

# 管理元数据的原理与应用

张晓林 曾蕾 毛军 李广建 刘炜

管理元数据的实质是对信息系统管理机制的规范、开放描述。本文将首先讨论管理机制及其规范化描述, 然后探讨由此形成的管理元数据的作用, 并通过 W3C 的 P3P 元数据分析管理元数据的应用框架和开放设计。

## 1. 管理机制及其政策化和规则化

管理机制作为广义概念, 是根据特定的本地化 (Localized) 和个性化 (Personalized) 的

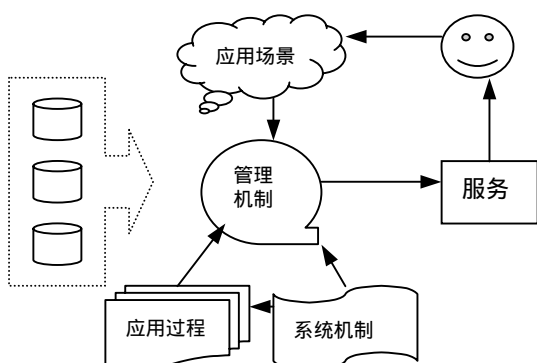


图 1

应用场景 (Context) 对裸对象 (Raw objects) 进行选择、组织、解释和表现以形成特定服务的系统方法与过程集合<sup>[1]</sup>。按照美国著名数字图书馆学者 Clifford Lynch 的说法, 管理机制实际上起着中间件的作用, 动态地实现上述方法与过程<sup>[2]</sup>。

这里, “裸对象”包括文献这样的信息内容对象、信息内容对象集合、完成特定任务的应用过程、提供选择控制的系统机制。管理机制可以表现为资源选择评价、质量控制、资源集成、资源调度、知识组织、使用控制、个性

化定制、版权管理、隐私保护、资源长期保护、流程控制、系统互操作管理等方面, 但它们的共性都在于根据具体的应用场景, 利用系统机制, 调整应用过程, 组合对象 (包括其内容与方式), 表现服务结果。

传统的, 管理机制表现为固化为专门的内部系统程序, 但实质上管理机制的核心是管理政策, 这些政策依据于一定的管理模式和描述这个模式中各个实体及其相互关系的管理词汇 (Vocabulary), 而且这些政策在一定应用场景中被现象化 (Operationalized) 为管理规则, 这些规则又由一定方式予以描述。描述方式可能是简单文本形式 (通过文件或网页信息予以公布), 可以被固化到计算机程序中 (例如使用控制程序), 也可以载入数据字典中供数据库系统调用后嵌入通用程序来控制运行过程。但长期以来, 在多数情况下管理规则被捆绑 (Hard-coding), 影响了管理规则及其所体现的管理政策的显现化和可调整性。但是, 当这些

管理规则以元数据进行描述时, 一方面可通过对元数据的修改来方便地实现管理机制的定制和修改, 另一方面也可支持智能代理对管理规则进行搜寻、解析, 并根据这些政策或规则来进行相应选择和和处理。

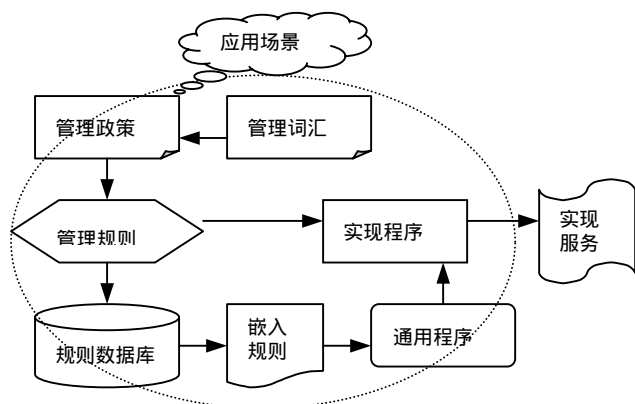


图 2

## 2. 描述管理政策与规则的管理元数据

传统的, 管理机制被认为是本地化问题, 人们往往从单一系统的角度来考虑和应用管理机制, 因此

即使有关于管理机制的描述, 也倾向于用内部的方法和专门的系统机制来实现。而且, 出于

对系统安全的担心甚至误解，人们也倾向于对用户和第三方屏蔽管理机制的实现原理和技术过程。但是，随着网络化的发展，信息服务本身已打破本地化局限，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方式和规范结构来描述数字图书馆系统的管理机制，形成管理元数据，通过这些元数据的发布、登记和传递，使得这些管理机制能够被广泛识别、理解和复用，以保障系统服务在整个网络空间的可用性和多个系统之间的互操作性。

(1) 将管理机制进行描述以形成管理元数据是元数据应用的深化。人们曾经习惯于将元数据纯粹看成是描述内容对象的数据集合。随着分布环境下信息系统互操作的现实发展，人们开始从用户在开放环境中甚至在业务流程中来有效利用分布、异构信息系统的角度，考虑如何描述信息系统的各个内容层次，包括内容对象、对象集合组织机制（体系组织和知识组织）、对象及其集合的管理机制、服务流程及服务流程管理控制以及信息系统本身，使得这些内容都能被规范地描述、交换、转换和理解。因此，通过元数据方式对信息系统的管理机制进行规范描述，成为在服务与管理层面促进互操作的有力工具。

(2) 管理元数据支持对系统服务的识别与选择。管理机制作为按照本地化和个性化应用场景对资源内容进行选择、组织、解释和表现以形成特定服务的系统方法与过程，直接制约着用户能否获得服务、是否信赖服务、如何控制服务。例如，用户依赖资源选择标准、服务效率规定、检索相关度评价标准等来判断是否可以信赖对象系统，倚赖使用控制策略（身份认证与使用授权规则等）和使用费用规则来了解能否使用相关服务，倚赖隐私保护政策、知识产权保护规定、个性化定制程度与方式等来决定是否使用这些服务。当这些管理政策被公开描述时，用户可以方便地进行了解和判断；当这些管理政策被描述为计算机可读的元数据时，用户可以借助代理系统自动了解相应政策；当用户选择服务的条件也用相同的元数据语言进行描述时，用户可以通过自己的代理系统自动识别和选择符合自己选择条件的服务。

(2) 管理元数据支持多个系统在服务层面的互操作。在一个开放和集成环境中，用户往往通过一个系统来连接和使用其他的系统，这些被连接的系统形成一个服务链，为用户提供逻辑化的服务，例如从文摘检索系统到全文文献系统及其参考文献到期刊联合目录系统到馆际互借系统到参考咨询系统，或者从一个整合检索系统到多个数据库系统。这些不同的系统都有自己的管理机制（例如身份认证或费用规定），但显然不应该让用户分别去面对一个个“环节”系统的管理机制，而应该通过管理机制间的互操作来传递和处理管理信息，例如用户在进入第一个“环节”系统时进行了身份认证，这个系统就应该将认证信息传递给下一个“环节”系统来进行第二个环节的认证；只有在前一个环节的认证信息不足以满足当前系统认证需要时才要求用户提供进一步（而不是重复）的认证信息。要做到这样的互操作，就要求各个系统之间能传递和识别管理政策及相应管理数据，而这又依赖规范描述。

(4) 管理元数据支持多个系统对管理政策及相应管理数据的共享。在系统之间进行互操作时，需要交换管理政策及相应管理数据，其中相当部分在各个系统都需要，因此可以进行复用。例如在使用控制中的用户信息、需求表达数据、使用条件数据等。如果各个系统都使用标准方式（或基于开放语言的描述方式），就可以解析其他系统的管理政策及相应管理数据，将其复用到自己的管理数据中，既减少输入要求又提高描述一致性。

当然，人们在讨论管理元数据时可能担心系统安全问题。但是，元数据所描述的是管理政策所规定的管理规则，并不是具体系统实现管理规则的技术机制，而且明确地区分管理规则和实现规则的技术机制还有利于对系统的安全控制。另一方面，提供明确、公开、公正和可靠的管理规则是任何一个系统得到信赖和利用的必要条件，在许多情况下，法律还要求公开有关管理政策，例如隐私保护、知识产权管理等。因此，建立描述管理机制的元数据并不会威胁系统管理的安全性，即使是管理数据传递，也可以通过一定的加密方式予以保护。

### 3. 管理元数据的应用框架

管理元数据的基本应用框架可以由图 3 表示。从服务系统角度，人们一方面根据管理环境

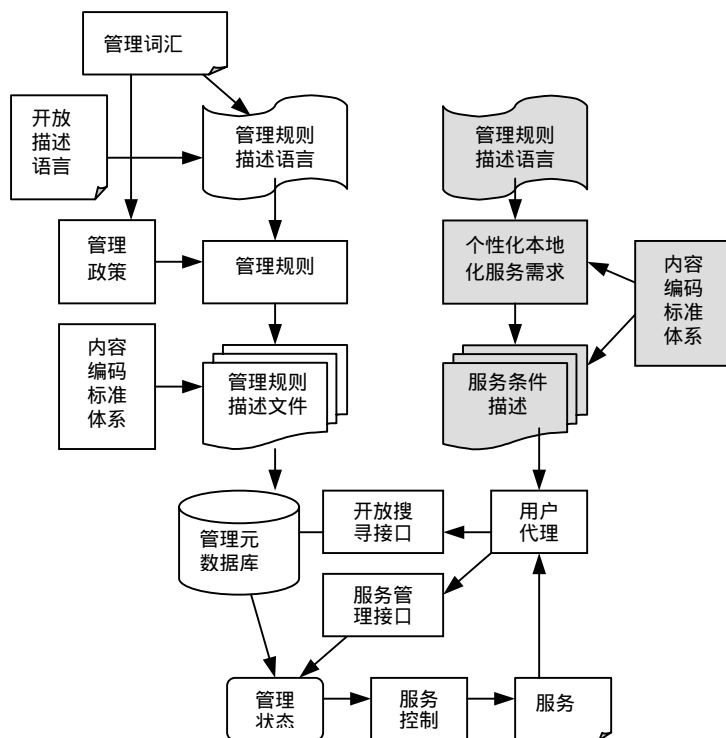


图 3

环境建立标准或通用的管理词汇，支持不同系统对管理对象、管理对象间关系和管理过程进行标准描述；进而，人们可根据具体应用系统和场景的管理政策，利用管理词汇析出相应的管理规则，准确表示管理对象及其类别，规范表达管理过程及其选择控制条件；同时，人们利用开放描述语言（例如 XML）定义管理规则描述语言，提供对管理规则进行描述的公共工具；然后，利用管理规则描述语言来描述具体的管理规则，形成管理规则描述文件，这些描述文件就是管理元数据。另一方面，人们可以利用同样的管理规则描述语言及其内容编码体

系，对个性化本地化服务需求进行描述（例如用户身份及其读取要求），形成针对具体需要的服务条件描述，也是用开放语言描述的用户服务条件“元数据”。

管理元数据可以置于系统的管理元数据库，供用户代理进行搜寻，然后将管理元数据与相应的用户服务条件元数据比较匹配来进行服务选择。管理元数据也可以供系统调用，与从服务接口获得的用户服务条件数据（以元数据形式或输入参数形式）进行比较匹配形成管理状态参数，驱动服务控制程序，提供相应服务。

一般来说，管理元数据不涉及各个系统利用管理元数据进行管理控制的具体方法。

### 4. 管理元数据开放设计

我们以 W3C 的 Platform for Privacy Preferences<sup>[3]</sup>为例说明管理元数据开放设计。

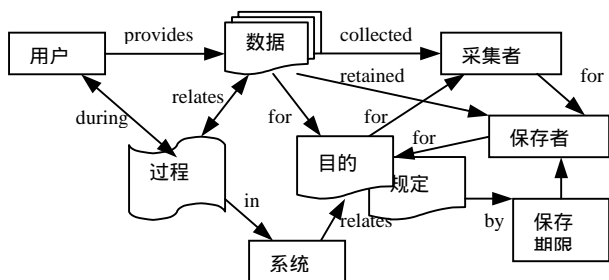


图 4A

(1) 数据模型。数据模型是元数据开放设计<sup>[4]</sup>的基础,通过对应用过程涉及实体及其关系的分析和分解,设计或复用不同的元数据模块来分别满足不同的逻辑功能,并在此基础上构建元数据逻辑框架。P3P 作为隐私保护政策元数据,它的目的是明确描述某个系统采集用户数据时的有关政策,以使用户判断这个政策是否符合自己的

要求。在这个过程中,涉及一系列实体、实体间关系和过程,对它们的描述就是 P3P 的数据模型,构成 P3P 设计和实施的基础(图 4A)。图 4A 中,用户在使用某个系统的某个功能过程中,主动(例如填表)或被动(例如系统跟踪点击对象和顺序)提供一定的用户数据,被采集者为了特定目的(例如个性化定制或第三方配送)或按照特定规定(例如法律规定)采集,并为了一定目的或按照特定规定由采集者自己或交一定的保存者保存一定期限。功能过程和用户数据之间、采集目的或规定和系统之间,都必须存在合理的关联。

(2) 文件模型。上述数据模型要通过一定的数据元素来实现,这些数据元素要被组织到一定的格式或文件之中。P3P 规定了两类文件来组织自己的元数据,如图 4B 所示,包括政策指向文件(Policy references)和政策文件(Policy)。政策指向文件以 URI 形式列出各个政策文件以及所应对的资源集合(Resources);而政策文件则描述数据收集者(Collector)及其关于隐私数据的保护声明(Statement),这些声明具体描述所收集的数据集(Dataset)、收集目的(Purpose)、作用(Consequence)、数据接收者(Recipient)、数据保存要求(Retention)等,并描述关于收集者和保存者的数据集。按照开放元数据原则,P3P 中的隐私数据集、收集者数据集、保存者数据集都可复用其他元数据所定义的数据集。

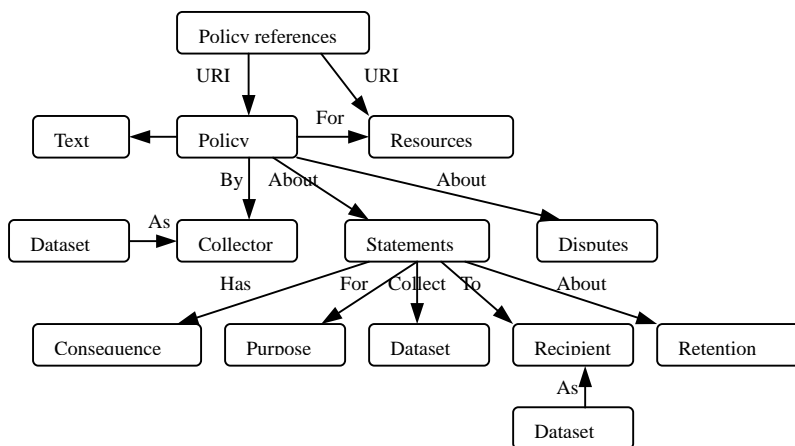


图 4B

(3) 开放描述工具。P3P 致力于定义一套规范的计算机可读的描述语言,为此 W3C 设计了相应的语言工具,包括:一组关于管理政策及其细节的词汇元素集(Vocabulary),一组常被收集的隐私数据集和收集者保存者数据集定义(Datasets,统称为 P3P Data Schema),一个利用上述词汇元素集描述政策

文件的 XML 格式(P3P Policy Schema),一种将政策文件与特定资源捆绑的方法(实际上就是政策指向文件的格式 P3P Policy Reference Schema),以及通过 HTTP 传输 P3P 元数据的方法。实际上,这也体现了管理元数据的基本设计方法,即主要是定义元数据语言(描述词汇及其 XML Schema)以及相应的应用机制。

(4) 元数据应用机制。一般的,用管理元数据语言描述的管理政策将作为管理文件,可以放置于系统公知位置,或在元数据登记系统注册,或嵌套在系统描述元数据中,或通过其他方式传递给第三方系统。例如,P3P 规定,网站的 P3P 政策指向文件应置于根目录下的通用位置/w3c/p3p.xml,也可以嵌入主页(利用 HTML/XHTML link 标签)或在 HTTP 应答中

嵌入 P3P 扩展头标来标明文件位置，而关于具体资源的政策文件则在政策指向文件中明确标识，从而保证用户代理准确获取政策指向文件和政策文件。

## 5. 其他管理元数据范例

目前，各类机构已定义了多种管理元数据描述语言及相应机制，其中有些相对比较成熟，有些尚待完善，但已反映出系统地建设管理元数据的努力<sup>[5]</sup>。

(1) PICS 资源评鉴元数据。信息资源评鉴信息指根据专家分析、用户反馈、使用量分析、链接量分析等方法对资源内容的主题范围、正确性、新颖性、权威性等的判断或评价。当评鉴信息通过开放元数据进行描述后，用户可通过智能代理查询和释读资源评鉴信息，并根据评鉴信息自动选择信息资源。W3C 定义了“互联网内容选择平台” PICS)<sup>[6]</sup>，提供了一个可适应多种应用的评鉴元数据应用模式和标记语言。

(2) ODRL 知识产权管理元数据。知识产权管理是指对数字化信息产品在网络中交易、传输和利用时各方权利进行定义、描述、保护和监控的整体机制。这个机制必须定义和描述谁拥有什么产品的什么权限、按照什么协议和交易方式将哪些权限在什么范围授予给谁，而相应的定义与描述信息可用开放描述语言，形成描述知识产权管理的元数据，系统可能自动进行记录、识别、解析，并据此进行权限控制。现已定义若干语言，如 ODRL<sup>[7]</sup>和 XrML<sup>[8]</sup>。

(3) 使用控制元数据。用户使用控制 (Access control) 包括身份认证、使用授权以及使用审计等功能，是分布用户环境下信息资源系统管理的核心机制之一。在传统情况下，信息系统采用内部程序模块固化 (hard coding) 使用控制政策，但在开放环境中，人们可利用开放语言标记使用控制政策，支持异构系统在服务层的互操作。有关研究包括 IBM 的 XACL<sup>[9]</sup>和 OASIS 的 XACML<sup>[10]</sup>。

上述管理元数据语言主要用于描述和标记管理政策，类似的还有个性化信息描述语言 PIDL<sup>[11]</sup>和安全条件标记语言 SAML<sup>[12]</sup>。除此之外，还有几类元数据语言与管理机制有关：资源集合元数据，对资源集合及其管理与使用机制进行描述，例如 RLSP CDS<sup>[13]</sup>、UKOLN SCD<sup>[14]</sup>等；长期保存元数据，对数字资源的长期保存历史、政策、技术条件和措施等进行描述，例如 CEDARS 等<sup>[15-16]</sup>；系统元数据，对资源或服务系统的基本接入条件、界面和输入输出要求等进行描述，例如 CC/PP<sup>[17]</sup>、WSDL<sup>[18]</sup>等；业务流程元数据，对业务操作过程及其相关实体和相关操作条件等进行描述，例如 ebXML BPSS<sup>[19]</sup>和 WSFL<sup>[20]</sup>，以及更高层的 UML/XML<sup>[21]</sup>和 5SL<sup>[22]</sup>。

## 6. 管理元数据开发要求

管理元数据作为开放描述机制，其描述对象是管理政策，描述结果是管理元数据，通过一定系统机制来实施用元数据表现的政策。管理元数据的应用需要以下工具：关于特定管理需求的管理词汇集，用于描述管理问题所涉及的实体、实体关系、过程和规则，最典型的是 INDECS<sup>[23]</sup>元数据词典；关于特定管理政策或规则的描述语言，例如前述的 PICS、XACL、ODRL、XRML、P3P 等管理政策描述语言；管理元数据的组织与传递机制，例如管理元数据文件结构和这些文件的发布、传递与调用要求。一般的，利用管理元数据来实现管理控制本身属于本地系统内部机制。

虽然管理元数据从整体上讲还处于探索和实验中，但利用元数据机制描述这些政策并通过元数据实现对这些政策的自动解析和互操作，将是开放数字图书馆体系的基本要求。因此，应该积极推动管理元数据的发展与应用，例如分析整个系统所涉及的管理需求领域（例如用户认证、个性化定制、参考咨询、馆际互借、资源选择、知识产权控制等），明确建立、规范描述和公开发布管理政策；选择若干互操作迫切、管理政策相对成熟、已有初步管理元数

据的管理领域，建立管理模式框架，逻辑描述管理实体与关系，建立公共的管理词汇集；选择开放的管理政策描述语言，建立相应业务管理的管理政策描述元数据；在数字内容对象和资源集合元数据中建立开放接口，嵌入或链接相应的管理元数据，从而可以内容对象或资源集合元数据对管理元数据进行调用和验证；管理元数据也应在公共元数据登记系统上发布，为了支持第三方系统发现、调用和解析管理元数据；无论管理元数据的具体内容和格式是什么，应该严格基于 XML 语言，应该采用通用的应用描述语言，保证管理元数据的开放性和可解析型。

## 参考文献

- [1] 张晓林. 元数据研究与应用. 北京: 北京图书馆出版社, 2002
- [2] Lynch, Clifford. Digital Collections, Digital Libraries, and the Digitization of Cultural Heritage Information. First Monday, 7(5), 2002. [http://firstmonday.org/issues/issue7\\_5/lynch/](http://firstmonday.org/issues/issue7_5/lynch/)
- [3] Platform for Privacy Preferences Project. <http://www.w3.org/P3P/>
- [4] 张晓林等. 数字图书馆的开放元数据机制. 研究报告, CSDL2002-15-2
- [5] 同[1]
- [6] Platform for Internet Content Selection (PICS). <http://www.w3.org/PICS/#Specs>
- [7] Open Digital Rights Language. 1.0. Dec. 21, 2001. <http://odrl.net/ODRL-10.HTML>
- [8] XrML 2.0 Technical Overview. Version 1.0. March 2002  
<http://www.xrml.org/reference/XrMLTechnicalOverview.pdf>
- [9] XML Access Control Language: Provisional Authorization for XML Documents, Oct. 16, 2000.  
<http://www.trl.ibm.com/projects/xml/xacl/xacl-spec.html>
- [10] Extensible Access Control Markup Language. <http://www.oasis-open.org/committees/xacml/>
- [11] PIDL: Personalized Information Description Language. W3C Note, Feb.9, 1999.  
[http://www.w3.org/TR/NOTE\\_PIDL](http://www.w3.org/TR/NOTE_PIDL)
- [12] Security Assertion Markup Language (SAML). <http://www.oasis-open.org/committees/security/>
- [13] RSLP CDS. RSLP Collection Description Schema. May 2000. <http://www.ukoln.ac.uk/metadata/rsllp/schema/>
- [14] UKOLN Simple Collection Description. Draft, August 2, 1999. <http://www.ukoln.ac.uk/metadata/cld/simple/>
- [15] Metadata For Digital Preservation: The Cedars Outline Specification. March 2000.  
<http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>
- [16] 梁娜, 张晓林. 关于数字资源长期保存的元数据. 四川图书馆学报, 2002 ( 1 )
- [17] Composite Capability/Preference Profiles (CC/PP): Structure and Vocabularies. W3C Working Draft, March 15, 2001. <http://www.w3.org/TR/CCPP-struct-vocab/>
- [18] Web Service Definition Language (WSDL) Version 1.2. W3C Working Draft, July 9, 2002.  
<http://www.w3.org/TR/wsd12/>
- [19] ebXML Business Process Specification Schema. Version 1.1, May 11, 2001.  
<http://www.ebxml.org/specs/ebBPSS.pdf>
- [20] Leymann, F. Web Service Flow Language (WSFL 1.0) May 2001.  
<http://www-3.ibm.com/software/solutions/webservices/pdf/WSFL.pdf>
- [21] UML2XML-Initial List of Requirements, Oct.9, 2001.  
<http://xml.coverpages.org/UML2XMLReq1-day2.pdf>
- [22] Goncalves, M. and Fox, E. 5S--A Language for Declarative Specification and Generation of Digital Libraries. Joint Conference On Digital Libraries, July 13-17, 2002. Portland, USA.  
<http://www.dlib.vt.edu/projects/5S-model/p117-goncalves.pdf>
- [23] The INDECS Metadata Framework: Principles, Model and Data Dictionary, June 2000.  
<http://www.indecs.org/pdf/framework.pdf>

原载于《图书情报工作》2003 年第 10 期  
作者署名：张晓林 曾蕾 毛军 李广建 刘炜