

# 基于内容结构视图的研究数据 元数据标准比较研究

□朱玲\*

**摘要** 元数据是保证研究数据(又被称作“科学数据”)获取和再利用的关键,国内外许多学科领域和学术社区都制定了该领域特定的研究数据元数据标准或规范。与文献资源元数据标准相较,研究数据元数据标准结构和元素设计往往更为复杂,为跨学科的数据分享和发现带来了障碍。文章应用一种内容结构视图,对八种典型的研究数据元数据标准元素进行比较分析,分析不同研究数据元数据标准的内容特征,并由此探讨研究数据元数据标准发展趋势,为元数据应用方案设计提供参考。

**关键词** 研究数据 科学数据 元数据标准 内容结构视图

**分类号** G254

**DOI** 10.16603/j.issn1002-1027.2019.06.013

## 1 引言

研究数据,是指科研人员通过观察、探测、实验、建模、仿真和调查等形式收集的,用于验证研究发现的,能够直接或间接产出学术成果的完整事实材料和数字产品,例如:文档、数据文件、问卷、模型、算法、软件或代码、图片、音视频资料等。

元数据,被定义为“描述数据的数据”。相应地,研究数据元数据是指描述研究数据集的创建者、上下文、空间、时间、起源、历史、数据质量等信息,以便支持数据使用者对数据的发现、选择、定位、获取、解释和再利用等功能,以及数据拥有者或管理者对数据的组织、管理、共享和长期保存等功能。

2011年,英国数字监护中心(Digital Curation Center)发布了《数字监护参考手册—科学数据元数据部分》(Digital Curation Reference Manual-Installation on Scientific Metadata),论述了科学数据元数据的特征和角色,指出元数据在数据仓储的管理和数据监护中扮演中心角色<sup>[1]</sup>。研究数据元数据在数据管理计划和公共资助研究项目研究数据的公众获取中也扮演着核心角色,例如,2011年美国国家自然科学基金会(NSF)新增的研究数据管理计划要求中明确把“将使用的数据和元数据格式标准”列

为重要建议<sup>[2]</sup>。2012年,欧洲研究图书馆协会(LIBER)发布的《图书馆开始从事研究数据管理的十条建议》中指出,图书馆应该参与元数据和相关数据标准的建设,同时应当提供研究数据相关的元数据服务<sup>[3]</sup>。美国密歇根大学校际政治及社会研究联盟(Inter-university Consortium for Political and Social Research, ICPSR)在其《维持领域数据仓储白皮书》(Sustaining Domain Repositories for Digital Data: A White Paper)中指出,元数据标准如保存元数据框架对于数据仓储的发现、检索、获取和科研群体标准的转变是非常必要的<sup>[4]</sup>。

## 2 研究数据元数据的功能需求

研究数据的产生和积累正在呈指数级增长,科学研究和科技创新越来越依赖于对海量数据的管理、分析和再利用。为了支持研究数据的发现、获取、解析和再利用,描述和呈现研究数据的元数据应该满足的功能需求可以从以下两方面进行探讨:

### 2.1 面向数据任务和数据活动的需求

研究数据元数据功能需求模型将元数据功能需求分为四个象限:数据管理、数据质量控制、数据发现和利用<sup>[5]</sup>。这四个象限分别支持不同的数据

\* 朱玲, ORCID: 0000-0003-2881-8050, 邮箱: zhul@lib.pku.edu.cn.



任务;数据管理象限支持元数据自动或半自动生成、数据管理、数据存储、数据安全、数据互通等任务;数据质量控制象限支持数据验证、复制等任务;数据发现象限支持数据分享(发布、引用)和数据发现(选择、唯一性确定、获取)等任务;数据使用象限支持数据分析、可视化等任务。

研究数据元数据对研究数据生命周期中的用户数据活动的支持,体现在为数据活动需要的信息设计相应的元数据模块和字段。例如,数据发现、数据唯一性确定、数据选择、数据获取等活动,属于与传统文献类似的一般性任务,元数据以描述性、身份标识、语义等信息内容为主,其中数据选择活动的需求还包括技术性、科研上下文、时空等信息内容。数据验证、数据分析属于科研任务,元数据主要以描述性、科研上下文、时空等信息内容为主。数据管理、数据存档属于数据监护任务,元数据主要以描述性、管理性、结构性和技术性等信息内容为主,并以身份标识、语义信息、科研上下文、时空等信息内容作为补充。数据发布、数据引用属于传播任务,元数据主要以描述性、身份标识、语义等信息内容为主,并以科研上下文、时空等信息内容作为补充。

## 2.2 面向不同服务对象的需求

针对不同的对象,研究数据元数据发挥的作用也不同<sup>[6]</sup>。对于数据用户而言,元数据是他们正确选择、使用、交换数据的不可缺少的工具。元数据通过对数据资源的内容、分类、质量和存储等的详细描述,回答了用户的一系列问题:有什么数据?数据怎么样?如何获取数据?怎么使用数据?对于数据管理者而言,元数据方便了他们集成各种数据库,为数据集(或库)建立目录,更有效地管理并维护海量数据。对于数据生产者而言,元数据的存在方便了数据的生产、加工和更新,并可以使数据归档更高效,使数据成为有生命力的资源,不必担心随着时间或者人员的变化而影响到数据的生产,数据的增值具有了持久性。研究数据元数据的最大用户是科研人员,元数据是他们发现数据、评价数据和使用数据的不可或缺的工具。

## 3 研究数据元数据标准发展现状

### 3.1 国内外研究数据元数据标准的发展和分布

元数据是保证研究数据获取和再利用的关键因素。国外许多学科领域和学者社区都采取了各种行

动制定和发展该学科需要遵守的元数据标准或规范,根据英国数字监护中心对目前国际上各学科的主流研究数据元数据标准的总结,其主要分布在通用学科、生物学、地球科学、物理学、社会科学与人文学科、跨学科等领域<sup>[7]</sup>。

与国外相比,我国研究数据元数据标准的制定起步较晚,但近年来在科学数据共享工程等项目推动下,取得了较快的发展。科学数据共享工程试点建立了9个学科领域(气象、地球系统、水文、海洋、地震、国土、农业、林业、人口健康数据共享)的科学数据共享平台。为了满足科学数据共享工程对元数据内容的基本需求,促进数据加工的规范化、标准化,科学数据共享工程办公室制定了《科学数据共享元数据内容》作为参考元数据标准<sup>[8]</sup>。以此为基础,各领域逐渐建立了各自的科学元数据标准或者元数据应用方案<sup>[9-15]</sup>。另外,中国科学院为了实现院内科学数据库群的资源利用、共享、交换和整合,制定了《科学数据库核心元数据标准》<sup>[16]</sup>,并为满足各学科领域的特殊要求,制定了元数据应用扩展方案的规则和方法。

### 3.2 研究数据元数据标准的复杂性

研究数据元数据的基本功能,是在研究数据共享和交换中提供快速查询和定位,通过元数据提供的信息描述,迅速查找所需信息的存放位置,发现原始数据的存在状况。除此之外,为了实现大型项目中学科领域内深度数据互通与合作的需求,研究数据元数据还需支持数据初步分析、数据质量评价和控制等功能。例如,生态学研究的空间叠加分析中,需要元数据提供分析所需的多层空间数据的范围、坐标体系、空间位置坐标等信息,对空间特征和属性进行合并。因此,研究数据元数据在基本发现信息之外,往往根据科研过程的需要进行了较多的扩展设计,如在时空范围、研究方法、数据加工处理过程等方面。

MODAL 框架是一种考察元数据标准实施文件组织结构和规模的研究方法,包括编码方式(Encodings)、结构类型(Structural Types)、元素规模(Extent)、实施文件数量、层级数量(Levels of Hierarchy)等。一项基于 MODAL 框架的研究对9种研究数据元数据标准的结构特征进行了分析<sup>[17]</sup>。结果表明,9种标准平均具有563个元数据元素(最少142个元素,最多1802个元素),9个实施文件

(最少 1 个文件,最多 25 个文件),5 个元素层级(最少 3 个层级,最多 10 个层级),大大超过文献资源元数据标准的复杂度。

的其他元数据元素放入其他元数据 (Miscellany) 模块。

#### 4 基于内容结构视图的研究数据元数据标准比较

##### 4.1 内容结构视图

研究数据元数据内容结构视图<sup>[18]</sup>(如图 1 所示)将跨学科、跨领域的研究数据元数据元素按照通用的结构模块进行分类和重组,这些结构模块组合成研究数据元数据的总体框架,数据集对象的综合信息透过此总体框架得以更为清晰地呈现。其中:识别元数据模块 (Identity Metadata) 提供支持识别功能和发现功能的基础信息,包含可以用于识别该数据集的所有实体信息,以及实体信息与该数据集的相互关系,例如数据集创建人、负责机构、数据集标识符等。语义元数据 (Semantic Metadata) 模块提供学科、主题、分类等信息,可以用于学科分类和链接到相关领域的其他数据。科研活动上下文 (Scientific Context) 模块包含工作流程、所用模型、数据来源、实验方法和参数等信息,这些信息可以被研究数据流程管理系统用于确定下一步数据分析操作。时间元数据 (Temporal Metadata) 和地理空间元数据 (Geospatial Metadata) 包含数据集所采用的时间和空间维度以及覆盖的范围。无法归于上面所列模块

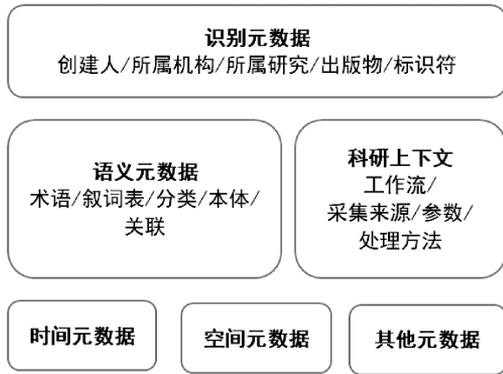


图 1 研究数据元数据内容结构视图

##### 4.2 八种典型研究数据元数据标准的结构视图分析

本文对三种通用学科研究数据元数据标准,即都柏林核心集 (Dublin Core)、DataCite Metadata Schema 和中国科学院科学数据库核心元数据标准,以及五种特定学科研究数据元数据标准 (生物学 Darwin Core、物理学 AVM、地球科学 CF、社会科学 DDI、人口健康科学数据共享元数据标准) 的元素或者元素分类应用上述结构视图进行分析,以便在此基础上识别和分析各个研究数据元数据标准的内容特征。结构视图分析结果如表 1 所示。

表 1 八种元数据标准内容的结构视图分析

研究数据元数据内容结构视图模块	识别元数据	科研活动上下文	语义元数据	地理/空间元数据	时间元数据
Dublin Core <sup>[19]</sup>	元数据元素:1. 贡献者 2. 时空覆盖范围 3. 创建者 4. 日期 5. 描述 6. 格式 7. 标识符 8. 语言 9. 出版者 10. 关系 11. 权益 12. 来源 13. 主题 14. 标题 15. 类型				
	√ 1,3,7,9,11,14	√ 12	√ 6,8,10,13,15	√ 2	√ 2,4
DataCite Metadata Schema V4.0 <sup>[20]</sup>	元数据元素:1. 标识符 2. 创建者 3. 标题 4. 出版者 5. 出版年 6. 资源类型 7. 主题 8. 贡献者 9. 日期 10. 语言 11. 替代标识符 12. 相关标识符 13. 大小 14. 格式 15. 版本 16. 权益 17. 描述 18. 地理位置 19. 资助项				
	√ 1,2,3,4,8,11,12,16,19	√ 13,14,15,17	√ 6,7,10	√ 18	√ 5,9
Darwin Core <sup>[21]</sup> (生物学)	元数据元素类:1. 数据集层面(含 19 个元素)2. 类群出现信息(含 21 个元素)3. 生物种群信息(含 7 个元素)4. 材料样本 5. 活标本 6. 保存标本 7. 化石标本 8. 事件(含 18 个元素)9. 人的观察 10. 仪器观测 11. 位置(含 44 个元素)12. 地质环境(含 18 个元素)13. 标识类(含 8 个元素)14. 类群(含 33 个元素)15. 测量或事实(含 9 个元素)16. 资源关系(含 7 个元素)17. 分类或编码体系名称(含有 43 个元素)				



研究数据元数据内容结构视图模块	识别元数据	科研活动上下文	语义元数据	地理/空间元数据	时间元数据
	√ 1,13	√ 2,3,4,5,6,7,8,9, 10,12,14,15	√ 1,16,17	√ 11,12	√ 8
AVM V1.2 rc1 <sup>[22]</sup> (天文学)	元数据标签分类:1. 核心类(创建者、权益、ID、出版时间、版本等)2. 语义类(主题等)3. AVM 特有语义类(主题、观测仪器等)4. 复合数据解析类 5. WCS 参考系信息基础类 6. WCS 参考系信息完整类				
	√ 1	√ 1,3,4,5,6	√ 2,3	√ 5,6	√ 1
CF V1.7 <sup>[23]</sup> (地球科学)	全局属性(即顶层描述性元数据元素)和变量属性分类:1. 文件描述类(标题、机构、来源、历史、参考资料、评论)2. 变量属性类(单位、标准名称、长名称、辅助数据、旗标)3. 坐标类型类(经度坐标、纬度坐标、高度坐标、时间坐标、离散轴)4. 坐标系 5. 标签(地理位置)6. 网格类(边界、气候统计)				
	√ 1	√ 2,3,4,6		√ 5	√ 3
DDI 3.0 <sup>[24]</sup> (社会科学)	元数据元素模块(支持整个数据统计生命周期):1. 存档 2. 比较 3. 概念组件 4. 数据采集 5. 都柏林核心元素 6. DDI 配置文件 7. 分组 8. 实例(根元素,包含子元素;引用、覆盖范围、分组、资源包、学习单元等)9. 逻辑产品 10. 物理数据产出 11. 物理实例 12. 学习单元				
	√ 1,5,8	√ 2,3,4,7,8,9,10,11,12	√ 5	√ 8	√ 8
人口健康科学数据共享元数据标准	元数据子集:1. 引用信息 2. 标识信息 3. 内容信息 4. 分发信息 5. 数据质量信息 6. 数据表现信息 7. 扩展信息(人口健康领域在《科学数据共享元数据内容》之外的扩展)8. 数据模式信息 9. 限制信息 10. 维护信息				
	√ 1,2	√ 5,6,7,8,9,10	√ 3		√ 1
科学数据库核心元数据标准 v2.2 <sup>[25]</sup>	元数据元素类:1. 描述类:标题 2. 描述类:URI 3. 描述类:简介/目标 4. 描述类:关键词/分类/类型/语言 5. 描述类:时空覆盖范围 6. 描述类:责任人 7. 描述类:来源/数据志/更新频率 8. 描述类:质量报告 9. 描述类:大小/格式 10. 描述类:资源关系 11. 描述类:权益/分享 12. 参考信息类 13. 联系信息类				
	√ 1,2,6,10,11,12,13	√ 3,7,8,9	√ 4	√ 5	√ 5

### 4.3 元数据标准内容特征分析

#### (1)元数据标准内容结构特征

Dublin Core 元数据标准是一种平行或并列结构,包含 15 个元素,有的元素具有对元素语义做进一步限定的修饰词,修饰词的语义包含于被修饰词(即元素)之中,对于不了解修饰词使用方法的用户来说,可以直接使用元素本身来描述资源。

DataCite Metadata Schema 元数据标准、Darwin Core 元数据标准、AVM 元数据标准可以看作是一种树状结构,其中只有叶节点才真正有价值。例如,Darwin Core 标准下的每一个元素类可以看作是“根”下面的一个“枝”,下面的元素是“叶”。Dat-

aCite Metadata Schema、AVM 的一级元素可以看作直接附属在“根”下面的“叶”,而二级元素可以看作附属在“根”下面一个“枝”上的“叶”。

科学数据库核心元数据标准、人口健康科学数据共享元数据标准可以看作是一种多层父子结构,其中最顶层的根元素属于复合型元素,下面可包含其他复合型元素或单纯型元素作为子元素。递归地,每一级的复合型元素都由一个或多个下一级复合型或单纯型子元素组成。

以上六种元数据标准重点都是从数据集层面的特征进行描述,而 CF 元数据标准和 DDI 元数据标准则更进一步引入了用于变量定义和描述的庞大的



变量属性集,支持专业领域的变量理解和操作。这两种元数据标准在数据集层面描述上可以看作平行或并列结构,在变量描述上可以看作是带有庞大修饰词体系(变量属性:例如变量单位、坐标、取值范围、测量误差、缺失插值等)的可重复元素的平行或并列结构。

## (2)元数据标准内容分布特征

从上文中的内容结构视图分析可以得出,通用学科研究数据元数据标准内容更大比例分布在识别元数据模块,而特定学科研究数据元数据标准内容更大比例分布在科研活动上下文模块,其中:

Dublin Core 元数据标准定义了 Web 资源都应遵循的通用的核心标准,其内容较少,也比较通用,描述对象并不限于研究数据,因此其元数据元素主要为识别类和语义类,较少覆盖科研活动上下文模块。

DataCite Metadata Schema 元数据标准、科学数据库核心元数据标准,这两种元数据标准以描述通用学科的研究数据为目标,其元数据内容主要分布在识别模块和科研活动上下文模块,但科研活动上下文的字段与学科无关。

Darwin Core 元数据标准、AVM 元数据标准、人口健康科学数据共享元数据标准,这三种元数据标准的特征是设计了较多的子类和字段对特定学科领域的科研活动上下文进行描述,描述主要以数据收集、观测等科研活动的参数为对象。

CF 元数据标准、DDI 元数据标准,这两种元数据标准也有较多的字段属于科研活动上下文模块,但其特征在于对本学科的数据、文件、变量等进行了较为充分的分层描述。

## 5 研究数据元数据标准发展趋势

### 5.1 支持跨学科数据分享和发现

近年来,各学科领域的研究数据元数据标准设计和实践取得了积极成果,但专门针对学科需求的元数据标准的增长也为跨学科的数据发现和再利用造成了一定的人为障碍,这些障碍体现在元数据语义信息和数据结构的多样性,以及对多学科的科学进程的干扰。

当代科学被认为是一个结合了专业化和多学科或跨学科的综合研究,越来越多的成果和突破来自

学科之间的边界区域,研究者需要越来越频繁地与其他学科领域交换信息、技术和工具。实现跨学科元数据的互通和分享,从而实现跨学科的数据管理、监护、再利用,是研究数据元数据标准发展的必然趋势。

从研究数据元数据元素的类别分布来看,分布极不均匀,例如科研上下文类别中的元素种数远超其他类别的元素种数;从研究数据元数据元素的统计频次来看,在不同标准中具有最高统计频次的元素都属于描述类别,且大部分与 DC 元素重叠;少数语义性的一般元素在标准中也出现较多的频次;其余的元素统计频次很少,共同形成了具有各自领域特定语义的长尾部分。为了支持跨学科领域的研究数据分享,研究数据元数据标准的设计和 implement 中,需要重点考虑模块化、可扩展性、灵活性、综合性、充足性、简易性等目标。

### 5.2 嵌入研究数据生命周期

研究数据的生命周期管理是确保其连续性的必要条件。如果元数据希望捕捉更加动态的结构,认识到什么信息在持续,而什么信息在数据随着时间的推移期间发生了什么变化,就必须基于对研究数据生命周期的理解和应用。例如,DDI 生命周期模型包含研究构想、数据收集、数据处理、数据分发、数据发现 and 数据分析等线性模块,以及作为可选路径的数据存档模块和作为反馈循环路径的数据重新规划模块。相应地,DDI 元数据标准的元素(Codebook)<sup>[26]</sup>也遵循这些模块内容来设计,例如,DDI 针对数据收集模块设计了 2.3 小节、2.5 小节和 4.8 小节的元素。通过将研究数据生命周期与元数据标准元素结合设计,DDI 可以使用该模型作为数据创建者、数据管理者和数据用户的培训工具,并组织 and 规划资源,策划计划和管理策略。

此外,为了在科研过程和数据管理的同时,收集、自动或者半自动地生成所需的元数据信息,有必要构建元数据基础设施服务。相应地,研究数据元数据标准的设计也需要与研究数据生命周期结合,为元数据基础设施服务的应用提供前提和一致性。

### 5.3 应用语义数据和语义工具

资源描述框架(Resource Description Framework, RDF)、网络本体语言(Web Ontology Language, OWL)、关联数据(Linked Data)等新技术的



发展,为元数据更深入和更有意义地表达和关联研究数据,更有力地支持数据驱动的科研活动提供了更多可能。研究数据元数据中的语义元数据扮演着两种角色:一是进行数据主题识别,二是作为主题分类依据,支持相似主题链接机制。目前,统一医学语言系统(Unified Medical Language System, UMLS)、国会图书馆主题列表(Congress Subject Headings, LCSH)等一些大型语义工具,已被转换为RDF等编码方式,这就为在元数据中应用这些语义工具,实现对数据的更加灵活和广泛的表达和链接功能提供了支持。例如,我们可以使用主题项的URI或者其他标识方式来代替用自然语言描述研究数据的主题,并且在数据、主题项之间通过主题关系建立RDF的三元组。通过这种方式,可以使得研究数据的发现、分类和关联变得更为智能和准确,避免经常返回无关数据的情况。

目前,应用语义数据和语义工具,发展关联数据形式的大规模研究数据语义元数据<sup>[27]</sup>,探索跨学科的数据发现和关联等<sup>[28]</sup>,都已取得可喜进展,是研究数据元数据标准发展值得期待的未来发展方向。

## 6 启示

研究数据元数据,是当代科学交流活动的重要元素。同时,元数据作为数据资源存放、管理和交换的重要手段,在科学数据共享和数据仓储平台建设中发挥着极其重要的作用,包括数据描述作用、数据发现作用、数据管理作用、数据互换互通作用、资源整合作用、数据引用作用和知识产权保护作用等<sup>[29]</sup>。

目前,各学科领域的研究数据元数据标准设计和实践取得了积极成果,但也面临着不少问题。例如,在跨学科数据分享和发现方面,已有研究数据元数据标准结构不一,且往往只考虑标准之内的统一设计和实施,在标准之间的信息互通和数据互操作上重视不足。满足学科数据交流和数据仓储平台建设的元数据方案时,设计时应着重考虑的目标包括:(1)实现跨学科领域的数据分享实践,具体目标包括模块化、可扩展性、灵活性、综合性、充足性、简易性、数据交换、数据检索、数据存档、数据发布等;(2)支持科研过程和数据管理活动,具体目标包括数据收集、数据处理、数据分发、数据发现和数

据分析等;(3)支持更智能的数据主题分类和实体识别,具体目标包括数据语义化、关联化等。

## 参考文献

- 1 Davenhall C. Digital curation reference manual - Instalment on scientific metadata[EB/OL].[2018-07-05]. [http://www.dcc.ac.uk/sites/default/files/documents/Scientific%20Metadata\\_2011\\_Final.pdf](http://www.dcc.ac.uk/sites/default/files/documents/Scientific%20Metadata_2011_Final.pdf).
- 2 National Science Foundation. Dissemination and sharing of research results[EB/OL].[2018-07-05]. <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>.
- 3 Birte C.D, et al.Ten recommendations for libraries to get started with research data management[EB/OL].[2018-07-05]. [http://libereurope.eu/wp-content/uploads/WGSC\\_20120801.pdf](http://libereurope.eu/wp-content/uploads/WGSC_20120801.pdf).
- 4 ICPSR. Sustaining domain repositories for digital data: a white paper[EB/OL].[2018-07-05]. [http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper\\_ICPSR\\_SDRDD\\_121113.pdf](http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf).
- 5 Qin J, Ball A, Greenberg J. Functional and architectural requirements for metadata: Supporting discovery and management of scientific data[C].International Conference on Dublin Core and Metadata Applications. 2012: 62-71.
- 6 赵华,王健.科学数据元数据功能与内容分析[J].科技管理研究, 2015(17):232-235.
- 7 Data Curation Center. List of metadata standards-disciplinary metadata[EB/OL].[2018-07-05]. <http://www.dcc.ac.uk/resources/metadata-standards>.
- 8 科学数据共享工程办公室.《科学数据共享元数据内容》标准培训教材[EB/OL].[2018-07-05]. <http://www.doc88.com/p-3704312750141.html>
- 9 人口健康科学数据共享标准规范课题组. 人口健康科学数据共享元数据标准(征求意见稿)[EB/OL].[2018-07-05]. <http://www.ncmi.cn/info/108/1052>.
- 10 王国复,徐枫,吴增祥.气象元数据标准与信息发布时间研究[J].应用气象学报,2005,16(1):114-120.
- 11 赵华,王健.元数据标准与我国农业科学数据元数据[J].中国科技资源导刊,2014(5):79-83.
- 12 朱星明,张行南,白婧怡,等.水利科学数据共享元数据理论的应用探讨[J].水利学报,2005,36(8):946-949.
- 13 王卷乐,陈义华.基于元数据管理的地球系统科学数据共享研究[C].中国地理信息系统协会年会,2004.
- 14 中国地震局.地震科学数据共享工程技术标准一元数据编写指南[EB/OL].[2018-07-05]. <https://wenku.baidu.com/view/98bab95c3b3567ec102d8af5.html>.
- 15 吴彬.生态科学数据元数据及其标准研究[J].中南林业科技大学学报,2010,30(12):75-79.
- 16 朱艳华,刘宁,胡良霖.科学数据库核心元数据标准的修订与实践[C].科学数据库与信息技术学术研讨会. 2010.



- 17 Willis C, Greenberg J, White H. Analysis and synthesis of metadata goals for scientific data[J]. Journal of the American Society for Information Science and Technology, 2012, 63(8): 1505—1520.
  - 18 同 5.
  - 19 Dublin Core Metadata Initiative. Dublin core metadata element set-version 1.1: reference description[EB/OL].[2018—07—05]. <http://dublincore.org/documents/dces/>.
  - 20 DataCite Metadata Working Group. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.0. [EB/OL].[2018—07—05]. <http://doi.org/10.5438/0012>.
  - 21 Biodiversity Information Standards. Darwin core terms: a quick reference guide[EB/OL]. [2018—07—05]. <http://rs.tdwg.org/dwc/terms/index.htm>.
  - 22 Robert H, Lars L.C., Adrienne G, et al. Astronomy visualization metadata (AVM) standard-version 1.2 rc1[EB/OL].[2018—07—05]. [https://www.virtualastronomy.org/AVM\\_DRAFTVersion12\\_rlh02.pdf](https://www.virtualastronomy.org/AVM_DRAFTVersion12_rlh02.pdf).
  - 23 Brian E, Jonathan G, Bob D, et al. NetCDF Climate and forecast metadata conventions-version 1.7[EB/OL].[2018—07—05]. <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.pdf>.
  - 24 Data Documentation Initiative. DDI\_3.0\_part\_I\_overview[EB/OL].[2018—07—05]. [https://bitbucket.org/ddi-alliance/ddi-lifecycle/downloads/DDI\\_3\\_0\\_2008-04-28\\_Documentation\\_XMLSchema.zip](https://bitbucket.org/ddi-alliance/ddi-lifecycle/downloads/DDI_3_0_2008-04-28_Documentation_XMLSchema.zip).
  - 25 中国科学院数据应用环境建设和服务项目组. 科学数据库核心元数据标准[EB/OL].[2018—07—05]. <https://wenku.baidu.com/view/db2808365a8102d276a22f51.html>.
  - 26 DDI structural reform group. DDI version 3.0 conceptual model-Data Documentation Initiative[EB/OL].[2018—07—05]. <https://www.ddialliance.org/sites/default/files/Concept-Model-WD.pdf>.
  - 27 Bizer C, Tom H, Tim B.L. Linked data—the story so far[M]. Semantic Services, Interoperability and Web Applications: Emerging Concepts, 2009: 205—227.
  - 28 Bechhofer S, et al. Why linked data is not enough for scientists [J]. Future Generation Computer Systems, 2013, 29(2): 599—611.
  - 29 王国复, 涂勇, 王卷乐, 等. 科学数据共享中的元数据技术研究[J]. 中国科技资源导刊, 2008, 40(1): 30—36.
- 作者单位: 北京大学图书馆, 北京, 100871  
收稿日期: 2018年7月7日  
(责任编辑: 关志英)

## Comparative Study of Research Data Metadata Standards Based on Content Architectural View

Zhu Ling

**Abstract:** Metadata is key for the discovery and reuse of research data (or scientific data). Specific metadata standards or norms have been developed in different subject areas and academic communities both at home and abroad. Compared with the literature resource metadata standard, the standard structure and element design of research data metadata are more complex in most cases, which brings obstacles to interdisciplinary data sharing and discovery. This paper applied a content structure view to compare and analyze the eight typical research data metadata standard elements, identified the content characteristics of different research data metadata standards, and explored the development trend of research data metadata standards for metadata application, and thus provides a reference for the design of metadata application profile.

**Keywords:** Research Data; Scientific Data; Metadata Standard; Content Architectural View.