

科技部科技基础条件平台工作重点项目 研究成果

项目名称：数字图书馆标准与规范建设

子项目名称：元数据标准规范开放登记系统研究

项目编号：2003DEA4T035

研究成果类型：研究报告

成果名称：元数据标准规范开放登记系统发展趋势

成果编号：CDLS-S08-001

成果版本：总项目组推荐稿

成果提交日期：2004年6月

撰写人：梁 娜（中国科学院文献情报中心）

张晓林（中国科学院文献情报中心）

项目版权声明

本报告研究工作属于科技部科技基础条件平台工作重大项目《数字图书馆标准与规范建设》的一部分，得到科技部科技基础条件平台专项资金资助，项目编号为 2003DEA4T035。按照有关规定，国家和《我国数字图书馆标准规范建设》课题组拥有本报告的版权，依照《中华人民共和国著作权法》享有著作权。

为了学习、研究和应用推广等目的，可以复制、转载、或在电子信息系统上镜像本报告。但在复制、转载或镜像时，必须在明显地方标明“科技部科技基础条件平台工作重大项目《数字图书馆标准与规范建设》项目资助”的字样，必须保证本报告的完整性，必须注明总项目组、子项目组和作者的真实名称。任何人不得以商业赢利的目的复制、转载、镜像、或以其他形式传递和发布本报告。

报告版权人不承担用户在使用本作品内容时可能造成的任何实际或预计的损失。

作者声明

本报告作者谨保证本作品中出现的文字、图片、声音、剪辑和文后参考文献等内容的真实性和可靠性，愿按照《中华人民共和国著作权法》，承担本作品发布过程中的责任和义务。科技部有关管理机构对于本作品内容所引发的版权、署名权的异议、纠纷不承担任何责任。

《数字图书馆标准与规范建设》课题组网站 (<http://cdls.nstl.gov.cn>) 作为本报告的第一发表单位，并可向其他媒体推荐此作品。在不发生重复授权的前提下，报告撰写人保留将经过修改的项目成果向正式学术媒体直接投稿的权利。

元数据标准规范开放登记系统发展趋势

目 录

| | |
|-------------------------------------|-----------|
| 1. 元数据登记系统的定义与意义 | 1 |
| 1.1 元数据管理..... | 1 |
| 1.2 元数据互操作..... | 1 |
| 1.3 元数据开放应用..... | 2 |
| 1.4 元数据复用..... | 2 |
| 2. 元数据登记系统的功能结构 | 3 |
| 2.1 元数据登记系统的基本类别..... | 3 |
| 2.2 元数据登记系统的基本功能..... | 4 |
| 2.3 元数据登记系统的登记内容..... | 5 |
| 3. 元数据登记系统的技术结构 | 7 |
| 3.1 元数据登记系统的标准模型..... | 7 |
| 3.2 人工登记与检索的范例系统..... | 10 |
| 3.2.1 ROADS Metadata Registry | 10 |
| 3.2.2 DESIRE Metadata Registry..... | 11 |
| 3.2.3 DCMI Metadata Registry | 13 |
| 3.2.4 Cores Metadata Registry | 14 |
| 3.3 基于自动发布和发现的MR | 16 |
| 3.3.1 ebXML Registry Services | 16 |
| 3.3.2 UDDI Registry | 17 |
| 3.3.3 GoXML Registry..... | 19 |
| 4. 元数据登记系统的发展趋势 | 20 |
| 4.1 依据标准模型..... | 20 |
| 4.2 基于WEB SERVICE的元数据登记系统 | 20 |
| 参考文献 | 22 |

1. 元数据登记系统的定义与意义

元数据登记系统 (Metadata Registry, MR) 是对元数据的定义信息及其编码、转换、应用等规范进行发布、登记、管理和检索的系统^[1-3], 支持开放环境中元数据规范的发现、识别、调用以及在此基础上的元数据转换、挖掘和复用。元数据登记系统可以发挥多重意义。

1.1 元数据管理

元数据登记系统就是要建立一个权威、可靠、可持续的和公共的元数据规范及应用信息的登记管理机制, 从而支持对元数据规范的应用。在数字图书馆系统中, 建立元数据规范登记系统, 对内部各种不同信息资源的数据格式、元素、元素语义定义、元素应用规则、元素映射关系等进行记载, 支持系统对这些资源的管理和整合利用。对于从外部引进的信息资源, 通过提供指向详细信息的链接, 可以在登记系统中直接利用资源提供者的数据规范。同时为了支持对元数据规范的管理, 元数据登记系统还应该包括元素字典、元数据格式定义、元数据描述模板、元素内容编码体系、元数据规范翻译、元数据扩展应用协议、元数据转换模板等, 为元数据规范的发展和应用提供全面的信息和工具。

1.2 元数据互操作

由于现实信息环境是由众多分布、异构、自主、变化的信息资源与服务系统构成的开放环境, 这些系统往往从各自角度来建立自己的元数据, 因此形成多种不同的元数据定义和应用规则。即使使用某种“标准”元数据格式的系统, 也可能各自进行了不同的扩展, 从而形成不同的应用格式。为了集成利用这些系统的资源与服务, 需要发现、解析、转换这些系统的元数据, 来支持跨系统的检索、获取和数据复用。

传统的进行元数据转换的方法是事先 (往往是人工地) 分析各个资源系统, 解析其元数据格式、建立相应的转换关系, 并将这些转换关系写入转换模块之中。但是, 这种方法要求事先获得和分析转换局部元数据格式本身可获得, 工作难度大, 尤其难以应付大规模异构环境和动态环境。如果各个资源系统能够将元数据格式通过一个公共登记系统予以发布, 将有效支持其它系统通过登记系统来发现和调用这个定义信息。元数据登记系统建立和发布专门的元数据格式与其他元数据格式的映射表和转换模板, 能支持第三方系统动态调用这些映射表和转换模板, 或支持第三方系统开发相应的转换模板, 并提供与这些映射表和转换模板的链接。

1.3 元数据开放应用

在开放环境里，一个应用系统不仅将利用众多其它资源和服务系统，同时还应该支持其它系统对自己的利用、支持系统间的互操作。因此，应用系统应该帮助其它系统发现、识别和利用自己的资源和服务，这就要求元数据格式的开放应用^[5]。开放应用实际包括开放描述、开放发布和开放转换，其中与元数据登记直接相关的是开放发布。

开放发布要求元数据定义以及所依据的标准、规范或协议等都是公开可查询和获取的，才能保证第三方可靠了解开放描述信息并基于开放描述进行互操作。发布方法包括在本地系统上公布，供人工查阅；在本地系统某个公知位置存放定义文件，供第三方系统自动搜寻；提交公共登记系统公布，供公共查询和调用。开放发布信息可以是文本文件，但应尽量以计算机可识别的 XML 文件方式公布定义文件及其支撑定义文件，支持自动解析、识别和利用元数据定义。开放转换也与开放发布有关，指采用专门元数据格式的应用系统应支持本地格式与其它格式（尤其是通用格式）的转换，本地系统应发布自己的映射关系表、转换模块和转换链接，支持对它们的共享。

1.4 元数据复用

尽管应用系统可以在本地系统上发布相关的元数据定义信息，但通过公共的元数据登记系统来发布和查询有利于更大范围和更方便经济的元数据定义信息的发现、复用和共享。元数据登记系统不仅登记元数据定义、映射表、转换模板，还支持相关支撑这些定义、映射表和转换模板的内容登记。它包括元数据所在领域的主题标记语言（例如 Chemical Markup Language^[10]、Mathematical Markup Language^[11]）、主题字典、概念集定义等，并建立它们之间的解析与推理链接，从而以更经济方便的方式支持更大范围内元数据规范信息的发现与复用。

综上所述，元数据登记系统作为公共、公开、可靠和可持续的机制，支持元数据定义信息的规范化发布和检索，对于复杂环境下的开放的数字信息服务机制具有重要意义。

同时我们还要区别元数据登记系统(Metadata Registry)和元数据库(Metadata Repository)，后者是对数字对象进行描述的实际元数据记录的集合，分布环境中常将从多个数字资源系统搜集而形成的元数据记录库称之为 Metadata Repository，例如 NSDL 体系中利用 OAI 元数据搜寻协议从多个数字对象库(Digital Repository) 搜集元数据记录组成元数据库^[12]。

2. 元数据登记系统的功能结构

2.1 元数据登记系统的基本类别

(1) 单一命名域的元数据登记系统

单一命名域 (Namespace) MR负责管理一个命名域的元数据, 例如DCMI Registry^[13]、ROADS registry^[14]等。它能够实现一般的元数据规范管理功能, 并且往往详细登记该名称域的应用进展、研究报告、发展动态等。这类MR承担了该命名域的元数据管理和应用推广的职责, 不但提供元数据规范编制单位关于其格式、元素和修饰属性的权威定义, 帮助人们查询、了解和验证元数据格式、元素及其使用要求, 还提供其他应用单位提供的应用协议、翻译格式和转换模板的公共登记, 形成该命名域元数据的研究、应用和发展的聚焦点。虽然它们可能被认为不是严格意义上的公共MR (因为只登记了一种元数据的相关信息), 但能够提供权威、系统和全面的信息, 便于其他系统对该命名域元数据的发现和利用。

(2) 跨命名域的元数据登记系统

跨命名域MR管理多个命名域的元数据格式, 如German Metadata Registry (GMR)^[15]、DESIRE Metadata Registry (DMR)^[16]和Schemas Registry (SR)^[17]。GMR对包括教育、医学、物理和数学等领域的元数据格式进行登记和管理, DMR则登记了ACORE、BIBILINK、DC、IEEE LOM、IMS、NC-UK、ROADS、VCARD、VTC等多个命名域的元素、内容编码体系、应用协议、转换模板等, 而SR按照DMR模式登记了AGLS、DC、DC-GOV、DC-LIB、RQS、ETB、EUN、FAO、IEEE LOM、CLD等命名域的元数据信息。

跨命名域 MR 按照统一的数据模式 (Data Model) 对不同命名域的元数据规范进行描述和管理, 尤其是对不同规范中可能的关联进行描述和管理, 从而可以对同一逻辑对象或逻辑关系在不同元数据格式中的表达情况进行分析比较, 例如同一概念在不同元数据格式中的表达形态、同一元素在不同元数据格式中的语义差异、不同元数据格式中关于同一对象或关系的描述方法等等。跨命名域 MR 并不为了维护某个命名域数据元素的权威性, 因为这样的权威已经存在, 而是通过对各种命名域和语种的元数据格式定义、元素语义、应用规则等的规范描述来支持元数据的识别、复用、映射转换操作。

(3) 分布式元数据登记系统

分布式元数据规范登记体系由多个元数据登记系统组成^[2], 各个系统可能针对专业 (应用) 领域建立, 例如地理空间MR、教育MR、音频视频对象MR, 或者按照元数据类型建立MR, 例如资源集合元数据、内容对象元数据、知识组织体系元数据、管理规则元数据等。当然还可以按照地域或机构等的划分来分工建立某个领域或某类元数据的MR。这些MR中有些可能是单一命名域MR, 另外一

些可能是跨命名域MR。由于各个数字图书馆体系可能建立自己的MR（甚至多个MR），在实际环境中也许不可避免地将面对分布式MR体系。这些体系可能是统筹规划建立的，因此它们可能有同样或兼容的数据模式和界面，允许用户采用相同或相似方法进行登记和检索，并有可能建立关联查询和链接机制。当然，分布式MR体系也可能是自然累积的产物，这时各个MR在内部组织和外部接口等方面都不相同，增加了使用的多样性，也增加了关联查询的难度。无论哪种情况，MR都应作为服务模块在分布服务登记系统中进行登记，支持第三方系统对它们的开放搜寻和调用。

2.2 元数据登记系统的基本功能

（1）登记操作。MR支持人工登记，通过上载模板可以直接将各个规范文件上载，MR可根据上载者填写的信息对上载文件进行分类和组织，并可依据模板信息进行自动发布。MR也支持自动登记，元数据维护机构可以直接利用MR的登记应用接口（API）上载要登记的有关定义信息。由于登记信息的复杂性，可能两种方式都应支持。MR将为所登记的定义文本建立相应的元数据，这些元数据可直接从上载模板中提取，或者从机器可读的定义文本中提取。

（2）数据管理。当一个元数据规范被登记时，产生两类信息：关于被登记规范的元数据描述信息和规范文本本身。前者一般用数据库方式予以组织，后者则通过结构化文本目录组织，两者之间通过元数据记录中的链接来连接。当然，MR允许某个登记机构（Registration Authority, RA）只登记关于某个规范的元数据，建立与存放在RA的实际规范文本的链接，但并不实际上载该规范文本。需要注意的是，对一个元数据格式，相关的多方面的定义信息或应用信息可能被不同的系统在不同的时间登记到MR（例如不同系统编制的关于同一种元数据格式的多个应用规范、某个应用系统新提出的元素内容编码标准等），因此MR的一个重要任务是建立一种机制，保证相关的规范文件被相关地链接。这意味着在上载登记时要按照MR数据模型来描述所上载的规范文本的类别、对象、关系等，同时MR要确认新上载文本与已有定义文本的关系，并通过有效的内部组织结构来标识这些关系。

（3）检索与发布。MR的发布往往通过一个WEB服务器实现，支持多种检索，尤其应该支持基于关联关系的检索。MR也应支持浏览，根据命名域和规范类型来浏览显示有关信息。检索和浏览可以是专门的系统机制，也可以采用公共的检索界面（例如用WSDL，Web服务描述语言^[18]描述），描述信息在公共服务登记系统登记，支持智能代理对它的发现、配置和检索。

（4）元数据解析与转换支持。MR利用所登记的元数据规范、元数据内容关联等方式，可以支持对元数据定义信息的逐层解析，例如元素来自什么命名域、

元素语义是什么、内容编码标准是什么、采用什么描述模本、编制指南是如何要求的、有无编制范例等。解析本身可以是人工的，即通过关联查询来分析了解；解析也可以是自动的，通过MR数据库所记载的关联关系自动揭示、链接、调用。在跨命名域MR中，可以利用各个命名域的元素字典、元素语义网络^[19]、概念集（Ontology）^[20]等，进一步追询元素语义定义和语义关系。通过查询描述同一概念或同一对象的元素名称，分析它们的语义对应程度，确定这些元素的转换关系；可登记和查询元数据格式间的转换关系或转换模板，支持应用系统直接复用有关转换模块；可以通过循证元素的语义定义链，利用共同的根概念集（Root Ontology）自动建立不同命名域中元素的语义关联，从而辅助构建元素转换关系。

（5）自我描述。MR本身是一种网络服务（Web Service），可按照规范形式（例如WSDL）进行开放描述，描述信息被登记到分布服务登记系统，供第三方系统查询和调用。

2.3 元数据登记系统的登记内容

元数据登记系统所能提供的功能，不仅取决于登记系统的具体操作功能，而且与所登记的元数据规范内容、所登记内容的描述方式、这些内容的关联形式有密切关系。

（1）元数据规范的内容层次

A. 元数据定义信息，包括元数据格式定义、元数据应用规范和定义者信息。元数据格式定义，作为最基本的定义，帮助人们了解关于特定元数据的权威信息。这个定义的主要内容包含元素内容定义（采用ISO 11179^[21-26]规定），也包括相应的结构与语法描述规则定义。元数据应用规范往往可以看成是一个应用格式，针对一个应用系统或领域，具体规定使用一个元数据格式或复用多个元数据格式中某些元素时的方法，例如是否选择使用某些元素（例如必备元素、可选元素、或者暂时不用的元素），如何规定元素的取值类型（Value domain）、取值范围（Value range）和内容编码体系（Encoding scheme），如何规定元素的出现基数（Cardinality）等。元数据定义者及其定义过程信息（又称Registration Authority信息），包括创建者、发布者、维护机构、维护机构地址与联系方式、登记机构、登记机构地址与联系方式及版本信息等。

B. 元数据使用信息，包括元数据内容编码体系、使用指南、编制模板信息、编制范例、翻译文本和转换模板等。元数据内容编码体系描述编制元数据所需要的相应标准和规范，包括自建编码体系和外部编码体系；自建的编码体系应提供独立的定义文本进行规范描述，外部编码体系应提供对定义文本的链接。元数据使用规则（或指南）规范地说明元数据应用的具体原则、方法和过程，这些规则可能来自有关应用单位的最佳实践，可以作为权威的应用指导。元数据编制模板

描述并链接可用于该元数据编制的程序模块，尤其是开放软件模块。元数据编制范例，一般采用最佳实践，通过丰富实例来说明元数据的应用方法，可能针对多个标记方式（例如XML DTD或RDF^[27]）或多个应用协议有多个范例。元数据翻译文本信息，为多语种环境下准确表达元数据定义和使用要求建立标准文本；这些文本可能在本地登记系统，也可能在其它系统上而被链接。元数据转换模块，描述或链接该元数据格式与通用元数据格式、其他主要领域元数据格式等的转换映射表定义，甚至可以指向转换程序模块。

C. 元数据语义定义信息，主要指元数据所依靠的语义体系，具体体现可能是元素字典（详细定义某个应用领域可能涉及的众多元素、属性或关系）、语义网络、概念集等。元素的基本语义一般在元数据格式中采用基于ISO/IEC 11179规定（10个属性）进行定义，但这种定义只能提供语义查询，难以支持语义关联处理。元素字典将元素作为对象进行登记，记载元素名称、唯一标识符、概念含义、应用范围、与其他概念或元素的层级和关联关系等，可以直接利用权威叙词表，也可以通过应用分析建立。例如INDECS元素字典^[28]和NISO TMI元素字典^[29]。语义网络侧重记载元素（或其所隶属的概念）与其他元素或概念的逻辑关系，例如作者与文献、人物与事件、河流与城市、疾病与医疗过程等，典型的范例是ULMS^[30]的Semantic Networks。概念集在描述概念层级体系的同时，描述概念属性、概念间关系和概念关系推理规则^[31]，例如OIL^[32]等。这些语义定义信息从不同深度支持对元素语义的查询、验证和推理。

（2）元数据规范内容的描述方式

在MR中，元数据规范信息可以是文本文件或数据库记录方式，供人工浏览。但也可以是计算机可识别形式。例如格式定义可采用XML DTD/Schema模本，元数据定义者信息可采用专门的基于XML的元数据规范，转换模板可采用XSLT^[33]语言描述，内容编码体系可采用基于XML的词表描述语言（例如VocML^[34]），编制范例可以直接用基于XML的记录描述形式，语义定义可采用基于XML的描述语言（例如XTM^[35]、OIL等），元数据编制模块、转换模块等软件工具，也可以用Web Service方式（例如WSDL语言）进行描述。

（3）元数据规范内容的关联形式

MR中（尤其是跨命名域MR）往往涉及许多相互关联的内容，一方面可以是同一元数据格式与涉及它的应用规范、内容编码体系、使用指南、编制范例、编制模块、转换模块、翻译文本等应用工具间的关联，支持人们查询“使用到这种格式的应用规范有哪些”、“这种格式的使用指南（或编制模块等）是什么”等问题。这种关联一般通过在相应登记信息中嵌入对相关应用工具的描述和链接，例如在应用规范中描述所复用的元数据格式、所采用的内容编码体系和使用指南、可供选择的编制模块、与其他元数据格式的转换模块等。一般地，MR将建

立一个数据模型来描述这些登记对象间的关系，并将这些关系通过数据库表关联或规范链接体现出来。另一方面，元数据内容关联可以是涉及不同元数据格式或应用格式间的内容，例如同一个名称元素可能出现在不同格式中，同类资源类型可能被不同格式以不同方式描述等等，支持人们查询“这个元素在其他元数据格式中是什么意思”、“有什么元数据格式用什么方法来描述这种资源”、“这个格式的这种元素与另一格式的某个元素是什么关系”等问题。这些关联可以通过数据库检索来表现，也可以通过基于RDF/RDFS^[36]的描述来体现。

为了支持人们对元数据定义与应用中的复杂关系的查询，所以提供对元数据规范内容的各种关联关系的描述和查询功能是非常重要的，而且这种功能应该支持对不可预见的关联关系的处理，这也正是当前 MR 发展的重要问题。

3. 元数据登记系统的技术结构

3.1 元数据登记系统的标准模型

ISO/IEC 11179 是数据元素描述和标准化 (Specification and Standardization of Data Elements) 的国际标准，指导各类登记系统对数据库和文件中数据元素的含义 (meaning) 和表述 (representations) 进行确切的描述，支持不同数据系统间唯一地标识、登记和转换数据元素。因此它为各类 MR 提供了一个基本框架。ISO/IEC 11179 分为六个部分：

(1) Part 1^[21]: Framework for the Specification and Standardization of Data Elements (数据元素规范说明和标准化的框架)，为数据元素概念提供一个总体描述框架。按照它提供的数据元素模型 (图 1)，一个数据元素由三部分组成：

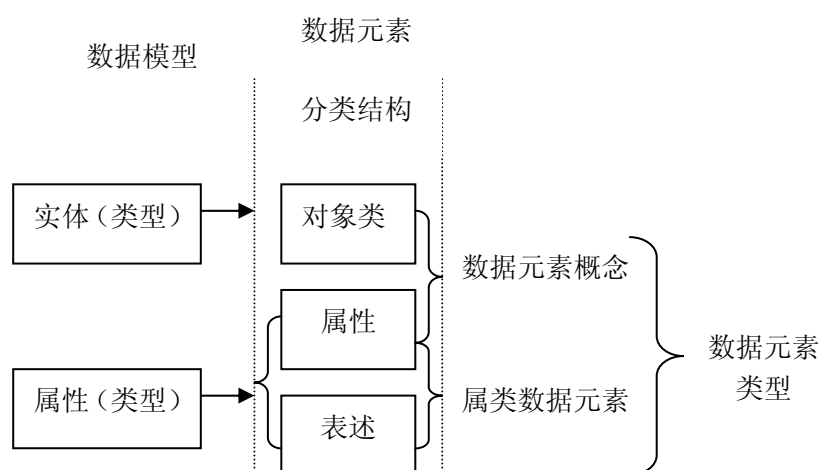


图 1 ISO/IEC 11179 数据元素模型

对象类 (Object Class)，关于客观世界中一类事物、概念或过程 (统称对象) 的抽象描述，同类对象属性和行为基本相同。

属性 (Property), 对象类所有成员的共有特性, 用以描述或区分对象的特质。

表述 (Representation), 数据元素允许的数据类型 (Data Type)、值域 (Value Domain)、测量单位 (Unit) 等。

对象类和属性的组合为数据元素概念 (Data Element Concept, DEC), 例如家庭收入, 可以有多种可能的表述方式 (例如按等级或按数额)。从这个意义上讲, 一个数据元素可以看成由 DEC 和表述方式组成。

(2) Part 2^[22]: Rules and Guidelines for Classifying Metadata (数据元素依据的分类体系), 描述如何把数据元素的各个组分 (对象类、属性、数据类型、值域、测量单位等) 与有关的分类体系或词表 (或称内容编码体系) 联系起来。ISO/IEC 11179 鼓励人们从成熟的编码体系中选择数据元素的组分对象, 从而保障数据元素定义的语义清晰性、逻辑一致性、属性或表述的继承性, 提高数据元素定义的效率和互操作性。这些编码体系包括关键词 (Keyword)、叙词 (Thesaurus)、分类与概念集 (Taxonomy/Ontology)。当从某个编码体系中选择某个词汇或类别来描述数据元素的某个组分时, 需要标明他们之间的关系, 因此该部分具体规定了这种关系的描述方法。

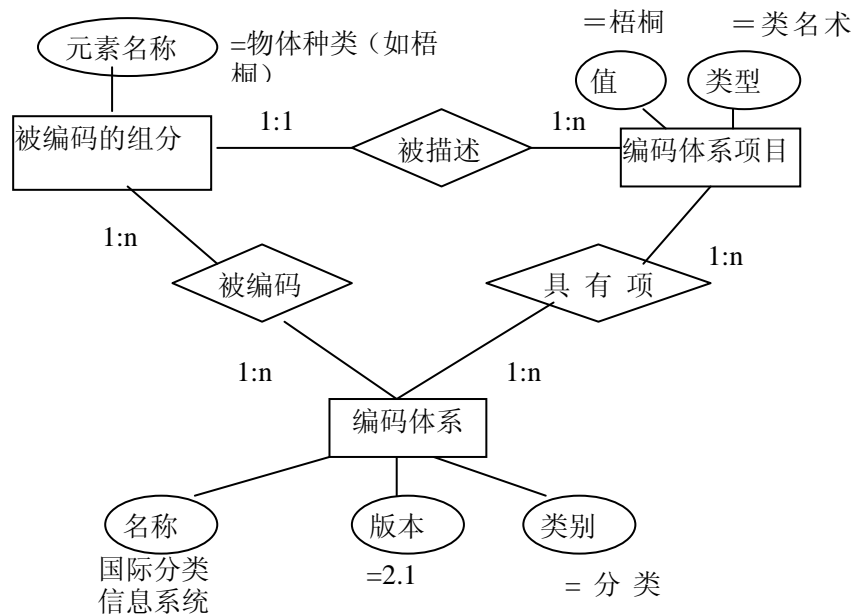


图 2 ISO/IEC 11179 数据元素及其相关关系

在逻辑上 (图 2), 标明被描述的组分 (Classified component)、编码体系类别 (Classification scheme type)、编码体系名称 (Classification scheme name)、编码体系版本 (Classification scheme version)、编码体系项目类别 (Classification scheme item type) 和编码体系项目值 (Classification scheme item value)。由于与数据元素相关的编码体系及其具体项目都被记录, 通过链接和浏览相关的分类结构, 可以支持对数据元素及其相关关系的查询和验证。

(3) Part 3^[23]: Registry Metamodel and Basic Attribute (元数据登记的元模型和基本属性)。metamodel元模型, 是描述模型的模型。一个元模型提供一个描述特定模型的结构和内容元素的机制。ISO/IEC 11179-3: 2003 就提供了一个描述 metadata registry 元数据登记 (简称MR) 的基本结构的元模型, 它的作用就是提供元数据登记系统中统一的概念、条目、值域和值义, 从而支持数据元素的共享和复用。

ISO/IEC 11179-3 定义了 MR 结构的六个部分, 包括 administration and identification region (数据元素管理责任机构与标识), 提供提交到登记系统的数据元素的标识、登记、提交机构、管理机构、责任机构以及相关机构的联系方式; naming and definition region (数据元素命名与定义) 负责管理 Administered_item (管理元素) 的命名和定义, 每一个管理元素都是在一个或多个 context 环境中命名和定义的, 这个环境就是数据元素的含义范围, 可以是一个信息系统、一个数据库或一个标准文档; in classification region (数据元素分类体系) 中, 数据元素可被零个或多个分类表划分归类, 分类表可以是分类法、关联网络、概念集或其他的术语系统, 它可以揭示数据元素的属性和关系, 利于对数据元素进行标识、命名和定义; data element concepts region (数据元素概念) 提供数据元素的概念域信息; conceptual and value domains region (概念域与取值域), 从某种意义上可看作是数据元素的逻辑编码集和物理编码集, 前者是支持 data element concept 数据元素概念, 后者支持 data element 数据元素; data elements region (数据元素) 则提供对数据元素的管理。

ISO/IEC11179-3 定义的一个完整的元数据登记系统的元数据模型, 如图 3 所示:

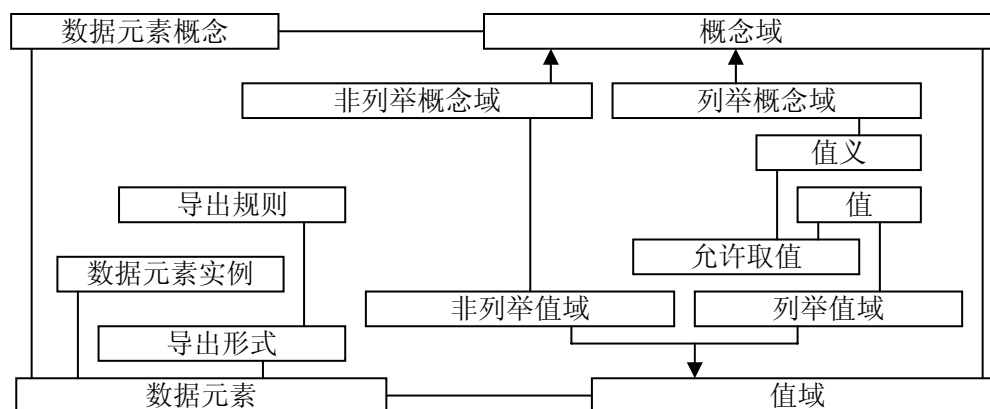


图 3 ISO/IEC11179-3:MR 模型

(4) Part 4^[24]: Rules and Guidelines for the Formulation of Definitions (数据元素定义形成的规则和指南), 包括一个数据定义规则以及一个指导原则: 前者指出一个数据定义必须是唯一的, 要用一个描述的短语或句子独立地声明该概念是何物, 包含一般的可理解的缩略语, 表达时不得嵌入其他数据元素或底层概

念的定义。后者建议数据定义应该清晰简要地声明概念的基本含义，能独立存在，表述时不能嵌入推理性、功能性用途，要避免循环推理，对相关定义使用一致的术语体系和结构。

(5) Part 5^[25]: Naming and Identifying Principles for metadata (数据元素的命名和标识原则)，为数据元素命名和设计非智能标识提供特定的规则和指南。对数据元素的命名与识别可以用5个具体属性来完成：name、context、registration authority identifier、data identifier、version identifier，其中name和context是成对出现的，一个数据元素在一个应用环境(context)中至少应有一个名字(Name)。每个元素在一个登记机构中应有一个唯一标识符，由registration authority identifier、data identifier、version identifier联合构成。

(6) Part 6^[26]: Rules and Guidelines for Registering Metadata (数据元素的登记)，描述了数据元素登记管理机构的功能和规则。一个登记人通过一个权威登记系统为自己的数据元素申请一个有效(Valid)的唯一标识符，从而实现该元素在该登记系统的登记。在登记过程中，数据元素登记可能经历若干登记阶段，例如：未完成(Incomplete)，临时登记，可能变化；被记载(Recorded)，正式申请登记的元素；被确认(Certified)，经过验证，符合登记的技术和管理要求；标准化(Standardized)，按照有关标准规范化处理登记的数据元素。数据元素的登记应该是非排外的，即每一个组织都可成为一个登记机构；应该是数据共享登记，即数据可以为内部或外部组织共享；应该是实用驱动，即实用性决定了元素登记的时效和有用性；应该是灵活的登记，即元素可以在不同等级登记。

许多组织将该标准作为他们的数据元素登记系统的标准，如美国的环境数据登记系统^[36]、健康信息库^[37]、人口普查团体的元数据库^[38]等。

3.2 人工登记与检索的范例系统

现有的基于人工登记和检索的MR，登记元数据的定义信息、元数据的内容编码体系、元数据描述信息及关联形式等，若某一系统要在该MR登记自己的元数据规范，可通过与系统管理员联系，或填写MR系统提供的上载模板来登记和发布元数据。MR通过一定的组织机制对登记者的元数据规范进行组织管理，并提供对登记信息的浏览。同时事先人为的分析资源系统、解析其元数据格式、定义不同元数据名称域之间的关系、定义元数据转换规则，以及MR的给定的检索界面、检索方法和检索参数，进行元数据的检索和转换。这里我们要谈到的元数据登记范例系统包括ROADS Metadata Registry、DESIRE Metadata Registry、Cores Metadata Schema Registry^[39]和DCMI Metadata Registry。

3.2.1 ROADS Metadata Registry[14]

ROADS (Resource Organisation And Discovery in Subject-based services, 主

题服务系统中的资源组织和发现)项目,致力于开发主题网关软件包、研究跨网关检索和互操作的方法、开发建设特定主题领域的资源目录。ROADS 系统建立了一系列主题网关,例如 ADAM、Biz/ed、EELS、Finish VL Project、History、NOVAGate、OMNI、SOSIG。

这些主题信息网关使用专门的元数据模板(Templates)来描述所标引的网络资源,不同的主题网关可能使用不同的元数据模板来描述相同或相似的资源类型。ROADS Registry 就是致力于记录和维护主题网关所采用的元数据模板、模板元素以及相应的目录化指南(Cataloguing Guidelines)。如果ROADS 系统用户需要建立新的元数据模板,必须将此模板在ROADS Registry 上登记;用户如果要在现有的模板中增加新的元素,也必须在ROADS Registry 上登记;用户可以通过登记系统来复用有关模板、或者复用模板中的某些数据元素;用户也可以通过对各个模板的查询、解析来支持跨网关检索中的元数据模板转换。

ROADS Registry 结构较为简单,只支持人工登记和人工查询。它登记了ROADS 可以支持的所有模板类型,共有16种,包括资源集合(COLLECTION)、数据集(DATASET)、文件(DOCUMENT)、都柏林核心集(DUBLINCORE)、图像(IMAGE)、项目(PROJECT)、资源(RESOURCE)等。在登记各个模板时,系统完整地记载了模板的有关属性(Attribute),包括 Template-Type(模板类型)、Template-Version(模板版本)、Title(标题)、Subject(主题)、Description(模板描述)、Owner(模板制定者或拥有者)、Date(日期)、Language(语言)、Source(来源)等等,以及采用相应模板的主题网关的信息,包括主题网关名称、联系方式、搜索范例、URL 以及简介。不同模板类型中包含相同元素,ROADS Registry 对这些元素的定义分类集中给出,例如机构(Organization)、个人(User)、代理(Agent)和登录控制(Access),实际上形成四个相对独立的元素组,供各个模板根据自己需要复用这些元素组中的相关元素。

3.2.2 DESIRE Metadata Registry[16]

DESIRE 元数据登记系统(DMR)是英国 JISC 支持、UKOLN 具体实施,提供关于跨命名域的元数据格式的定义信息。它登记了22个名称域(Namespace)、35个登记机构(Registration Authority)、16个基本语义单位(Semantic Units)、517个元数据元素(Elements)、14个数据元素表述级(Representation Classes)、106个 Schemes、17个应用规范(Application Profile)和19个词汇表(Glossary)。

(1) DMR 的基本模型

DMR以ISO/IEC 11179为基础,它定义了一个数据模型(Dada Model),全面描述各类元数据规范对象。在这个数据模型(图4)中,最基本的对象是命名域(Namespace),作为唯一地定义元数据元素的限定空间,例如DC命名域^[34]、

LOM命名域^[35]等。每个命名域定义了一组数据元素（Elements），它们在特定的命名域内具有唯一的含义，例如DC命名域的Title和INDECS命名域的Title就表达不同的语义。一个命名域的元素集合构成元数据词汇（Metadata Vocabulary），提供了描述某种内容对象的词汇集合。一个应用规范（Application Profile）往往针对特定应用领域或应用系统的需要，从多个命名域中选取元素来构成新的元素集合，形成新的元数据词汇。为了规范元素的内容描述，往往针对命名域的元素推荐内容编码体系（Schemes）或针对应用规范的元素规定所采用的内容编码体系。

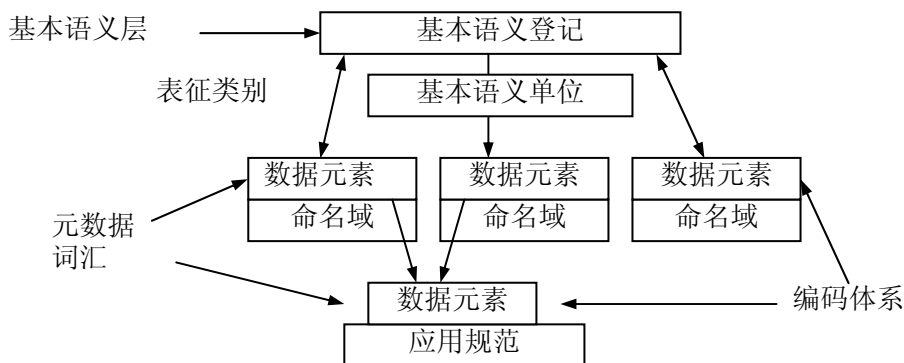


图 4 DMR 数据模型

当多个命名域的元素进行转换时，可以两两建立转换关系，但这种方法成本过高。DMR 采用了另一种方法，建立了一个基本语义层（Semantic Layer），通过基本语义登记来提供一系列唯一的概念，作为所有命名域的基准概念空间。与这个基本语义层相连的是基本语义单位（Basic Semantic Units），提供对概念进行具体表示时的表征类别（Representation Class），例如名称、文本、代码等，从而将一个基本语义层的概念具体地表示用某个表征类别表示的元素。基本语义层的主要作用是提供一个公共转换工具。当要将命名域 A 的元素 X 转换到命名域 B 时，首先根据命名域 A 与基本语义层的转换关系，将 X 转换为基本语义层的对应基本概念 Y，然后根据基本语义层与命名域 B 的转换关系将 Y 转换为命名域 B 的元素 Z。这样，每个命名域都只需要建立与基本语义层的转换关系。当然，这种方法的转换精度可能有局限，因为基本语义层的概念定义必须非常精细才能保证满足各个命名域可能包含的复杂语义情况，但越是精细就越容易产生“测不准”现象。DMR 只有一个语义层，这种方法减少了复杂性。该系统目前只能提供人可以识别的版本，但它也为今后发展机器可理解的版本打下了一定的基础，人机均可识别是这类系统的发展方向。

（2）DMR 的基本功能

A.浏览。根据上述模型，DMR 对各类元数据规范对象进行登记，并可分类浏览。例如，元数据登记机构（Registration Authorities）、命名域概念（Namespace Concepts）、命名域（Namespaces）、语义单元（Semantic Units）、元素（Elements）、

表征类别(Representation Class)、内容编码体系(Schemes)、应用规范(Application Profiles)。

B.检索。DMR 提供对各类登记的元数据规范的分类检索(包括完全匹配检索和包含字段检索)。例如对登记机构(Registration Authorities)可以按 ID、名称 Name 和网址 URL 进行检索,对元素(Elements)可以按其数据类型(Datatype)、ID、名称 Name、链接 URL、定义 Definition 等。

C.登记。需要注意的是,当一个元数据规范在 DMR 进行登记时,需要申明自己的登记机构和命名域,描述自己的元素,建立自己元素与基本语义层概念的关系以及相应的表征类别,提出相关的内容编码体系等等。一个应用规范也可如此登记,只是要说明自己所含元素来自什么命名域等。

D.映射。DMR 的特色就是提供了不同命名域之间的映射。系统给出了 22 个名称域之间的映射。在系统中,进入到映射页面(Generate a Cross-walk),在 from 和 to 两个框中各选一种映射类型,选择确定后系统就自动进行操作。映射的结果分为三栏,左边是 from 框选择类型的所有元素,中间是语义层的元素,右边是 to 框选择类型的元素,整个映射关系就表现出来了。

3.2.3 DCMI Metadata Registry[13]

DCMI (Dublin Core Metadata Initiative) Metadata Registry (简称 DCMI MR)是由 DCMI 组织建立的一个元数据登记系统,其目的就是开发一个提供关于 DCMI 词汇和词汇中术语间关系的权威信息的元数据登记。它基于 RDF Schema 来定义和描述元数据标准、应用规范和相关语义集的系统工具,支持 DC 命名域权威化的元数据规范及其应用管理。

DCMI MR 是基于 RDF 的元数据登记应用。它利用 RDF (资源描述框架)规定了系统中描述特定资源的特定属性的基本数据模型,利用 RDF Schema 定义了 DC 的元素、限制属性和 DC 类型。它建立起了 RDFS 描述的元素关系,从而支持基于 RDFS 描述的检索。例如可以查找元素 X 与元素 Y 的关系 (find x in relation to y)、查找属于特定元素项的类 (find classes containing the following term)、查找元素 X 的细化项 (find refinements for term x)、查找元素 X 已登记的编码体系 (find registered encoding schemes for term x) 等。

DCMI MR 可以浏览和检索登记的 Schemas、类 Classes、属性 Properties、元素 Elements、限制属性 Qualifiers、控制词汇 Controlled Vocabulary Terms、词汇与编码规则 Vocabulary & Encoding Schemes。

DCMI MR 有四大系统功能模块:

查询模块提供检索和浏览功能。用户在“查询”文本框中输入检索式,并选择是否需要检索式中的英文对检索结果中的英文大小写敏感,是否要求检索结果“确切的词组”匹配、“所有项”匹配或者“任意项”匹配。例如我们查询“title”,

并使这次查询结果对检索式中的“title”不区分英文大小写，要求匹配确切的词组。检索结果为 2 条，分别为 DC terms 集合中可选（alternative）属性集合中的“title”（<http://purl.org/dc/terms/alternative>）和 DCES 核心元素集定义中的“title”（<http://purl.org/dc/elements/1.1/title>）。对于“title”的定义有英语、德语，标注有英语和繁体中文，内容描述有英语、意大利语和德语。

特选模块提供用户界面样式、用户界面语言和返回检索结果语言选择的功能。系统提供两种用户界面样式：标准界面和 RDF 界面，两个界面的不同就在于标识检索结果的标号和支持的检索类型的不同。前者是默认的用户界面，不要用户一定要了解 RDF，为匹配的检索结果标识出“题名”、“定义”、“描述”、“标识符”等。RDF 界面是面向 RDF 专家的，它不仅能检索登记系统中的不同元素集中的元素信息，而且能检索它们的 RDF schema 文档，标识出检索结果的 RDF 标记，如“标识符”的 RDF 标记“rdfs:label”、“定义”的 RDF 标记“rdfs:comment”、“描述”的 RDF 标记“dc:description”。

管理模块提供对系统的管理。其中载入 RDF 数据模板（Import RDF）的作用是将 RDF 数据装载入登记系统的数据库中。可以直接加载 Web 上的文件（利用 URL），也可以浏览本地 RDF 文件将文件路径载入模板的文件栏中，同样可以加载一个目录中的所有文件将目录路径放入载入模板的目录栏中，然后选择目标模式（target model）登记（Registry）提交即可。它的资源管理模板（resource management），列出了系统已有的资源 URL 列表，点击其中一个资源，可以查看该资源已有的多种语言的描述。根据需要可以删除、更新现有的描述，也可选择国家代码（RFC 3066）来增加某种语言的资源描述。它的语言支持模板（language support）支持管理员依据国家代码（RFC 3066）和系统需要来选择更新、删除和增加用户界面和元数据表述的语言。属性编辑（property editor）模板，主要为程序员提供 Java 属性文件的编辑，将属性文件的路径粘贴在模板的文件栏中提交即可。登记与管理远程服务（registry and manage remote service）模板，用于把系统客户端和远程服务连接起来供登记系统发现和查询远程服务，只需将远程服务的 WSIL 文档粘贴在模板的文件栏中提交即可；执行远程服务（execute remote service）模板，列出系统中的远程服务的 URL 列表，点击 URL 可查看远程服务描述文档，目前系统中还没有任何远程服务载入。

帮助模块为用户提供系统使用的帮助信息，帮助了解系统四个功能模块的作用和使用，并辅以例子和图示说明。系统也对多语种环境下的信息的表示做了解释。帮助模块还为开发人员给出了 DCMI MR 系统的应用接口的各个部分做了说明。

3.2.4 Cores Metadata Registry^[39]

与 DMR 密切相关的是 Cores Metadata Schema Registry(简称 Cores Registry)

项目是 UKOLN 组织建立的致力于创建一个统一的共享的数据模型来发布和发现数据元素集 (element sets) 和应用规范集 (application profiles)。

Schema 是标准组织制定的用于定义元数据项并以一个命名域发布的工具。从某一或某些命名域选取部分元数据项组合起来用于某一特定应用的 Schema 称为应用规范 (Application Profile)。从一定角度讲, 元数据登记系统存储和管理的信息就是不同类型的 Schema, 那么获取机器可读 Schema 对于登记系统自动发布信息 and 登记系统间互相发现和共享信息等应用很关键。

Cores Registry 系统(见图 5)由 RDF 数据库(RDF database)、Web 浏览器(www browser)、登记模块 (registry)、schema 创建工具 (schema creation tool^[40]) 和一系列应用程序接口 (APIs) 组成。系统提供一个用户可读界面来浏览 schema 的内容, 提供一个应用程序接口 (API) 来上传和下载 schema。Schema 创建工具通过请求 (query) API 和上传 (upload) API 与登记系统交互, 它不要求应用者一定是 RDF 专家或者熟悉 XML 句法, 利用 schema 创建工具就可以创建 schema, 并以 RDF/XML 形式在 RDF 数据库存储 schema, 然后提交给登记服务器。

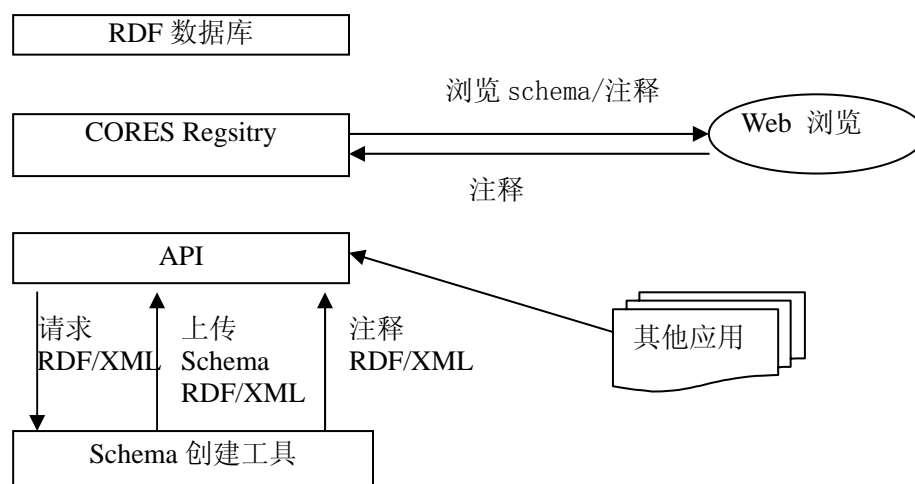


图 5 CORES Registry 系统框架

CORES Registry 的核心模块之一就是 schema 创建工具, 它是用 Java 语言写成的能用于多种系统平台, 以 Java Swing 工具做的图形化用户界面, 以 Jena RDF toolkit 支持 RDF, 客户端与登记系统的交互是通过 HTTP 协议, 数据交换格式是 RDF/XML。Schema 创建工具提供了一个简单的用户界面来创建和编辑 RDF Schema, 创建后的 RDF Schema 可以本地存储也可以提交到登记系统 (需要是验证用户)。它不仅能够定义数据元素、应用规范和编码规则 (encoding scheme), 而且还通过它检索窗口查询登记系统并提供检索出来的数据元素、编码规则为其他应用所复用, 而这种复用只需要通过拖一放操作到目的应用集即可实现。

3.3 基于自动发布和发现的 MR

基于人工登记和检索的MR对元数据发现的传统方法是人工地分析各个资源系统，解析其元数据格式，从而支持元数据的检索与转换。而MR的发展趋势是支持元数据规范的自动发布与发现。这样的MR在元数据登记上通过登记API（应用程序接口）来支持自动登记，元数据维护机构的系统可以直接利用该API上载要登记的有关的元数据定义信息；在数据管理上，以一种有效的MR数据模型将登记元数据规范、登记元数据描述信息、登记元数据应用信息等进行登记管理，并通过链接等机制将他们之间的关系组织好；在检索与发布上，通过Web服务器实现浏览与查询，支持多种检索，尤其是基于关联关系的检索；MR还要进行自我描述，并在公共服务登记系统登记，支持智能代理、第三方系统对自己的发现、配置和检索；对元数据的自动解析与转换的支持，通过MR数据库所记载的关联关系自动揭示、链接、调用，如在跨命名域MR中，可以利用各个命名域的元素字典、元素语义网络（Semantic Web）、概念集（Ontology）等，进一步追寻元素语义定义和语义关系，通过查询描述同一概念或同一对象的元素名称，分析它们的语义对应程度，确定这些元素的转换关系；可登记和查询元数据格式间的转换关系或转换模板，支持应用系统直接复用有关转换模块，可以通过关联元素的语义定义链，利用共同的根概念集（Root Ontology）自动建立不同命名域中元素的语义关联，从而辅助构建元素转换关系。这里我们要谈到的基于自动发布和发现的元数据登记范例系统包括ebXML Registry Services^[41]、UDDI Registry^[42]和GoXML Registry^[43]。

3.3.1 ebXML Registry Services[41]

ebXML 是基于 XML 的开放环境下电子商务机制，ebXML Registry 的基本机制由图 3（摘自 ebXML Registry 网站）描述。如图 6 所示，系统 A 为实现特定商务功能，向 ebXML 登记系统查询并请求特定商务模式（及其过程、交换信息、文件等）的标记文件，利用这些标记文件建立符合 ebXML 规范的商务系统，然后将自己的商务意向（Business Profile）提交给 ebXML 登记系统，向外界公布；系统 B 通过查询 ebXML 登记系统，发现符合自己商务需求的系统 A（通过 A 的商务意向文件），系统 B 就与系统 A 利用各自的商务意向文件进行“谈判”，建立商务合作协议（Business Agreement），双方利用商务合作协议进行具体商务操作。

为了有效实现上述机制，ebXML 定义了一系列协议，包括：

- （1）ebXML Business Process Specification Schema（BPS）商务流程定义；
- （2）ebXML Collaboration Protocol Profile 商务要约（CPP）；
- （3）ebXML Collaboration Protocol Agreement 合作协议协定（CPA）；

(4) ebXML Catalog of Common Business Processes 商务流程目录 (Catalog of CBP);

(5) ebXML Registry Service Specification 登记服务定义。

ebXML Registry 在整个 ebXML 机制中起着核心的作用, 支持分布的电子商务系统进行登记、发现和调用。ebXML 登记系统所登记的对象包括 Core Components (包括 CBP 和基本商务数据元素)、商业文件格式、BPS、CPP、CPA 等。提供商业服务的企业可从登记系统检索有关的商务数据元素、商务过程、商业文件, 利用这些工具构建自己的商务流程 (BPS) 和商务要约 (CPP), 并将它们登记到登记系统。那么需要商业服务的企业可检索登记系统, 发现合适的卖家, 根据卖家的 CPP 制定合适的 CPA, 通过 ebXML 消息服务传送给卖家, 卖家接受建议的 CPA, 双方可依据 CPA 配置各自的系统, 从而建立电子商务互操作环境。

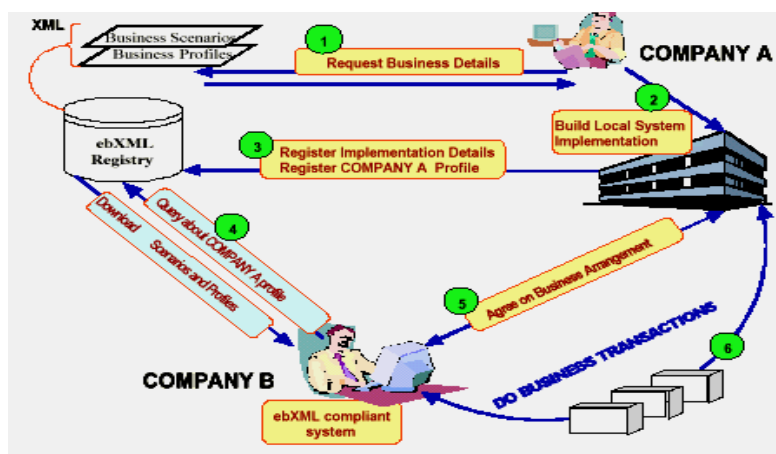


图 6 ebXML 基本机制

3.3.2 UDDI Registry^[42]

UDDI (Universal Description, Discovery, and Integration通用描述、发现与集成^[44]) 是Web Service体系中的登记服务机制。UDDI提供了一种基于分布式的商业登记中心机制, 可以维护企业及其服务的全球目录。它采用XML方式描述企业及其服务。从概念上来说, UDDI商业登记所提供的信息包含三个部分: “白页” (White Page), 包括了企业地址、联系方式、已知企业标识、所提供的商业服务等; “黄页” (Yellow page), 包括了基于标准分类法的企业行业类别及按照类别列出的企业; “绿页” (Green Page) 则包括了关于该企业所提供的Web商业服务的技术信息, 其具体形式可能是指向详细技术文件的链接 (例如URL)。

UDDI主要定义了一套建立UDDI登记系统的规范, 其中主要涉及内部信息模型和外部接口规范, 人们可采用UDDI规范来建立自己的登记系统。UDDI内部信息模型定义有关的登记信息结构, 由XML Schema描述, 从而支持丰富的数据类型、开放描述方式及其按信息模型对数据进行验证的能力。这些信息在传递时被

封装在SOAP^[45]包里，从而可利用HTTP协议传递，并可利用XML-capable^[46]系统进行解析。UDDI外部接口定义了有关API，建立了信息交互框架，提供发布和检索各种Web服务描述信息的能力。API也是基于XML Schema的，所以可以被方便地识别和应用。UDDI登记系统可以是逻辑上集中、物理上分布的，由多个根节点组成，相互之间按一定规则进行数据同步。当一个企业在一个UDDI登记系统注册后，其登记信息会被自动复制到其它UDDI根节点，于是可以实现“一次注册、分布发现”。

UDDI 所登记的信息（图 7）主要包括：

商业实体信息 **businessEntity**，这可以被看成是一个关于商业机构的标准描述元数据，作为发布和发现的核心，包括 **businessKey**（商业实体标识）、**name**（名称）、**description**（描述）、**businessServices**（商业服务集）、**identifierBag**（企业或商品标识集）、**categoryBag**（分类集，包括行业代码、产品代码，地理位置代码和商业标识代码）等元素。

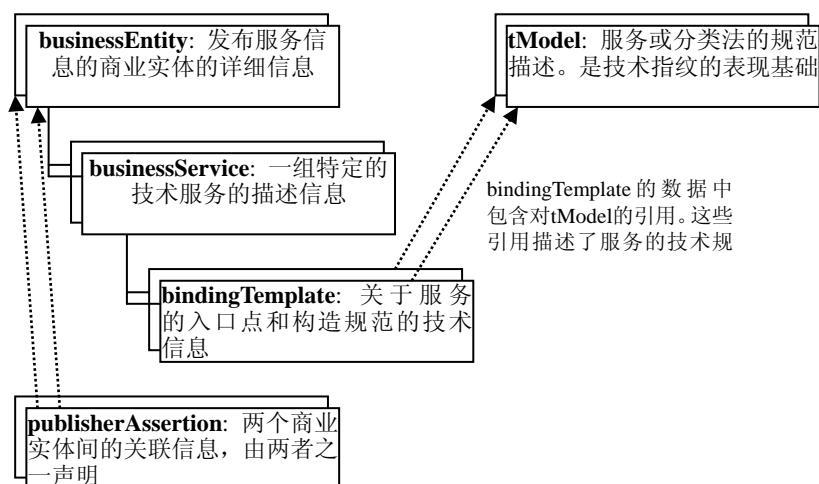


图 7 UDDI 登记信息

商业服务信息 **businessService**，对 Web 服务进行商业业务描述，是商业实体的子结构，一个商业实体可以有多个商业服务。商业服务也是元数据容器，通过 **serviceKey**（商业服务标识）、**businessKey**（商业实体标识）、**name**（名称）、**description**（描述）、**bindingTemplates**（绑定模板）、**categoryBag**（分类集）等元素，对特定企业所提供的特定服务进行描述。这些商业服务可以加以分类，使其可以按不同的行业、产品、服务类型、地域等来划分。

技术绑定信息 **bindingTemplate**，对于每一个商业服务，存在一个或多个技术实现方式，这些方式的技术描述通过技术绑定信息来实现，包括应用程序连接远程 Web 服务并与之通讯所必须的信息，例如 **bindingKey**（绑定标识）、**serviceKey**（服务标识）、**description**（描述）、**accessPoint**（接入地址，提供商业服务的远程服务器地址）等。

调用规范信息 tModel，它提供一种引用机制，包括服务名称、发布服务的组织以及指向有关技术规范的 URL 指针（UDDI 定义了一个利用 URL 绑定 Web 服务的框架，使每个企业能够集中地维护自己的调用规范）。技术规范包括信息格式、响应机制、传输协议、安全机制等。这个引用信息可被看作是提供这项服务的公司的承诺，承诺他们已经实现了一个与所引用的调用规范相兼容的服务。这个引用信息被嵌入到相应的技术绑定信息中，从而支持第三方系统利用这些调用规范来了解技术要求和配置系统。

关联声明 publisherAssertaion，用以描述两个商务实体之间的关联信息。

3.3.3 GoXML Registry[43]

GoXML Registry是Golbal XML^[47]的登记模块，主要支持跨Web Services和 ebXML Registry的登记系统机制。它包括四个主要部分:Registry engine（登记引擎），用于商务服务、过程、消息和词汇的分类、存储和管理；Repository（存储库），用于存储和检索登记服务所需信息；Web-based Registry Client（基于web的登记客户端），用于登记系统的结构配置、内容管理、浏览和查询；A Registry Services API（登记服务API），对登记引擎和ebXML登记服务提供程序化控制。在GoXML Registry的结构（图 8）中，它的登记客户端和登记服务器由Web服务接口和ebXML接口相连。

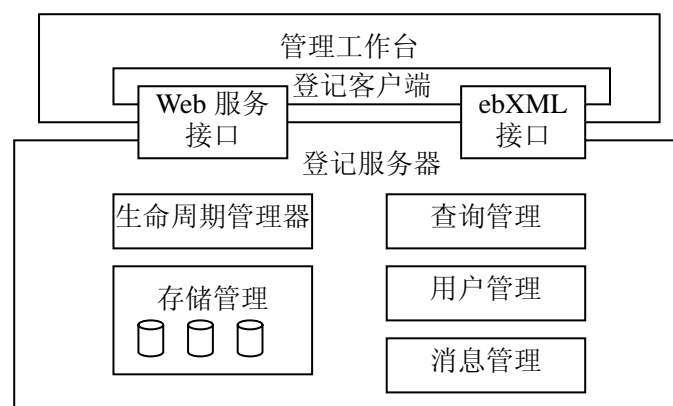


图 8 GoXML Registry 结构

登记服务器要素包括：生命周期管理，管理登记对象的整个生命周期，从登记项的提交、分类、联系、驳回和迁移；用户管理，管理注册用户登记、授权和认证；查询管理，为相关对象间的浏览检索提供发现和组织功能；存储管理，控制登记对象的物理存储和安全保护，主要是通用的关系型数据库；消息管理，提供了 ebXML 消息的安全可靠传递；管理工作台，基于网络的管理机制，作为登记服务器的客户，通过启动客户端的有关工作方式进行管理；登记服务接口，支持三种登记系统接口类型（Web Services、ebXML、JAXR^[39]），支持在 .NET 和 Java 环境下的 B2B 企业对企业 and A2A 应用对应用的应用集成。Windows

98/NT/2000/ME/XP、Solaris、Linux 都宣称支持GoXML Registry。

4. 元数据登记系统的发展趋势

元数据登记系统应用在数字图书馆、开放档案信息系统、门户网站、电子商务、数字科研、权限管理等方面。在数字图书馆服务体系中，元数据登记系统绝大部分是需要人工来识别元数据进行登记操作的，以 DCMI MR 为代表的元数据登记系统发展到 v3.0.0 版本，也只能部分支持元数据自动发现功能（WSIL 描述文档驻留系统）。要支持元数据登记系统在分布的数字信息环境下应用，必须对元数据进行有效的管理，包括元数据交换、元数据共享与复用、使用统一工具，依据元数据登记标准模型，并在此基础上进行元数据扩展及开发满足需求的本地化应用。那么自动发现和复用元数据，与更多更广泛应用领域的登记系统交互和共享数据，支持自动发布和共享元数据成为了数字图书馆元数据登记系统的发展趋势。

4.1 依据标准模型

现存的绝大多数 MR 都不是基于 ISO/IEC 11179，所以不同 MR 间的共享和协作显得困难重重。例如系统间进行元数据交换时，要么自行设计一种中介数据模型作为交换模板，要么借助其他交换格式却在一定程度上难以满足交换双方的元数据要求。而在开放信息环境下，一个 MR 应支持跨系统和跨命名域的元数据获取和共享。重要的是，一个 MR 必须给予用户系统中所登记数据的可理解的含义、属性、标识符等来支持数据共享。我们不主张一个元数据登记系统就是一个包括所有在系统中存储的可找到的元数据的数据库。希望能有一个标准的结构化的工具或模型来帮助构建一个支持元数据定义、命名、语义解析、分类聚合及属性管理的系统，ISO/IEC 11179 就提供了这样的模型。

在 ISO/IEC 11179 中，定义了元数据的基本属性、语义关系、分类体系和元数据所有者信息，以结构化的方式将元数据组织成了精确的可靠的可检验的数据记录，使元数据用户（包括其他第三方系统）能够对元数据的含义、表述、使用有一个共同的理解和认识。那么在它们的系统之间可以建立不同元数据元素集合的交换和参照、不同概念集的内容定义和参照、不同元数据应用规范的数据模型、不同元数据交换模板的设计和说明等。

4.2 基于 Web Service 的元数据登记系统

元数据如何登记到系统以及让其他系统或机构查找到它并加以利用是一个很重要的问题。前面我们所讨论的基于自动登记和检索的 MR 都是通过应用程序接口 API 进行元数据发布和查询，支持元数据的自动关联与循证，系统之间可

通过链接互相调用，形成了分布式元数据登记体系。在这样的分布系统中，每一个元数据登记系统自身也是分布式登记系统，要在分布服务登记系统登记，使用 UDDI 进行说明，并通过唯一标识符进行解析。同时在分布式服务登记系统中，还存在集成定制、整合检索、使用控制、服务评鉴等系统，那么还存在内部登记系统间的互操作问题，元数据登记系统对自身的描述说明将支持其他系统对它的开放搜寻和调用。

Web Service 分布服务体系（图 9）就提供这样分布服务登记机制，其中分布的信息系统及其具体功能、功能组都被视为服务系统。它们以规范的 XML 消息传递方式（例如 SOAP 协议）通过远程程序调用或数据单元传递进行交互。这些服务系统可以作为服务请求者调用其它服务系统，也可作为服务提供者被其它服务系统调用。利用基于 XML 的服务描述语言（例如 WSDL），对这些服务系统的操作类型、输入输出数据流、数据类型、与传输协议和数据格式及安全控制机制等的捆绑方式、网络位置等进行规范描述，描述结果是一个规则的 XML 文件，成为被描述服务系统的标准界面。为了支持公共登记和开放发现，Web Service 体系利用了 UDDI 机制来进行 WEB 服务描述文件的登记、管理和检索服务。

同样元数据登记系统也是服务系统，可以通过多种形式（WSDL、WSIL 网络服务检查语言^[48]）将 WEB 服务描述文件进行发布（Publish），使其他系统发现、解析、获取元数据登记系统中的元数据规范信息，支持元数据定义与应用中的复杂关系的的查询。即使是元数据编制模块、转换模块等软件工具，也可以用 Web Service 方式（例如 WSDL）进行描述。

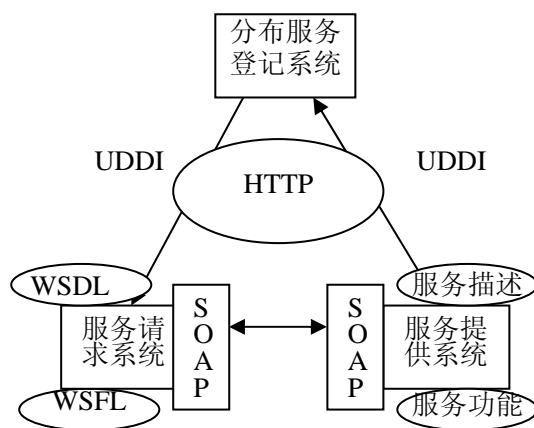


图 9 web service 分布服务体系

不仅仅是 Web Service 支持系统间元数据发现和共享，XML 这种结构化标记语言，可以通过标记数据元素和属性来建立与元数据的映射关系，同样可以获取描述资源、服务、系统等元数据。类似于我们讨论的范例系统 ebXML Registry，它虽然不涉及数字对象描述元数据，但如果我们将关于基本商务过程（CBP）、商务流程（BPS）、商务要约（CPP）和商务协议（CPA）的描述也看成元数据的话，ebXML 登记系统同样是 MR，而且由于它是基于 XML 进行自身定义、支持”

“元数据”自动登记和检索，对于一般 MR 系统设计也有重要借鉴意义。

参考文献

- [1] 张晓林.元数据研究与应用.北京图书馆出版社, 2002.5
- [2] Bianchi, C. and Petrone, J. Distributed Interoperable Metadata Registry. D-Lib Magazine, V.7 (12), Dec., 2001
<http://www.dlib.org/dlib/december01/blanchi/12blanchi.html>, 检索日期 2004.6.02
- [3] Heery, R. and Wagner, H. A Metadata Registry for the Semantic Web. D-Lib Magazine, V.8 (5), May, 2002.<http://www.dlib.org/dlib/may02/wagner/05wagner.html>, 检索日期 2004.6.02
- [4] Neuroth, H. and Koch, T. Cross-browsing and cross-searching in a distributed network of subject gateways: Architecture, data model, and classification. 19 May 2001.
<http://www.stk.cz/elag2001/Papers/HeikeNeuroth/HeikeNeuroth.html>, 检索日期 2004.6.02
- [5] 张晓林.数字图书馆建设中的开放描述技术.现代图书情报技术, 2002.3
- [6] Breiteneder, C. J. et al. Metadata Mining in Legacy Data Sets.
<http://www.computer.org/conferences/meta96/breiteneder/IEEE.fmk.html>, 检索日期 2004.6.02
- [7] Namespaces in XML, W3C Recommendation, Jan. 11, 1999.
<http://www.w3.org/TR/REC-xml-names/>, 检索日期 2004.6.02
- [8] Extensible Markup Language (XML) 1.0 (Second Edition) W3C Recommendation Oct. 6 2000
<http://www.w3.org/TR/2000/REC-xml-20001006>, 检索日期 2004.6.02
- [9] XML Schema. <http://www.w3.org/XML/Schema>, 检索日期 2004.6.02
- [10] Chemical Markup Language Site. <http://www.xml-cml.org/>, 检索日期 2004.6.02
- [11] Mathematical Markup Language (MathML) Version 2.0. W3C Recommendation 21 February 2001.<http://www.w3.org/TR/MathML2/>, 检索日期 2004.6.02
- [12] NSDL Metadata Primer. Technical Issues. NSDL Metadata Repository.
<http://metamanagement.comm.nslib.org/outline.html>, 检索日期 2004.6.02
- [13] DCMI Open Metadata Registry Prototypes.
<http://wip.dublincore.org:8080/registry/Registry>, 检索日期 2004.6.02
- [14] The ROADS Metadata Registry.
<http://www.ukoln.ac.uk/metadata/roads/templates/>, 检索日期 2004.6.02
- [15] German Metadata Registry. 检索日期 2004.6.02
<http://www.mpib-berlin.mpg.de/dok/metadata/gmr/gmr1e.htm>
- [16] DESIRE Metadata Registry.<http://desire.ukoln.ac.uk/registry/>, 检索日期 2004.6.02
- [17] Schemas Registry. <http://www.schemas-forum.org/registry/>, 检索日期 2004.6.02
- [18] Web Service Description Language. <http://www.w3.org/TR/wsdl>, 检索日期 2004.6.02
- [19] Semantic Web. <http://www.w3.org/2001/sw/>, 检索日期 2004.6.02
- [20] Ontology. <http://www.w3.org/2001/sw/WebOnt/>, 检索日期 2004.6.02
- [21] ISO/IEC 11179-1: 1999. 检索日期 2004.6.02
http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm
- [22] ISO/IEC 11179-2: 2000. 检索日期 2004.6.02
http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm
- [23] ISO/IEC 11179-3: 2003. 检索日期 2004.6.02

- http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm
[24] ISO/IEC 11179-4:1995. 检索日期 2004.6.02
http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm
[25] ISO/IEC 11179-5: 1995. 检索日期 2004.6.02
http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm
[26] ISO/IEC 11179-6: 1997. 检索日期 2004.6.02
http://isotc.iso.ch/livelink/livelink/fetch/2000/2489/Ittf_Home/PubliclyAvailableStandards.htm
[27] Resource Description Framework (RDF)
<http://www.w3.org/RDF/>, 检索日期 2004.6.02
[28] The <indecs> Metadata Framework: Principles, model and data dictionary. June 2000.
<http://www.indecs.org/pdf/framework.pdf>, 检索日期 2004.6.02
[29] NISO Draft Standard. Data Dictionary. Technical Metadata for Digital Still Image. July 5, 2000.
<http://www.niso.org/pdfs/DataDict.pdf>, 检索日期 2004.6.02
[30] UMLS SEMANTIC NETWORK.
<http://www.nlm.nih.gov/research/umls/META3.HTML#s3>, 检索日期 2004.6.02
[31] Requirements for a Web Ontology Language. W3C Working Draft 07 March 2002
<http://www.w3.org/TR/2002/WD-webont-req-20020307/>, 检索日期 2004.6.02
[32] The Ontology Inference Layer OIL。
<http://www.ontoknowledge.org/oil/TR/oil.long.html>, 检索日期 2004.6.02
[33] XSL Transformations (XSLT) Version 1.0. W3C Recommendation 16 November 1999
<http://www.w3.org/TR/xslt>, 检索日期 2004.6.02
[34] Vocabulary Markup Language (VocML) . Dec. 28, 2000.
<http://xml.coverpages.org/vocML.html>, 检索日期 2004.6.02
[35] XML Topic Maps (XTM) 1.0. TopicMaps.Org Specification. Aug. 16, 2001.
<http://www.topicmaps.org/xtm/1.0/>, 检索日期 2004.6.02
[36] Environmental Data Registry. <http://www.epa.gov/edr/>, 检索日期 2004.6.02
[37] United States Health Information Knowledgebase.
<http://www.ushik.org/registry/x/>, 检索日期 2004.6.02
[38] Building a Metadata Repository to Support the 2002 Economic Census.
<http://www.unece.org/stats/documents/2002/03/metis/4.add.1.e.pdf>, 检索日期 2004.6.02
[39] Cores Registry. <http://www.cores-eu.net/registry/>, 检索日期 2004.6.02
[40] Schema Creation Tool. <http://cores.dsd.sztaki.hu/download.html>检索日期 2004.6.02
[41] OASIS/ebXML Registry Information Model V2.5. 检索日期 2004.6.02
<http://www.oasis-open.org/committees/regrep/documents/2.5/specs/ebxml-2.5.pdf>
[42] UDDI Regsitry. 检索日期 2004.6.02
http://uddi.org/taxonomies/UDDI_Registry_tModels.htm
[43] GoXML Registry. <http://www.xmlglobal.com/prod/registry/>, 检索日期 2004.6.02
[44] UDDI Specifications. <http://www.uddi.org/specification.html>, 检索日期 2004.6.02
[45] Simple Object Access Protocol (SOAP) 1.1 W3C Note 08 May 2000.
<http://www.w3.org/TR/soap/>, 检索日期 2004.6.02
[46] XML capable. 检索日期 2004.6.02
<http://wwwcs.upb.de/cs/ag-engels/Papers/2001/LohmannICSE01.pdf>
[47] Gobal XML Web Service Architecture 检索日期 2004.6.02
http://www.getdotnet.com/team/XMLwebservices/gxa_overview.aspx

[48] WSIL (Web Service Inspection Language). 检索日期 2004.6.02
<http://www-106.ibm.com/developerworks/webservices/library/ws-wsilspec.html>