廖義誥

廖義論

第十三代

第十三代

廖安京

廖安閻

廖安周

廖安立

廖安用

第十四代

第十四代

廖邦鎮

廖邦錄

廖邦鏡

廖邦乾

廖邦坎

廖邦艮

廖邦震

廖邦巽

廖邦離

廖邦坤

廖邦兌

廖邦富

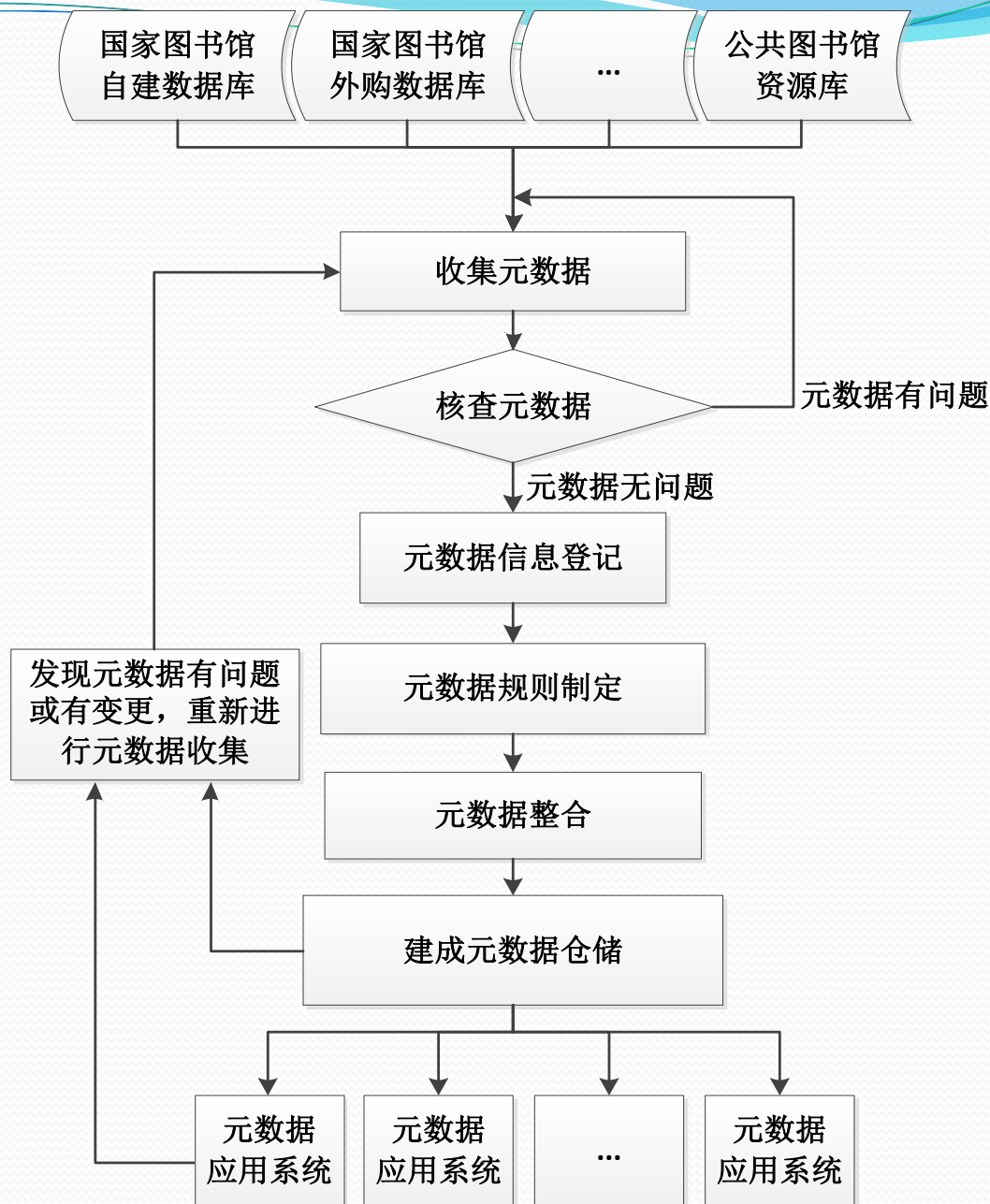
元数据仓储建设及应用

元数据仓储的建设

意义:

- ◆ 实现对不同格式元数据的一致性存储，能够对元数据进行集中管理和利用。
- ◆ 元数据经过转换，格式相对统一，并可以按照整合者的需求重新建立各种分类体系。这种整合机制有利于实施数据挖掘、知识发现等服务。
- ◆ 为用户提供一个单一、透明的数据使用环境，简化了用户访问信息的复杂度，提高了元数据的访问速度和整合系统的性能。

元数据仓储建设流程图



数据情况

元数据仓储目前已经整合的数据包括国家数字图书馆自建资源、外购资源合计**100多个**资源库，**超过2亿条**元数据。

中文资源方面：

自建资源：有中文图书、博士论文、民国文献、地方志、古籍文献、各类资源库等；

外购资源：包括清华同方、维普、万方、方正阿帕比、龙源期刊网、北大法意、国研网等资源服务商提供的资源库。



中文资源

数据情况

元数据仓储目前已经整合的数据包括国家数字图书馆自建资源、外购资源合计**100多个**资源库，**超过2亿条**元数据。

外文资源方面：目前已整合到文津系统14个数据库的元数据。

EARLY AMERICAN IMPRINTS

EEBO
EARLY ENGLISH BOOKS
ONLINE

The MAKING of the
MODERN WORLD

OXFORD JOURNALS

Eighteenth Century
Collections Online

The MAKING of
MODERN LAW
Legal Treatises 1800 - 1926



元数据仓储建设及应用

元数据仓储的应用

现有搜索引擎：

OPAC检索

针对实体资源的检索，检索内容范围小。

联邦检索

Metalib，可实现跨库检索，但响应速度和用户体验还不太理想。

新型检索系统

Primo, Summon，可实现一站式检索，用户体验较好，对中文数据库的兼容性还需改进。

元数据仓储建设及应用

基于元数据仓储的搜索服务——文津搜索：

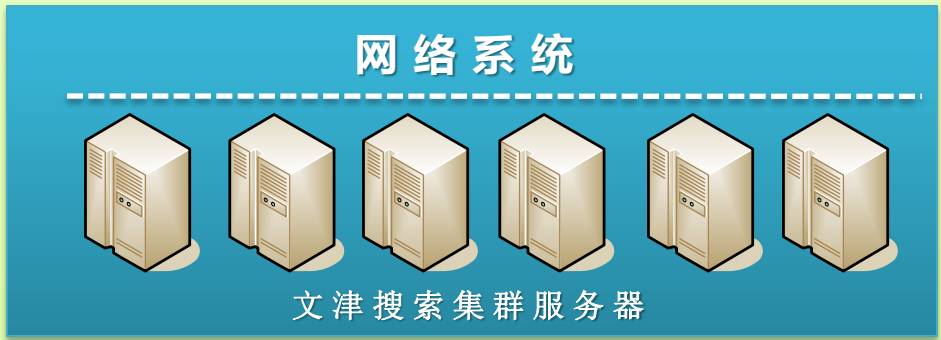
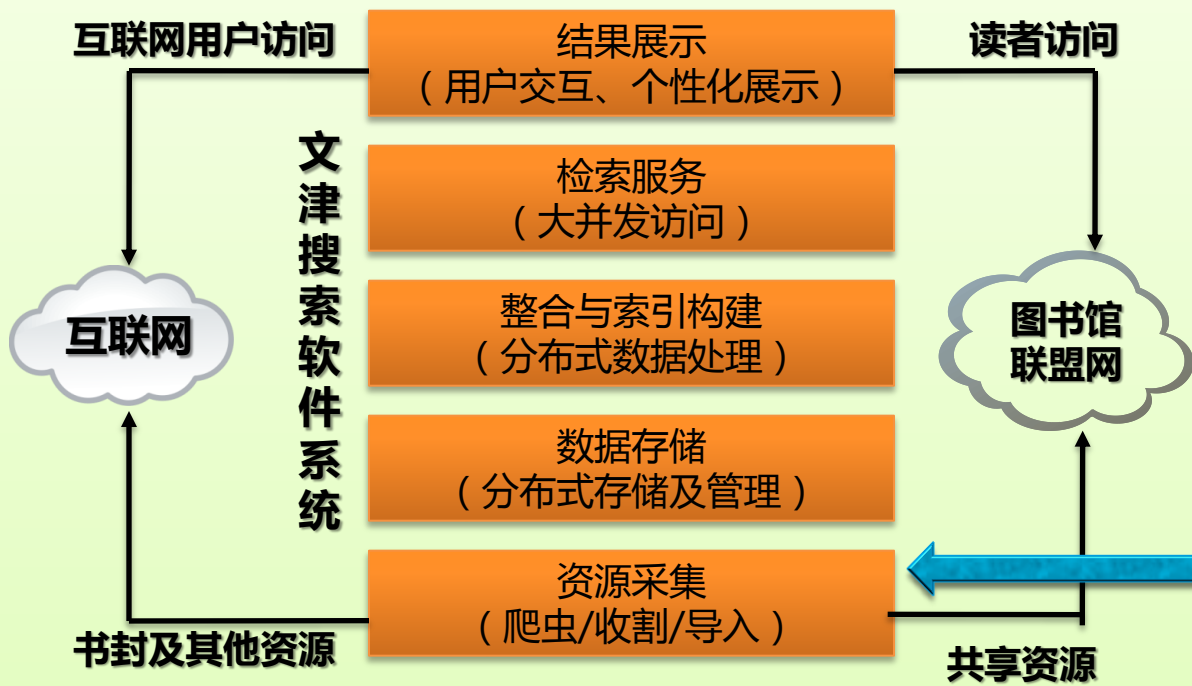
- ◆ 搜索引擎，资源检索、揭示和获取的服务平台；
- ◆ 资源整合和服务整合的平台；
- ◆ 国家数字图书馆工程中的一个核心子项目。
- ◆ 可以实现资源的统一检索、统一揭示和统一获取。
- ◆ 网址：<http://find.nlc.gov.cn/>

元数据仓储建设及应用

文津搜索系统建设目标：

整合国家数字图书馆自建或其它方式获取的数字资源，建立和维护高性能分布式索引，并结合基于互联网的信息收集和数据分析挖掘技术，向用户提供一个统一的实时高效、精准权威的数字图书馆数字资源元数据搜索服务平台。

文津搜索系统总体架构





文津搜索系统运行平台：

包括40个机柜及机柜间综合布线

400台服务器

(前端及缓存 54台，核心库检索 152台，扩展库检索 68台，索引构建 22台，数据存储和处理 80台，爬虫 24台)

4台中高端Unix小型机，2台高端Unix小型机

网络设备及网络安全设备

对数百台设备运行状态进行监控和运维支撑系统

1、资源一站式检索

[国图首页](#) [个人中心](#) [登录](#) [注册](#) [意见反馈](#) [帮助](#)



资源尽享
文化津梁

支持一站式的检索

图书、论文、期刊报纸信息一步获取

Q 搜索

高级检索
联邦检索

[全部](#) | [图书](#) | [古文献](#) | [论文](#) | [期刊报纸](#) | [多媒体](#) | [缩微文献](#) | [文档](#) | [词条](#)

 搜索热词榜

文津搜索系统检索内核支持拼音检索、检索提示、二次检索、同义词检索。

2、资源分类检索

国图首页 个人中心 登录 注册 意见反馈 帮助



文化津梁
资源尽享

全部字段 ▾

Q 搜索

高级检索
联邦检索

全部 | 图书 | 古文献 | 论文 | 期刊报纸 | 多媒体 | 缩微文献 | 文档 | 词条

🔥 搜索热词榜

文津搜索系统把所有资源分为了图书、古文献、论文、期刊报纸、多媒体、缩微文献、文档、词条8个一级分类，支持针对不同分类，提供专业检索字段，快速准确查找资源。

3、高级检索



资源尽享
文化津梁

图书

高级检索

1

选择

选择高级检索方式

>

3

选择高级检索范围

>

4

检索

选择文献类型

选择文献类型

指定文献类型：

选择高级检索方式

检索类型： 逻辑检索 在全文中检索

点击添加检索条件后，下方增加布尔逻辑命令和新增检索语句，最多增加3条逻辑检索项。

选择高级检索范围

出版年份： 年 - 年

只显示： 可提供全文

支持出版时间范围、提供全文的检索功能

检索

4、提供多维度的导航过滤

天津 数字图书馆


全部 图书 | 古文献 | 论文 | 期刊报纸 | 多媒体 | 缩微文献 | 文档 | 词条


高级检索 联邦检索


Q 搜索


全部检索字段 包含检索词的结果 排序: 相关性 1/5500 下一页

找到相关结果约 55000 个, 当前显示: 1-10

1.  **数字图书馆**
文献类型: 专著
著者: 王大可
出版年份: 2002 出版社: 天津人民出版社
来源数据库: 馆藏中文资源
[摘要](#) [目次](#) [馆藏信息](#) [查看其他版本或分册](#)

2.  **数字图书馆**
文献类型: 专著
著者: 高文
出版年份: 2000 出版社: 清华大学出版社
来源数据库: 馆藏中文资源
[摘要](#) [目次](#) [馆藏信息](#)

3.  **数字图书馆**
文献类型: 专著
著者: 吴志荣
出版年份: 2000 出版社: 学林出版社
来源数据库: 馆藏中文资源
[摘要](#) [目次](#) [馆藏信息](#)

4.  **数字图书馆**
文献类型: 专著
著者: 清华大学图书馆

[查看其他版本或分册](#)

查看指定类型 命中数

- 图书
 - 专著
- 缩微文献
 - 缩微文献
- 文档
 - 报告
 - 法律文书/...
- 论文
 - 期刊论文
 - 学位论文
 - 会议论文
- 词条
 - 年鉴词条
 - 工具书词条
- 多媒体
 - 计算机文件
 - 视频
- 期刊报纸
 - 报纸
 - 期刊

缩小检索范围

4、提供多维度的导航过滤

查看指定类型 命中数

- 图书
 - 专著 (362)
- 缩微文献
 - 缩微文献 (13)
- 文档
 - 报告 (588)
 - 法律文献/... (16)
- 论文
 - 期刊论文 (61858)
 - 学位论文 (1194)
 - 会议论文 (795)
- 词条
 - 年鉴词条 (474)
 - 工具书词条 (12)
- 多媒体
 - 计算机文件 (27)
 - 视频 (20)
- 期刊报纸
 - 报纸 (2761)
 - 期刊 (2)

缩小检索范围

全文

- 可提供全文 (67277)

年份

- 2014 (1)
- 2013 (2478)
- 2012 (4769)
- 2011 (6034)
- 2010 (6770)

著者

- 秦珂 (74)
- 张晓林 (62)
- 张冲 (59)
- 张秀梅 (56)
- 刘燕权 (45)

语种

- 汉语 (66848)
- 其他 (1270)
- 英语 (16)

来源数据库

- 维普中文科技... (30641)
- 知网-中国学... (27505)
- 龙源电子期刊 (3712)
- 知网-中国重... (2761)
- 馆藏中文资源 (845)
- 知网-中国重... (795)
- 国务院发展研... (585)
- 知网-中国优... (575)
- 知网-中国年... (474)
- 方正阿帕比电... (77)

根据检索结果，动态生成导航过滤项

5、列表页汇集展示

找到相关结果约 720 个，当前显示: 1-10

1.



数字图书馆与文化共享工程

文献类型: 专著

著者: 陈雪樵

出版年份: 2008

出版社: 中国环境科学出版社

来源数据库: 馆藏中文资源

[摘要](#) [目次](#) [馆藏信息](#)

[查看其他版本或分册](#)

2.



数字图书馆与文化信息共享工程的...

文献类型: 期刊论文

著者: 伍清霞

出版年份: 2011

来源数据库: 知网-中国学术期刊全文数据库

来源: 图书馆学刊, Journal of Library Science, 2011, 01

[摘要](#) [目次](#) [馆藏信息](#)

[查看其他版本或分册](#)

3.



从数字图书馆与文化多样性的关系...

文献类型: 期刊论文

著者: 陈庆丽

出版年份: 2007

来源数据库: 维普中文科技期刊数据库

来源: 图书馆学刊, Journal of Library Science, 2007, 005

[摘要](#) [目次](#) [馆藏信息](#)

[查看其他版本或分册](#)

5、列表页汇集展示

1.



数字图书馆与文化共享工程

文献类型: 专著

著者: 陈雪樵

出版年份: 2008

出版社: 中国环境科学出版社

来源数据库: 馆藏中文资源

[摘要](#) [目录](#) [馆藏信息](#)

2.



数字图书馆与文化共享工程

文献类型: 专著

著者: 陈雪樵

出版年份: 2008

出版社: 中国环境科学出版社

来源数据库: 馆藏中文电子图书

[摘要](#) [目录](#) [馆藏信息](#)

1.



数字图书馆与文化信息共享工程的...

文献类型: 期刊论文

著者: 伍清霞

出版年份: 2011

来源数据库: 知网-中国学术期刊全文数据库

来源: 图书馆学刊, Journal of Library Science, 2011, 01

[摘要](#) [目录](#) [馆藏信息](#)

2.



数字图书馆与文化信息共享工程的...

文献类型: 期刊论文

著者: 伍清霞

出版年份: 2011

来源数据库: 维普中文科技期刊数据库

来源: 图书馆学刊, Journal of Library Science, 2011, 001

[摘要](#) [目录](#) [馆藏信息](#)

对检索结果的
数据按照规则
进行汇聚。

6、数据挖掘及数据关联



数字图书馆

文献类型：专著

责任者：高文等

出版、发行者：清华大学出版社

出版发行时间：2000

来源数据库：馆藏中文资源

分享到：

文献传递

延伸检索

到百度中查看相关资源

到Google中查看相关资源

更多相似资源

- 河北省公益性数字图书馆信息隐私权问题探究
- 数字图书馆系统开源软件的比较分析
- 数字图书馆热的冷思考
- 数字图书馆的组织管理
- 移动数字图书馆集成服务的体系结构与表现形式

“更多相似资源”是通过责任者、出版项进行关联得到

详细信息

摘要

目次

馆藏信息

所有责任者：高文等著

所有题名：其他题名信息：原理与技术实现

标识号：ISBN：7-302-04079-6

出版、发行地：北京

关键词：数字图书馆--研究 数字图书馆

语种：Chinese 汉语

分类：中图分类：G250.76

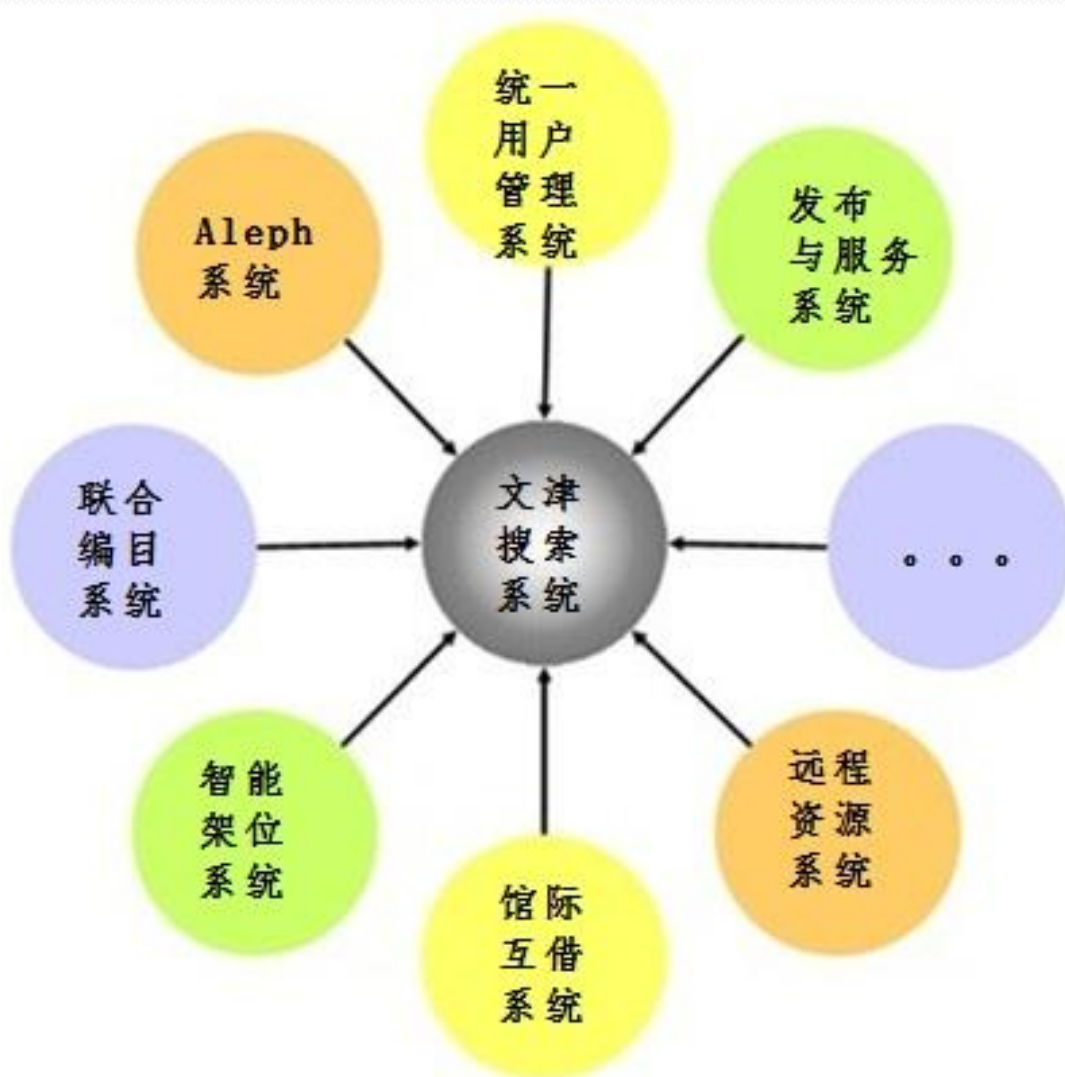
载体形态：474页

更多相关资源

- 航天国际合作项目进度风险管理研究
- Working out the inevitable
- 情报学研究进展
- 数字图书馆操作与实务
- 红楼梦

“更多相关资源”是通过对用户检索行为进行数据挖掘得到

7、整合多个系统的服务



8、整合统一用户管理系统



用户登录

账号:

密码:

验证码:

4729

登录

重置

返回

用户注册

密码找回

百万雄师过大江

文献类型：专著

9、联编数据的应用

书目查询 >> 书目相关信息

[重新检索](#)

[机读目录格式](#)

[返回上页](#)

文献所属馆

青岛市图书馆

[预约该书](#)

[预借该书](#)

书目相关信息	
【ISBN】	978-7-5358-3337-2 价格： CNY6.50
【语种】	chi
【题名】	百万雄师过大江 bai wan xiong shi guo da jiang / 虎郎著,
【出版发行】	出版地：长沙 出版社： 湖南少年儿童出版社 出版日期： 2007
【载体形态】	113页 照片，地图 21cm
【其它题名】	渡江战役始末记
【主题】	儿童文学 革命故事 中国 现代
【主题】	儿童文学
【主题】	革命故事
【中图分类】	I287.51 版本： 4
【主要著者】	虎郎 hu lang 著

[查看馆藏信息](#)

[查看预订信息](#)

[发表书评](#)

[收藏](#)

本书共有0条读者书评

馆藏信息

索书号	条码号	馆藏状态	借出时间	应还时间	所在馆	所在馆位置	流通类型	卷册信息	借阅次数	续借次数
I287.51/145	1083312	在馆			青岛市图书馆	少儿剔存	少儿剔存		2	0
I287.51/145	1083313	在馆			青岛市图书馆	少儿剔存	少儿剔存		2	0
I287.51/145	1083314	在馆			青岛市图书馆	少儿剔存	少儿剔存		2	0

青岛市图书馆

提供文档递送服务.

馆藏信息。

10、整合馆际互借系统

申请信息

版权声明

国家图书馆文献提供中心提供的原文复制件只能用于个人学习、研究或者学校课堂教学、科学研究的目的。如果用户超出著作权法规定的合理使用的范围使用文献，要为发生的版权侵权行为承担侵权责任。

同意 不同意

服务方式

复制（不提供外借服务）！

此服务为有偿服务！

特别提醒：

1. 缩微中心库所藏文献暂不提供读者借阅服务。
2. 需要复制古籍善本及拓片、舆图等特藏，请咨询sbyls@nlc.gov.cn。

文献信息

文献类型	<input type="text" value="期刊论文"/>	作者	<input type="text" value="陈臣"/>
出版物名称	<input type="text" value="图书馆论坛, Library Tribune,2013,001"/>		
卷期号	<input type="text" value=""/>	版本	<input type="text" value=""/>
文章名称	<input type="text" value="基于定制服务的数字图书馆云服务平台与云服务模式研究"/>		
出版地	<input type="text" value=""/>	出版者	<input type="text" value=""/>
出版年	<input type="text" value="2013"/>		
页码信息	<input type="text" value=""/>		
ISSN	<input type="text" value="1002-1167"/>	ISBN	<input type="text" value=""/>
条码	<input type="text" value=""/>	语种	<input type="text" value="中文"/>

馆际互借信息

申请来源	<input checked="" type="radio"/> 国内 <input type="radio"/> 国外		
有效时间	<input type="text" value=""/> YYYY-MM-DD(如果申请超过该时间还未被处理，则申请自动结束！)		
文献传递方式	<input type="text" value="e-mail方式"/>	联系方式	<input type="text" value=""/>
费用限制	<input type="text" value=""/> 元*	急迫程度	<input checked="" type="radio"/> 普通 <input type="radio"/> 加急

II、搜索热词榜



根据文献历史检索量情况建立热词排行、热点阅读、人气资源库、文津云图。

12、用户个性化设置

我的检索历史

我的检索历史

我的搜索设定

可查看历史检索词、检索文献类型、检索字段、执行时间

[清空检索历史](#)

检索词	检索文献类型	检索字段	执行时间
红楼梦	全部	全部字段	2014-04-15 09:48:32
国家数字图书馆	全部	全部字段	2014-04-15 09:48:09
数字图书馆	全部	全部字段	2014-04-15 09:47:53

我的检索历史

我的搜索设定

支持默认网页显示语言、保存检索历史、封面显示默认展示模块设定。

默认网页显示语言：

是否保存检索历史： 保存 不保存

检索结果是否展现封面信息： 展现 不展现

职业：

兴趣：

确定



THANK YOU!
QUESTIONS?