基于MDR2023的元数据值域语义约束注册标准化模型研究*

袁满　何玲通　袁靖舒　李洪欣
（东北石油大学计算机与信息技术学院，大庆 163318）

摘要：元数据注册（Metadata Registry，MDR）是数据治理中元数据精确表达语义的必要前提。通过全面系统地分析国内外的MDR系统，发现国内外MDR更多关注基本数据元素，数据语义约束注册方面的研究缺乏。因此，首先基于ISO/IEC 11179:2023（MDR2023）系列标准提出元数据语义约束外延分类模型，明确元数据语义约束范围，并选取其中的值域语义约束详细研究；其次，基于MDR2023标准提出元数据值域语义约束注册元模型，为元数据语义约束注册提供标准化且完整的注册算法流程，从而为元数据值域语义约束注册提供解决方案；最后，以石油领域著名的POSC标准为需求背景，对其中的值域语义约束进行注册，据此实现石油领域元数据值域语义约束的标准化，验证提出的元数据值域语义约束注册元模型的合理性和可行性。提出的元模型对于其他领域数据治理具有普适性。

关键词：元数据；值域；语义约束；元数据注册；注册元模型；数据语义标准
中图分类号：TP391；G254　DOI：10.3772/j.issn.1673-2286.2024.02.007
引文格式：袁满，何玲通，袁靖舒，等. 基于MDR2023的元数据值域语义约束注册标准化模型研究[J]. 数字图书馆论坛，2024，20（2）：70-81.

数据作为数据治理的核心要素，已成为企业重要资产和基础战略资源，而元数据作为企业数据的“DNA”，其重要性不断显现，随之而来的元数据标准化成为行业的核心议题。在标准化元数据之前，需要制定元数据标准规范。目前我国已颁布的元数据标准有《数据中台 元数据规范》（T/ZAII 035—2022）[1]、《科技平台 元数据标准化基本原则与方法》（GB/T 30522—2014）[2]、《管理元数据规范》（WH/T 52—2012）[3]等。同时，各行业、组织也纷纷推出元数据标准及其实施指南，例如《电子图书元数据规范》（WH/T 65—2014）[4]、《卫生健康信息数据集元数据标准》（WS/T 305—2023）[5]、《基础教育教学资源元数据实施指南》（JY/T 0610—2017）[6]、《信息与文献 文件（档案）管理元数据 第2部分：概念化及实施》（GB/T 26163.2—2023）[7]等。

当前，我国在元数据领域已建立多项标准和指南。但对于部分传统行业而言，元数据的国家标准和行业标准仍处于起步阶段。国内广泛采用的GB/T 18391系列标准因更新滞后，已无法完全适应新技术的发展需求，且在元数据描述、限制和扩展方面缺乏规范性指导。相比之下，国际上的元数据注册（Metadata Registry，MDR）系列标准ISO/IEC 11179:2023（以下简称“MDR2023”）提供了更为先进的注册框架和管理方案[8]。由于国内外的MDR系统大多基于早期MDR标准，几乎未考虑元数据语义约束注册，尤其是在值域语义约束上，而MDR2023标准增加了对语义约束的规

收稿日期：2023-12-08
*本研究得到东北石油大学人才引进科研启动经费资助项目“标准驱动的业务规则知识组织模型研究”（编号：2023KQ17）资助。

定，因此，基于此标准进行值域语义约束的研究，对于推动元数据语义约束注册的发展具有重要意义。

本文提出基于MDR2023的元数据值域语义约束注册元模型（Metadata Value Domain Semantic Constraint Registration Metamodel，MVCRM）及算法。首先采用系统论思想，从元数据语义约束整体及其相关概念间的关系入手，构建元数据语义约束外延分类模型，在此基础上，重点分析元数据值域语义约束，为构建MVCRM奠定基础；其次，提出基于MDR2023的元数据值域语义约束注册流程和规则，并构建元数据值域语义约束注册系统（Metadata Value Domain Semantic Constraint Registration System，MVCRS）；最后，结合石油领域的POSC（Petrotechnical Open Standards Consortium）标准进行应用验证。本文可为数据治理带来3个方面的益处：一是提高数据质量和语义的准确性、一致性，通过元数据值域语义约束的统一注册，定义和管理元数据语义，标准化格式，减少语义歧义；二是采用统一的标准MDR2023对元数据值域约束进行注册，有助于企业实现数据互联互通，打破“数据孤岛”，便于企业使用统一的标准进行数据关联整合和分析；三是促进数据治理效率提升和数据资产发展，通过统一标准、元数据命名和值域范围，降低数据理解沟通成本，使不同人员想法达成一致。最终，使行业和企业受益，使传统企业的数据治理效率显著提升。

1　MDR及相关研究现状

1.1　MDR2023

MDR2023是由国际标准组织（International Organization for Standardization，ISO）和国际电工委员会（International Electrotechnical Commission，IEC）共同制定的一项国际标准，旨在解决数据注册、数据语义理解、数据共享和互操作问题。MDR系列标准可追溯至1994年，目前MDR2023为第四版。以往废止的标准对于MDR2023同样具有重要参考价值，因为MDR系列标准是相互关联的，而不是独立的，MDR2023与其他标准的引用关系如图1所示。此标准

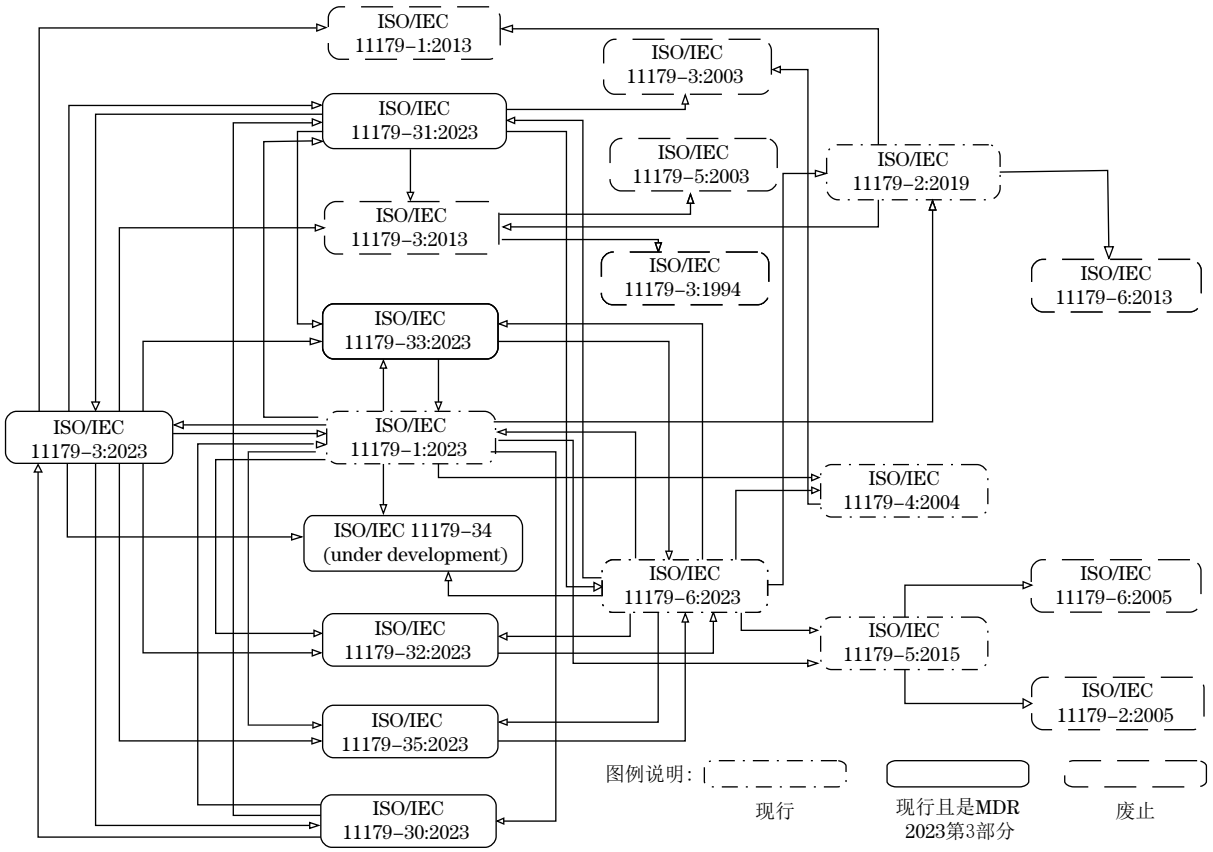

图1　MDR系列标准间引用关系

提供了理解和关联MDR各部分的方法，为元数据和MDR的概念性理解提供基础。其目的是通过精确注册元数据来确保准确理解数据语义。MDR2023不仅能够描述数据语义约束、表达数据和管理元数据信息，还通过严格规范的注册和管理方式，从源头精确控制和表达数据语义，并提高元数据的质量。

MDR2023系列标准共6个部分，其中核心部分是第3部分，该部分主要描述了注册系统的元模型、基本属性以及MDR的基本概念模型。MDR2023取消并取代了ISO/IEC 11179:2013（以下简称“MDR2013”），并在MDR2013的基础上进行修订，主要变化包括：MDR2023在MDR2013的基础上为基本注册的公共设施制定了元模型；为了便于管理，将第3部分划分为多个模块；简化用于描述元模型的统一建模语言（Unified Modeling Language，UML）；增加大量有关于元数据语义约束的内容；重构一些软件包以减少依赖性；在注册项目之间增加通用映射设施[9]。

1.2 国内外MDR研究与应用现状

当前，国外MDR的应用处于领先地位。根据笔者汇总的国外典型MDR系统（见表1）可以看出，MDR主要应用于医学、农业、政府、国防、教育等领域。此外，国外有诸多学者对MDR展开研究。Ulrich等[10]将MDR和GraphQL查询语言结合，并为元数据存储库引入统一的查询接口，以更好地促进元数据语义交换。Kim等[11]创建了4个新的约束来扩展现有的语义类型，以提高临床领域语义的完整性和互操作性。Hegselmann等[12]

表1 国外典型MDR系统汇总

领 域	地 区	机构/组织	名 称	功能（部分）
医学	美国	美国国家医学图书馆	USHIK	实现元数据的注册、维护、查询、映射、浏览，提供医学相关资源，以促进医学领域的知识共享
	美国	美国国家癌症研究所	caDSR	为癌症临床领域提供数据标准管理、元数据管理和标准化工具，以及信息共享服务
	美国	川崎病研究中心	KDDB	致力于收集、储存和分享关于川崎病的相关信息，并提供数据集、数据字典和分析工具
通用	美国	美国国家科学数字图书馆	OMR	发现、创建、访问和管理元数据方案、元数据数据集、模式、应用配置文件、交叉引用和概念映射
	英国	英国图书馆与信息网络办公室	DESIREMR	提供元数据格式的定义信息，以及元数据发现、检索、共享、交换、导入导出、拓展和导航工具
	国际	都柏林核心元数据倡议	DCMI MR	供用户认证、注册、提交、编辑、检索元数据，促进现有元数据的发现、重用、扩展和新词表创建
综合	澳大利亚	澳洲健康及福利研究所	METeOR	提供MDR模板，用于维护数据标准，存储、管理、分发、查看、检索和发布元数据
	英国	英国图书馆与信息网络办公室	CORESR	支持对注册的元数据规范、元素集、元素/属性的修改、添加、删除、审核，实现多领域、多元数据方案注册，促进元数据语义的共享和应用
国防	美国	美国国防部	DoDMR	为国防领域的数据资源提供元数据管理、标准化、检索、应用支持
环境	美国	美国国家环境保护局	EDR	推动机构之间的信息共享，同时提供权威的环境元数据信息
地理	美国	美国联邦地理数据委员会	FGDCCR	提供地理空间标准，推进国家空间数据基础设施建设，实现地理元数据的注册、管理和检索
司法	美国	美国司法部	GJXDM	实现法律信息标准化、数据整合和交互操作以及数据安全和隐私保护，管理公共安全领域数据的元数据模型
政府	美国	美国司法部、美国国土安全部	NIEM	提供统一的语言和框架，推动机构间的数据共享、信息交流和互操作，减少语义障碍
教育	英国	英国联合信息系统委员会	IEMSR	支持用户注册和认证，跨多个元数据标准比较和评估元数据，提供教育领域的元数据标准化方案，专门用于发布、导航和共享英国联合信息系统委员会的标准元数据框架
农业	国际	联合国粮食及农业组织	FAOVESTR	实现农业数据管理、检索和标准化，维护农业领域的信息资源，使农业类词表能够被检索、访问和使用
商务	联合国	联合国贸易促进及电子商务中心、结构化信息标准促进组织	ebXML MR	让全球企业在互联网上进行商务交易，同时让利益相关者共享业务流程整合的信息
图书馆	欧洲	欧洲图书馆	TEL MR	提供包含多个元数据方案的通用元数据模型，用于实现各欧洲国家图书馆和图书馆组织的信息互操作

基于MDR开发了元数据存储库，实现了医疗元数据自下而上的标准化。Arienti等[13]提出了一种针对新冠疫情（COVID-19）的临床环境注册系统的开发方法，实现了对COVID-19人群健康状况的科学评估和分析。Ferenz等[14]为了使能源研究软件更易于查找，按照FAIR原则使用元数据并对其进行描述，实现了更高查找性能的MDR。

尽管国内的MDR研究尚处于发展初期，但国内已涌现一系列MDR平台，如国家科技图书文献中心（National Science and Technology Library，NSTL）的MDR系统、HiTA知识服务平台以及亿信元数据管理平台等。这些平台的建立标志着国内MDR领域的积极进展。虽然与国际先进研究相比，国内研究仍存在一定差距，但国内学者正致力于填补这一鸿沟，不断推进相关研究工作。例如，刘旭[15]在概念系统注册元模型的基础上提出了概念系统注册方法，实现了语义注册与管理。袁靖舒等[16]定义了MDR概念系统向OWL本体映射的规则，验证了MDR概念系统和本体表示的可行性。田中贺等[17]提出了一种基于标识跟踪的数字内容资源注册服务系统，实现了非结构化数字资源的MDR。黄安琪等[18]提出了结构化数据MDR的标准和方法，有效解决了各类常用数据库的注册问题。为了对目前国内外MDR研究现状进行更直观、更全面的展示，在表1的基础上进行扩充，使用分词工具Jieba，从功能、领域和地区3个维度对国内外MDR研究进行分词处理，对分词结果进行词频统计，选取词频排名前50的词汇绘制词云图，如图2所示。

由图2可见，MDR的研究重点分布在医学领域和美国地区。此外，现有MDR系统功能丰富，已实现对元数据、概念、术语、数据元、结构化和非结构化数据等的标准化注册、管理、导入导出、维护、增删改查、检索、互操作和共享等，主要面向元数据和基本数据元素。目前缺乏对元数据语义更深入细致的研究，尤其是对元数据语义约束的研究存在不足。然而，对元数据语义约束的研究对于保证数据质量和数据互操作性至关重要。通过注册元数据语义约束，可以确保数据语义的一致性、准确性和完整性，提升数据的可信度和可用性，有助于解决数据歧义和不一致问题，使数据被更规范、更有效地利用。

图2　MDR研究词云图

2　约束相关概念关系与约束外延

2.1　约束与相关概念的关系

为确保MVCRM构建过程中语义准确，避免概念混淆，本节总结约束的相关概念及关系（见图3）。其中本体是对共享概念模型形式化、规范化、明确的说明，由类、关系、属性、公理和实例5种元素组成，与MDR中需要注册的数据信息相对应。在最简单情况下，本体是以包含关系为主的概念分类层次结构。在复杂情况下，本体还包括了概念之间的多种关系约束[19]，MDR中同样存在这些约束。公理是本体中永真的断言，是逻辑推理与关系描述的起点，用于定义概念的约束条件。本体中任何一种组成元素的本质都是公理，如类公理、注释公理、属性公理等。为了清晰地表达本体中的类、关系、属性和实例，需要对这些元素进行约束。约束是对接受某项断言必须成立的一种形式化的声明，用于限制

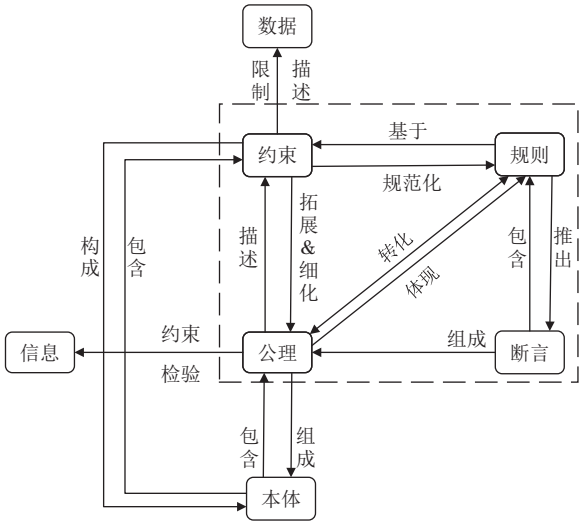

图3　约束与其相关概念的关系模型

数据或描述数据特征[20]。规则建立在约束之上，是依据特定形式的陈述得出的逻辑推论，类似于If-Then（前因-后果）形式的声明。规则和公理可相互转化，如在本体中，业务规则可以通过公理表达。断言是断定一个特定前提为真的陈述。本体中的断言包含了规则，并且基于推理规则可以得出断言，断言又组成了公理，因此，公理是断言的集合[21]。

2.2 元数据语义约束外延分类模型

为明确MVCRM中的元数据语义约束组成要素，本节基于MDR2023并结合本体语义约束，提出元数据语义约束外延分类模型，如图4所示。MDR2023主要将元数据语义约束分为概念约束（Concept_Constraint_Set）和关系约束（Relation_Constraint_Set）两大类，概念约束又分为数据元约束（Data_Element_Constraint_Set）、数据元概念约束（Data_Element_Concept_Constraint_Set）、可枚举概念域约束（Enumerated_Conceptual_Domain_Constraint_Set）、值域约束（Value_Domain_Constraint_Set）和允许值约束（Permissible_Value_Constraint_Set）五大类，关系约束又分为链约束（Link_Constraint_Set）、链端约束（Link_End_Constraint_Set）、关系角色约束（Relation_Role_Constaint_Set）和数据元值上下文约束（Data_Element_Values_in_Context_Constraint_Set）四大类[22]。虽然MDR2023对元数据语义约束的分类较为全面，但还有待细化。

通过调研相关文献发现本体约束与元数据语义约束有颇多相似之处，二者可以融合，以丰富和细化元数据语义约束类别。例如，林汝坤等[23]将本体中的约束公理划分为值约束（Value_Constraint_Set）和基数约束（Cardinality_Constraint_Set），其中，值约束限制属性值域的具体值，基数约束限制属性值域的取值个数。由于属性本质上是元数据的一种表达形式，值约束和基数约束同样可用于元数据的值域。然而，元数据语义约束中并没有这两种约束，故将本体的语义约束融入元数据语义约束，以扩展和细化其外延分类。因此，允许值约束又细分为值约束和基数约束，其中值约束又可细分为量词约束（Quantify_Constraint_Set）和指定值约束（Has_Value_Constraint_Set），量

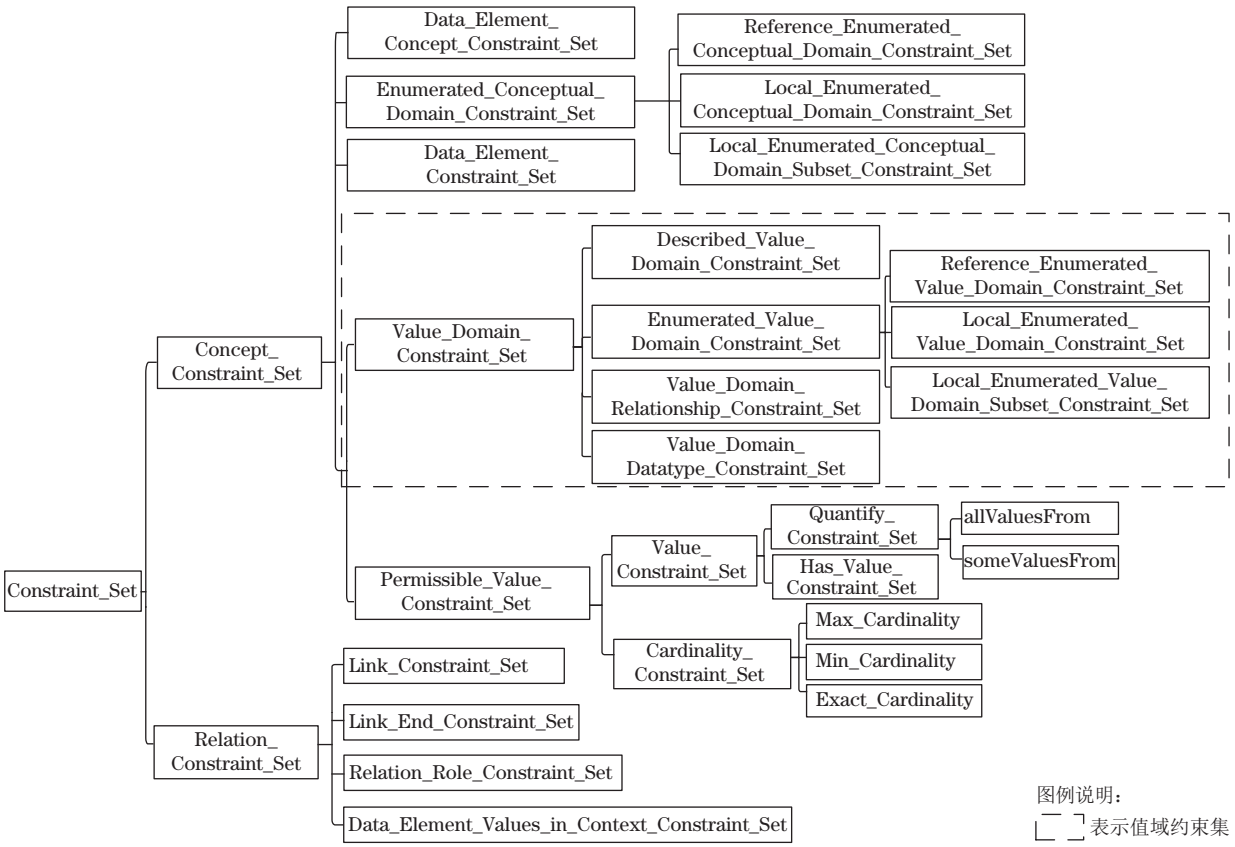

图4 元数据语义约束外延分类模型

词约束主要指全称量词（allValuesFrom）和存在量词（someValuesFrom）。基数约束包括最大基数（Max_Cardinality）、最小基数（Min_Cardinality）和准确基数（Exact_Cardinality）。此外，结合本体约束，值域约束又可细分为值域关系约束（Value_Domain_Relationship_Constraint_Set）和值域数据类型约束（Value_Domain_Datatype_Constraint_Set）等。

元数据值域语义约束用于规定元数据属性的合法取值范围，是元数据语义表达准确一致的重要前提，也是数据治理的关键环节。由图4可知值域约束在该模型中所占比例显著，表明值域约束在元数据语义约束中具有重要地位。在概念系统中，类具有属性，这些属性具有值域。属性可以用元数据来表达，由此可推出元数据也具有值域。例如，Person（人）类，可以有元数据Age（年龄），Age的值域是整数类型（int）。实际上，大多数约束都涉及对值域的限制，无论是在数据元的取值还是允许值方面，都需要对值域进行明确的约束。因此，值域约束在元数据的语义约束中起着至关重要的作用。

3　MVCRM及算法构建

3.1　MDR2023概念系统元模型

MDR系列标准为机器精确理解数据语义、获取数据语义成分提供了通用的框架，其中MDR概念系统元模型是该框架的核心元模型，通常用于组织和表达概念及概念体系，管理MDR中的概念及关系。MDR2023又提出了新的概念系统元模型（见图5）。相比MDR2013中的概念元模型，更加简洁，同时加入了概念约束和关系约束，用于表示概念自身的约束与概念之间的约束，使得概念系统的语义更加丰富和严谨[24]。MDR2023概念系统元模型由九元组组成。①Concept_System表示概念体系集合，概念体系是概念及概念间的关系和约束构成的集合。②Concept表示概念集合，概念是某些特性的唯一组合形成的知识单元。③Relation表示关系集合，用于表达概念之间的联系。④Relation_Role表示关系角色集合，说明参与某关系的概念在此关系中承担的角色。⑤Link表示链接集合，指概念间存在的某种链接，由相同的元组Link_End组成。⑥Link_End表示链端集合，一个链端是一个概念，与一个关系角色组成一个配对。⑦Assertion表示断言集合，在一个概念体系中，一个断言是判断一个断定或假设为真的逻辑命题或语句，Link是Assertion的子集。⑧Concept_Constraint_Set表示概念约束集合，是约束集（Constraint_Set）的子类，用于表示概念自身的约束。⑨Relation_Constraint_Set表示关系约束集合，扩展了约束集，其约束条件适用于概念系统包中使用关系类的约束，用于表达概念间的关系约束。

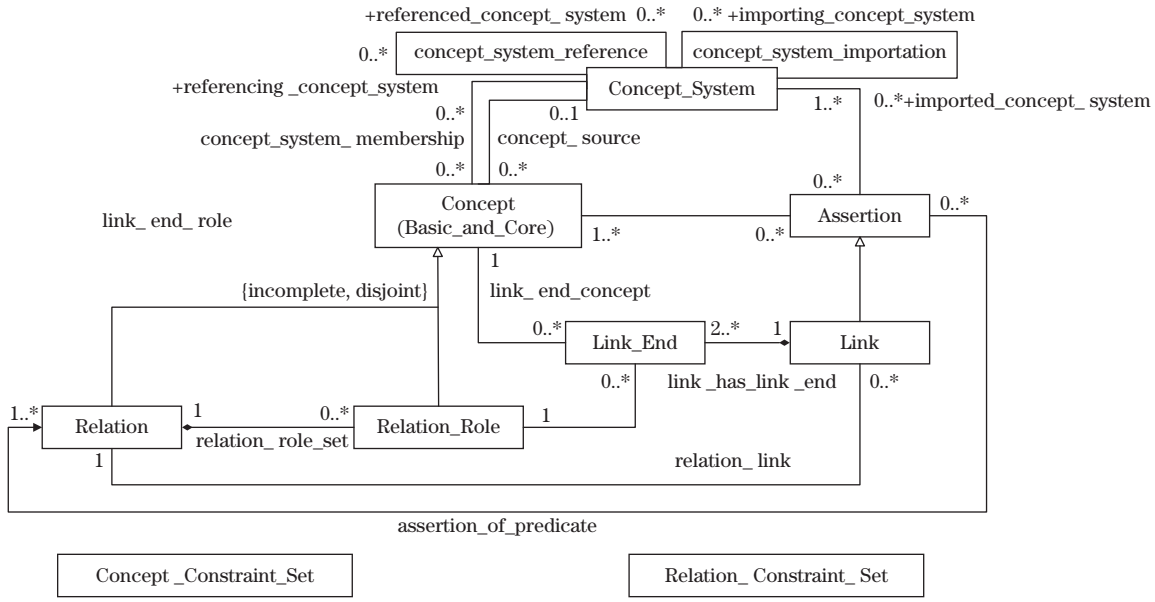

图5　MDR2023概念系统元模型

3.2 MVCRM

为了确保元数据语义的一致性和互操作性，MDR2023定义了多个元模型。其中的概念系统元模型描述了概念间的关系表示方法，以及概念系统中的两大基本约束类别。概念域与值域元模型解决了概念域与值域的注册与管理问题，该元模型详细描述了数据元的概念域与值域之间的对应关系[15]，并对其中需要进行约束的属性进行了说明。因此，在MDR2023的概念系统元模型和概念域与值域元模型的基础上进行结合并扩充，并融合2.2节所述的元数据语义约束构建基于MDR2023的MVCRM，如图6所示。在图6中，用斜体表示的类为抽象类，这意味着只能实例化它们的具体子类。MVCRM为元数据语义值域约束提供标准化注册框架和必备的约束语义注册要素，提升数据治理过程中元数据语义表达的标准化程度，并维护了数据语义的合法性和一致性。

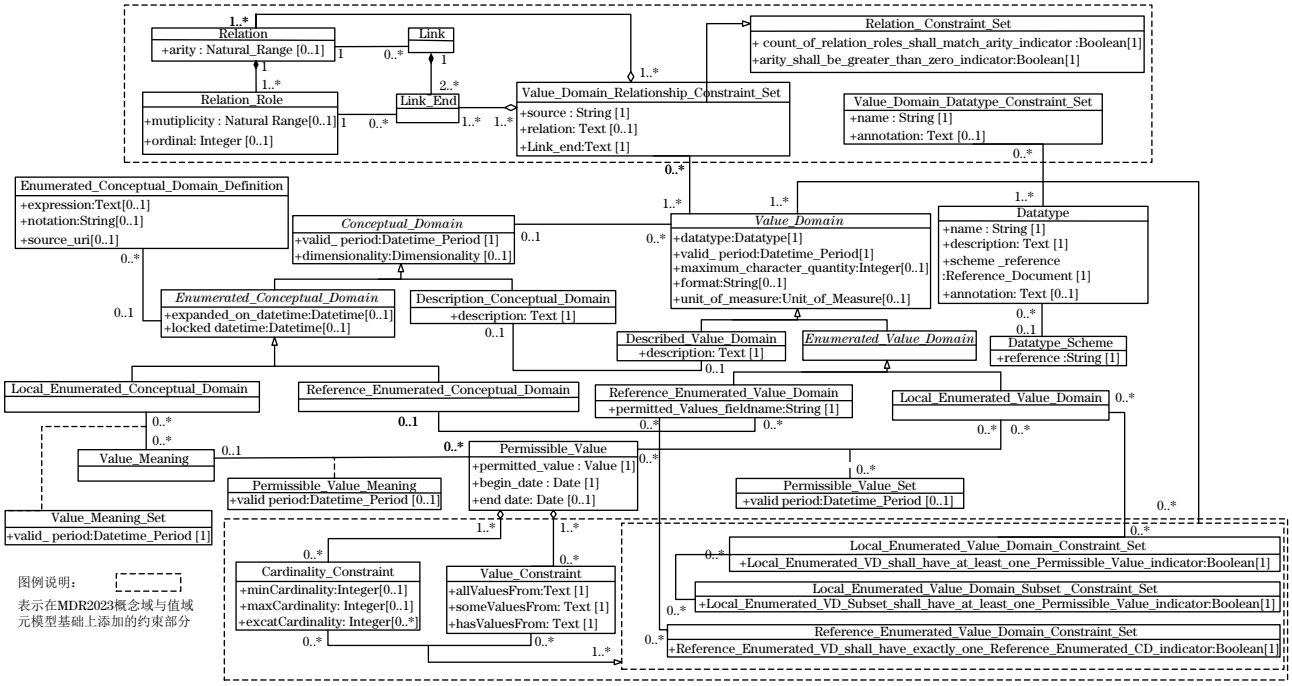

图6 基于MDR2023的MVCRM

MVCRM提取了概念系统元模型中的Relation、Relation_Role、Link_End、Link、Relation_Constraint_Set和Concept_Constraint_Set来表示值域关系约束和其他约束，保留了概念域与值域元模型所特有的要素及其属性，即概念域（Conceptual_Domain）、值域（Value_Domain）、值含义（Value_Meaning）和允许值（Permissible_Value）。值域约束作用于元数据的具体值，而描述型值域没有明确的值，因此，将基数约束和值约束融入可枚举值域的允许值。此外，还增加了值域之间的关系约束和数据类型约束，即融入MDR2023概念系统元模型中的Relation来表示值域之间的语义关系约束，最终实现了对MVCRM的设计。一个完整的MVCRM由5种语义要素构成。①Conceptual_Domain表示概念域的集合，一个概念域通常用来表示一组值含义（Value_Meaning）。概念域包括可枚举概念域（Enumerated_Conceptual_Domain）和描述型概念域（Description_Conceptual_Domain）。②Value_Domain表示值域的集合，值域是一组允许值的集合，用于表达一组允许值，描述了数据被表达的方式。数据类型以及可能的计量单位都与值域相关。具体包括可枚举值域（Enumerated_Value_Domain）和描述型值域（Described_Value_Domain）。可枚举值域又细分为本地可枚举值域（Local_Enumerated_Value_Domain）和引用型可枚举值域（Reference_Enumerated_Value_Domain）。其中，引用型可枚举值域与引用型可枚举概念域相对应。③Value_Meaning表示值含义的集合，一个值含义是一个真实值的一般含义。④Permissible_Value表示允许值的集合，一个允许值是一个值含义的指称。⑤ValueDomain_Constraint表示值域约束的集合。ValueDomain_Constraint融合了概念系统元

模型中的概念约束和关系约束，并在此基础上结合2.2节进行细化，得到ValueDomain_Constraint包括本地可枚举值域约束（Local_Enumerated_Value_Domain_Constraint_Set）、本地可枚举值域子集约束（Local_Enumerated_Value_Domain_Subset_Constraint_Set）、引用型可枚举值域约束（Reference_Enumerated_Value_Domain_Constraint_Set）、值域关系约束（Value_Domain_Relationship_Constraint_Set）和值域数据类型约束（Value_Domain_Datatype_Constraint_Set）。值域约束对属性的值域进行限制，进而规范化概念域和值域的语义表达，避免不合法数据的出现。

3.3　元数据值域语义约束注册规则

为确保数据治理中注册数据语义的准确性和一致性，元数据值域语义约束注册需要遵守相关规则。ISO/IEC TR 20943-3:2004（以下简称“20943-3”）描述了一组用于实现MDR内容一致性的程序，目的是通过保持注册的值域及其属性一致，确保对数据元属性以及值域属性有共同的理解，以便元数据在注册系统之间共享。20943-3是基于ISO/IEC 11179的标准，用于对值域及其组件进行概念化，以便标准化建立高质量的元数据，对于数据治理的标准化具有重要意义[25]。

20943-3描述了基于MDR进行元数据值域语义约束注册所需遵守的基本规则，见表2[25]。尽管MDR2023系列标准全面给出了元数据概念系统的注册和管理方案，但其中缺乏值域注册的详细方法。因此，要构建MVCRM，还需参考20943-3，为元数据值域语义约束注册的标准化流程和规则的制定提供方法论。

3.4　元数据值域语义约束注册算法流程

在实现过程中，需要从应用角度对MVCRM进行梳理，并结合20943-3总结出元数据值域语义约束注册算法（Metadata Value Domain Semantic Constraint Registration Algorithm，MVCRA）流程，如图7所示，具体描述如下。

表2　元数据值域语义约束注册约束规则

规 则	解 释
一个值域应有一个关联的概念域	一个概念域可以独自存在，但一个值域应有一个关联的概念域，即先定义概念域才可以确定值域
本地可枚举值域最少有一个允许值	本地可枚举值域和本地可枚举值域子集应至少具有一个允许值指示器
值含义需要维护	值含义是将不同的可枚举值域中不同的值与相同含义进行关联的方法，一些值域会不断变化，因此值含义需要维护
概念域和值域的子类不是排他的	概念域和值域的子类之间是非互斥关系，即子类是可以同时存在的，例如一些值域或概念域可以同时有一个可枚举部分和一个不可枚举的部分
数据类型需要文档化	数据类型是注册的重要部分，文档化数据类型确保每个值都有一个明确的定义，还可用作标准化规则的参考，以确保所有注册内容都符合同一标准
数据值用计量单位描述	计量单位用来描述数据值以确定不同来源数据之间的相似性，是任何计量文档所必需的部分
一个值含义与一个允许值关联不是必须的	值含义集合不必与任何可枚举的值域中的允许值一一对应
一个概念域是值含义的一个集合	概念等价可枚举值域不必是基本等价的，但是概念等价的域可能有同数量的允许值
概念等价可枚举值域的允许值数量是不确定的	组织、内外部用户和标准规定的需求都会对允许值的数量产生影响

Step 1：确定用于值域注册的属性是否存在一个相匹配的概念域。注册值域的前提是必须存在一个概念域，并且每个值域只能与一个特定的概念域相关联。

Step 2：如果一个合适的概念域已经存在，需要确定值域本身是否存在。

Step 3：理解值域的什么值是允许的。针对不可枚举值域，需要全面理解定义允许值的规则；而对于可枚举值域，则需要理解值的结构和含义。

Step 4：判断值域是可枚举值域还是不可枚举值域。如果是可枚举值域，除了注册值域基本信息之外，还需注册值域的每个允许值的值和值含义。如果是不可枚举值域，则需要注册描述、维度、计量单位等属性。

Step 5：如果是可枚举值域，则需要注册可枚举值域的数据类型约束、基数约束和值约束等约束信息。

Step 6：判断该值域对应的概念域是否与其他概念域相关。当值域的一个集合注册为一个分类时，该概

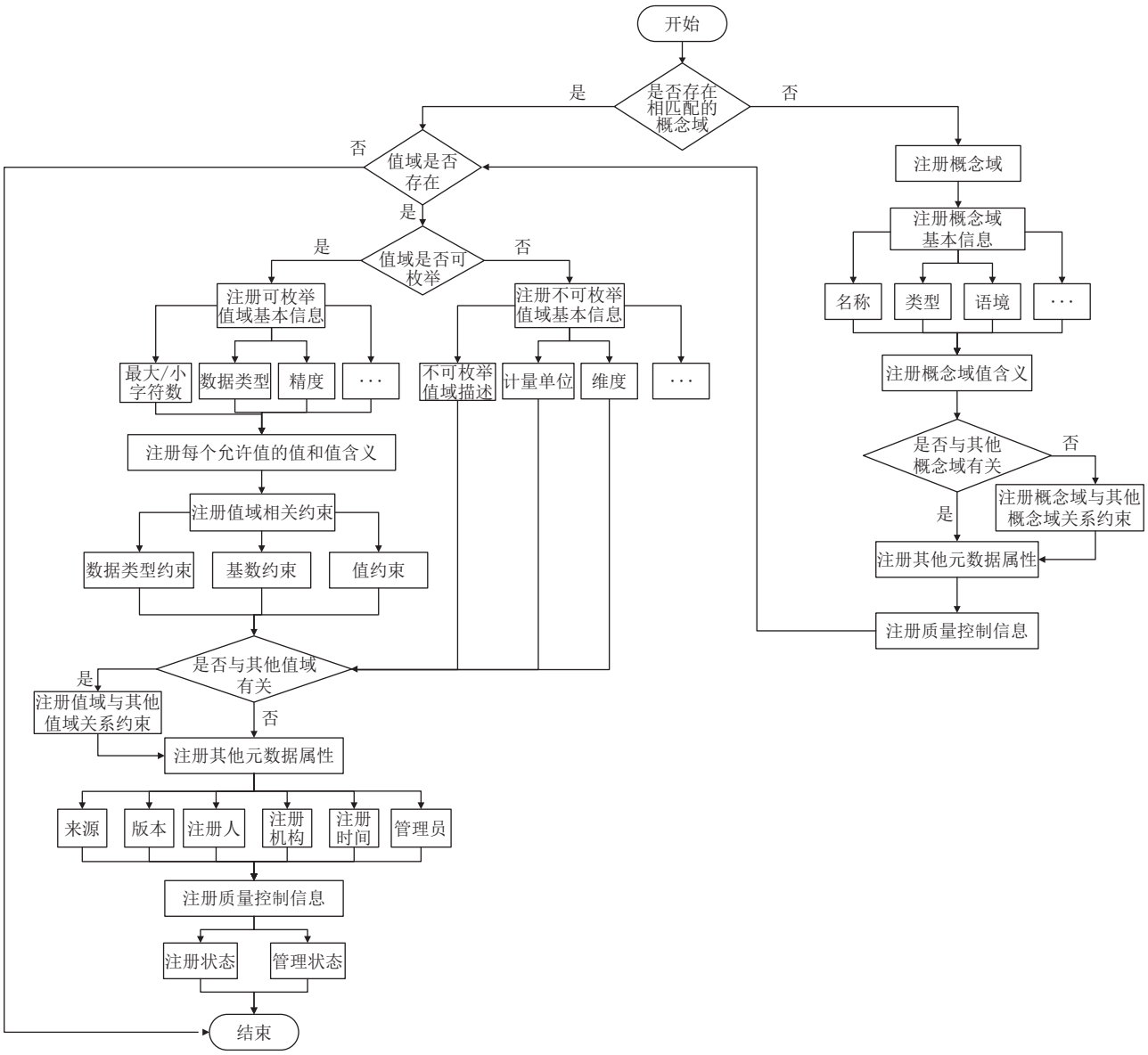

图7　MVCRA流程

念域就与其他概念域相关。

Step 7：判断值域是否与其他值域存在关系。当值域的一个集合注册为诸多分类时，则该值域与其他值域相关。

Step 8：注册概念域和值域的其他元数据属性，包括来源、版本、注册机构和管理员等。

Step 9：注册质量控制信息。

4　MVCRS构建与应用

以石油领域为例，结合石油领域著名的POSC标准中的值域约束进行研究，构建标准化MVCRS，用于应用和验证MVCRM。POSC是目前国际上权威的石油数据标准化组织。POSC发布了代表性的数据模型标准Epicentre3.0[26]，该标准按照面向对象思想组织数据，其核心思想是“对象—活动—特性”[27]。Epicentre3.0详细定义了石油领域的专业词汇，并对实体中的数据属性的值域、实体间的关系和数据类型等进行了详细描述和规范约束，对石油领域有重要应用价值[28]。但该标准中的约束较为隐蔽和分散，该标准并没有明确指出具体包括哪些约束，需要进行约束提取和映射。通过将Epicentre3.0中的约束提取并映射成2.2节中的约束，从而确定构建MVCRS中需要注册的属性和约束等信息。

以Epicentre3.0中的“井”（well）这一实体的外延即“井类别”（well_classification）这一值域为例，构建以well_classification为实例的元数据值域语义约束注册实例图，明确MVCRS需要注册的约束、属性以及值域间的关系等信息，如图8所示。

结合上述实例图与MVCRA进行元数据值域语义约束的注册。在实际注册过程中，可以简化并合并相关步骤。在确定概念域已注册的情况下，即可按照图7中的步骤进行注册，以图8为例。第一步，注册well_classification这一值域的基本信息，包括数据类型、精度和标识符等信息。第二步，注册值域的允许值和相关约束，well_classification是可枚举值域，本例中注册其允许值为well_production（生产井）、well_ exploration（探井）和well_abandonment（报废井），并注册well_classification的基数约束、值约束以及值域关系约束。第三步，注册值域的其他元数据属性，包括来源、版本、注册机构等信息。第四步，注册质量控制信息，包括注册状态和管理状态。实践中需要注册的值域约束有很多，但基本注册流程类似，可在本文的基础上进一步拓展以适应其他领域的元数据值域语义注册。

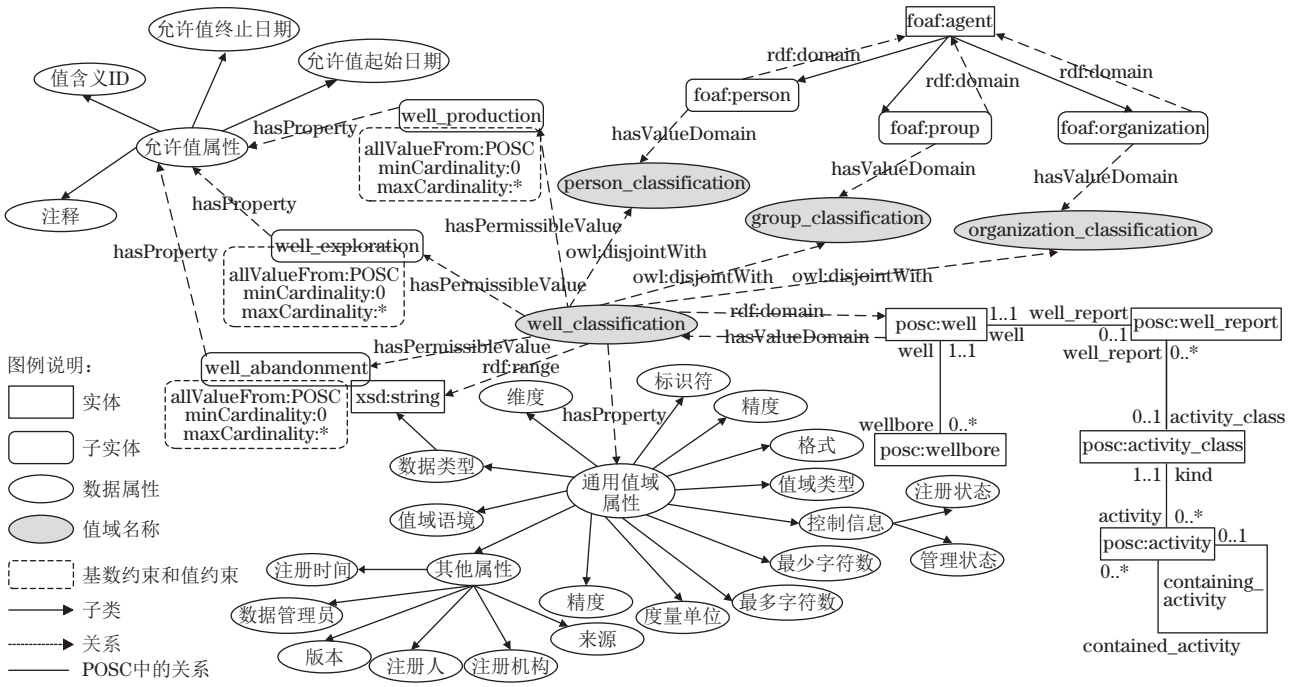

图8　元数据值域语义约束注册实例

由MVCRS在Epicentre3.0中的应用可见，通过注册值域的基数约束、值约束和关系约束等方式，能够丰富和规范元数据语义，确保元数据语义格式、语义范围和语义关系结构的准确性和完整性。因此，验证了基于MDR2023的MVCRM及MVCRA的可行性和合理性。

5　结语

本文主要解决目前数据治理中数据语义约束注册缺乏的问题。通过研究MDR2023相关标准和国内外MDR研究现状及应用，整合了国际上各领域常用的MDR信息。从理论底层探讨了语义约束与相关概念的关系，并提出元数据语义约束外延分类模型，构建了基于MDR2023的MVCRM，并制定了详细的MVCRA流程。根据该模型和算法流程构建了MVCRS，并用石油领域的POSC标准验证了方案的可行性和有效性。本文提出的标准化模型及算法具有普适性和拓展性，可适用于各领域的元数据语义约束知识表示与管理。未来研究将深化并扩展元数据语义约束外延分类模型，进一步完善其他类型约束的注册模型及注册流程，最终目标是建立一个完善且标准化的元数据语义约束注册系统，为数据治理中元数据语义的标准化奠定基础。

参考文献

[1] 浙江省物联网产业协会. 数据中台 元数据规范：T/SHMHZQ 035—2022[S]. 北京：中国标准出版社，2022.

[2] 国家质量监督检验检疫总局，中国国家标准化管理委员会. 科技平台 元数据标准化基本原则与方法：GB/T 30522—2014[S]. 北京：中国标准出版社，2014.

[3] 中华人民共和国文化部. 管理元数据规范：WH/T 52—2012[S]. 北京：中国标准出版社，2012.

[4] 中华人民共和国文化部. 电子图书元数据规范：WH/T 65—2014[S]. 北京：中国标准出版社，2014.

[5] 中华人民共和国国家卫生健康委员会. 卫生健康信息数据集元数据标准：WS/T 305—2023[S]. 北京：中国标准出版社，2023.

[6] 中华人民共和国教育部. 基础教育教学资源元数据 实施指南：JY/T 0610—2017[S]. 北京：中国标准出版社，2017.

[7] 国家市场监督管理总局，国家标准化管理委员会. 信息与文献 文件（档案）管理元数据 第2部分：概念化及实施：T/CVIA 128-2023[S]. 北京：中国标准出版社，2023.

[8] Information technology-Metadata registries（MDR）-Part 1：framework：ISO/IEC 11179-1：2023[S]. Geneva：International Organization for Standardization，2023.

[9] Information technology-Metadata registries（MDR）-Part 3：metamodel for registry common facilities：ISO/IEC 11179-3：2023[S]. Geneva：International Organization for Standardization，2023.

[10] ULRICH H，KERN J，TAS D，et al. QL^{4}MDR：a GraphQL query language for ISO 11179-based metadata repositories[J]. BMC Medical Informatics and Decision Making，2019，19（1）：45.

[11] KIM H H，PARK Y R，LEE S，et al. Composite CDE：modeling composite relationships between common data elements for representing complex clinical data[J]. BMC Medical Informatics and Decision Making，2020，20（1）：147.

[12] HEGSELMANN S，STORCK M，GESSNER S，et al. Pragmatic MDR：a metadata repository with bottom-up standardization of medical metadata through reuse[J]. BMC Medical Informatics and Decision Making，2021，21（1）：160.

[13] ARIENTI C，CAMPAGNINI S，BRAMBILLA L，et al. The methodology of a “living” COVID-19 registry development in a clinical context[J]. Journal of Clinical Epidemiology，2022，142：209-217.

[14] FERENZ S，NIEßE A. Towards improved findability of energy research software by introducing a metadata-based registry[J]. Ing.grid.，2023，1（2）：3837.

[15] 刘旭. MDR概念体系注册元模型标准研究[D]. 大庆：东北石油大学，2020.

[16] 袁靖舒，李洪奇. MDR概念系统注册及本体表示标准化研究[J]. 西南石油大学学报（自然科版），2020，42（6）：174-180.

[17] 田中贺，满春涛，关虎，等. 数字内容资源登记注册服务系统设计与实现[J]. 哈尔滨理工大学学报，2021，26（4）：119-124.

[18] 黄安琪，苗放，杨文晖，等. 基于数据架构的结构化数据注册引擎设计[J]. 计算机与现代化，2022（5）：82-89，95.

[19] BORST W N. Construction of engineering ontologies for knowledge sharing and reuse[EB/OL]. [2023-11-12]. https://research.utwente.nl/en/publications/construction-of-engineering-ontologies-for-knowledge-sharing-and-.

[20] NOY N F，SINTEK M，DECKER S，et al. Creating semantic web contents with Protege-2000[J]. IEEE Intelligent Systems，2001，16（2）：60-71.

[21] 张德海. 本体学习的认知模型[M]. 北京：科学出版社，2017.

[22] Information technology-Metadata registries（MDR）-Part 31：metamodel for data specification registration：ISO/IEC 11179-31：2023[S]. Geneva：International Organization for Standardization，2023.

[23] 林汝坤，刘芳，戴长华，等. OWL本体建模中约束公理的应用[J]. 计算机工程，2006，32（16）：193-194，223.

[24] Information technology-Metadata registries（MDR）-Part 32：metamodel for concept system registration：ISO/IEC 11179-32：2023[S]. Geneva：International Organization for Standardization，2023.

[25] Information technology-Procedures for achieving metadata registries contentconsistency-Part 3：value domains：ISO/IEC 20943-3：2004[S]. Geneva：International Organization for Standardization，2004.

[26] Petrotechnical Open Software Corporation. POSC Specifications Epicentre Version 3.0[EB/OL]. [2023-02-21]. http://w3.energistics.org/archive/Epicentre/Epicentre_v3.0/index.html.

[27] 袁满. 石油数据组织与分析[M]. 东营：中国石油大学出版社，2016：126-130.

[28] YUAN J S，LI H Q. Research on the standardization model of data semantics in the knowledge graph construction of Oil&Gas industry[J]. Computer Standards & Interfaces，2023，84：103705.

作者简介

袁满，男，博士，教授，研究方向：知识工程与数据标准化。
何玲通，女，硕士研究生，研究方向：语义约束。
袁靖舒，男，博士，讲师，通信作者，研究方向：知识工程，E-mail：yuanjingshu@nepu.edu.cn。
李洪欣，女，硕士研究生，研究方向：业务规则。

Research on the Metadata Value Domain Semantic Constraint Registration Model Based on MDR2023

YUAN Man HE LingTong YUAN JingShu LI HongXin
(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, P. R. China)

Abstract: Metadata registry (MDR) is a necessary prerequisite for the precise expression of semantics in metadata during data governance. This paper conducts a systematic analysis of MDR systems both domestically and internationally, finding that MDR focuses more on the registration of basic data elements, and there is a lack of research in the area of data semantic constraint registration. Therefore, this paper first proposes a metadata semantic constraint extension classification model based on ISO/IEC 11179:2023 (MDR2023) standards, clarifies the scope of metadata semantic constraints, and selects the value domain semantic constraints for detailed study. Secondly, the metadata value domain semantic constraint registration metamodel is proposed based on MDR2023 standards, providing a standardized and complete registration algorithm process for metadata semantic constraint registration and a solution for registering metadata value domain semantic constraints. Finally, with the famous POSC standard in the petroleum field as the demand background, the value domain semantic constraints are registered, thereby realizing the standardization of the value domain semantic constraints of the metadata in the petroleum field, and verifying the rationality and feasibility of the metadata value domain semantic constraint registration metamodel proposed in this paper. The metamodel proposed in this paper is generalizable to other domains of data governance.

Keywords: Metadata; Value Domain; Semantic Constraint; Metadata Registry; Registration Metamodel; Data Semantics Standard

（责任编辑：王玮）