科学元数据标准的现状、特点与改进建议

邱春艳 陈可睿
（曲阜师范大学传媒学院，日照 276826）

摘要：通过网络调查和对比分析，发现现有科学元数据标准类型丰富多样，涉及领域广泛；不同标准所含元素异中有同；标准之间通过元数据文件格式、映射关系实现对数据共享的支持。进而分析归纳出以下特点：科学元数据标准多面向以定量研究为主的学科领域，且多局限于科学研究或数据管理的某一阶段，所含元素差异明显，在实际应用中多通过开发衍生标准或元素复用扩宽适用范围。经研究，在科学元数据标准开发构建过程中，科学元数据标准应从开放科学建设、数据共享实际需求、科学研究整体性等多个角度综合考虑进行设计和实践应用。

关键词：科学数据；科学元数据；元数据标准；数据共享
中图分类号：G250　DOI：10.3772/j.issn.1673-2286.2022.12.002
引文格式：邱春艳，陈可睿. 科学元数据标准的现状、特点与改进建议[J]. 数字图书馆论坛，2022 (12)：10-18.

随着科学研究的不断深入，数据体量迅速增长，数据内容不断丰富，科学研究呈现出数据密集型的特点。科学数据已经成为最有价值的战略资源之一，是科技创新的关键要素[1]，也是全球科学体系基础设施的重要组成部分。2020年9月，英国政府为实现数据驱动型创新领导者的目标发布《国家数据战略》（*National Data Strategy*），提出四项核心能力和数据领域的五个优先任务[2]。美国国立卫生研究院（National Institutes of Health，NIH）于2020年10月发布《NIH数据管理和共享最终政策》（*Final NIH Policy for Data Management and Sharing*），确立了最大限度公开和共享由NIH资助或开展的科研项目所产生的科研数据[3]。无论是数据创新驱动、数据开放共享还是开放科学建设，都离不开科学数据资源的支撑，这就需要对科学数据开展更高效的组织和管理。科学元数据，一些学者亦称之为科学数据元数据，它是对科学数据外部特征和内部特征的详细描述[4]，能够对科学数据的识别、追踪、获取等提供线索，为科学数据共享提供支撑。科学元数据标准的构建和应用，应当满足领域内的存储要求、资源特点以及用户需求。本文通过网络调查和文献调研，对当前不同学科领域的典型科学元数据标准进行对比分析，归纳总结当前不同领域科学元数据标准使用现状与特点，以期为进一步推动我国科学数据的开放共享提供参考。

1 科学元数据标准的概念及分类

1.1 科学元数据标准的概念

元数据（Metadata）是描述信息资源或数据对象的数据，其最本质、最抽象的定义就是关于数据的数据（Data about Data）[5-7]。而元数据标准是构建元数据体系的重要环节，是评价管理数字资源的结构化数据，集数字资源的属性、图形、数值、影像等多种元数据元素，用于对数字资源进行生产管理和加工处理，记录数据处理过程所产生的数据参数[8]。

英国数据策展中心（Digital Curation Centre，DDC）指出科学元数据“是一系列诠释科学数据的文件，提供必要的辅助信息来发现、解释、理解、评估和使用数据”[9]。黄如花等[10]认为科学元数据“以科学数据为描述对象的元数据，是对科学数据开展描述、组织、出版等工作的重要工具”。李善青等[11]认为科学元数据“是对科学数据的用途、结构、过程和环境的规范化描述，使科学数据更容易被发现、解释、理解、评估

和共享使用”。由此看出，科学元数据很大程度上是对研究数据的描述，是聚焦于具体学科实践的一种元数据，其构建需要特定的学科领域知识作为支撑。因此，科学元数据标准是一系列诠释科学数据的结构化数据集，提供必要的信息线索来发现、解释、理解、评估和使用科学数据，是对科学数据展开描述、组织和揭示等工作的重要工具，以科学数据重用和解释为目的。

1.2 科学元数据标准的分类

通过对上述科学元数据标准概念的解读，本文认为凡是用于科学数据资源描述、组织、管理和揭示的元数据标准均属于科学元数据标准的范畴。对既有的科学元数据标准实践进行调查可知，目前科学元数据标准主要有两种情形：一种为将通用领域的元数据标准用于科学数据资源，以综合性数据资源平台为主；另一种为科学数据资源管理与共享实践中新制定或生成的元数据标准，以学科领域数据资源平台为代表。

此外，目前对于科学元数据标准的分类，应用较为广泛的是将其划分为通用元数据标准和学科领域元数据标准[12,13]。通用元数据标准可以对一般研究数据进行描述，不针对特定学科领域的科学数据，如Dublin Core、DataCite Metadata Schema等标准。面向学科领域的元数据标准是针对某一特定学科领域内科学数据特点构建的具有较强专指性的元数据标准，如生物科学领域的Darwin Core、地球科学领域的ISO 19115等。因此，本文将科学元数据标准的类型划分为通用元数据标准和学科领域元数据标准两类。

2 科学元数据标准的现状分析

科学技术的迭代更新带来科学数据体量的快速膨胀，1991年Diederich等[14]提出倡议要求针对不同学科领域的特点构建面向学科领域的元数据标准，以便于科研人员更加精准、高效地查找信息。科学元数据标准逐渐呈现学科领域多、研究视角广的趋势。本文主要对DDC[15]、Re3Data.org[16]以及FAIRSharing[17]中收录的生物科学、地球科学、物理科学和人文社会科学等不同学科领域的科学元数据标准进行多角度对比分析，相关标准如表1所示。

表1 不同学科领域科学元数据标准

生物科学	地球科学	物理科学	人文社会科学
Darwin Core	ISO 19115	Astronomy Visualization Metadata（AVM）	Data Documentation Initiative（DDI）
Access to Biological Collection Data（ABCD）	Content Standard for Digital Geospatial Metadata（CSDGM）	International Virtual Observatory Alliance Technical Specifications	Open Archives Initiative Object Reuse and Exchange（OAI-ORE）
W3C HCLC Dataset Description	Directory Interchange Format（DIF）	Crystallographic Information Framework（CIF）	Text Encoding Initiative（TEI）
Observ-OM	Astronomy Visualization Metadata（AVM）	Core Scientific Metadata Model（CSMD-CCLRC）	Encoded Archival Description（EAD）
DatA Tag Suite	Observations and Measurements	NeXus	MIDAS-Heritage
Dryad Repository Metadata	Agricultural Metadata Element Set（AgMES）	Observations and Measurements	城市地理空间信息共享与服务元数据标准（CJJ/T144-2010）
Open Microscopy Environment XML（OME-XML）	Common Information Model（CIM）	Flexible Image Transport System（FITS）	科技平台用户元数据（20132793-T-306）
Protein Data Bank Exchange Dictionary and the Macromolecular Crystallographic Information Framework（PDBx/mmCIF）	Climate and Forecast（CF）Metadata Conventions	Standard for Documentation of Astronomical Catalogues（SDAC）	标准文献元数据（GB/T22373-2021）
Minimum Information for Biological and Biomedical Investigations（MIBBI）	气象数据集核心元数据（QX/T39-2005）	Macromolecular Crystallographic Information Framework（mmCIF）	政府信息公开目录系统实施指引

续表

生物科学	地球科学	物理科学	人文社会科学
ISA-Tab	海洋信息元数据（HY/T136-2010）	中国科学院科学数据库核心元数据标准	基本数字对象描述元数据标准
Ecological Metadata Language（EML）	地理信息元数据（GB/T19710-20005）	-	-
Genome Metadata	土壤科学元数据	-	-
农业生产技术信息元数据标准（APTIM）	-	-	-
中国农业科学院农业科技信息核心元数据标准（ASTICM）	-	-	-

2.1 科学元数据标准的分布

2.1.1 科学元数据标准的学科领域分布

面向学科领域的科学元数据标准是根据学科领域特点和所属学科科研人员实际研究需要构建的，主要涉及可以产生大量数据集的自然科学领域，如物理化学、生物科学、地球科学、农学林学等学科。而在社会科学和人文学科领域的科学元数据标准数量相对较少，调查发现主要涉及档案学、统计学、社会行为、经济学等学科。

2.1.2 科学元数据标准的开发组织分布

科学元数据标准的构建和发展离不开国际组织的支持。如制定DataCite Metadata Schema[18]的DataCite国际联盟（the DataCite Metadata Consortium），开发DDI[19]的国际数据文档倡议联盟组织（Data Document Initiative Alliance，DDIA）等，这类机构通常是某一行业的权威国际组织，对行业内数据的描述和操作进行规范。能够在保证行业内部数据互操作的同时，便于不同行业间的数据互访问。

除了国际组织对所属学科领域科学数据描述的规范，政府部门、研究机构和高校在科学数据描述的标准化和一致性进程中也发挥着不可替代的作用。例如，构建目录交换格式（Directory Interchange Format，DIF）标准[20]的美国国家航空和宇宙航行局（National Aeronautics and Space Administration，NASA），建立Dryad科学数据仓储元数据标准的美国国家进化分析中心等机构[21]等。随着数据共享观念的普及，越来越多来自不同领域、不同组织的科研人员加入科学元数据标准的构建中。

2.2 科学元数据标准的元素设置与语义约束

通用科学元数据标准在元素的描述、结构和约束性等方面的一致性要求为跨领域科学数据共享奠定了基础，而当前大多数科学元数据标准的建设实践主要围绕特定独立学科领域展开，且科学元数据的具体应用场景和建设机构存在用途与目的的差异，因此其设计需求和应用目标大不相同，使得不同标准所包含元素存在明显差异。但从语义层面，不同科学元数据标准的元素存在相通性，为基于元数据的数据共享提供可能。

2.2.1 通用元数据标准的元素设置

通用元数据标准的元素设置多从科学数据资源的生命周期揭示与数据来源追溯、数据资源的重用性以及互操作性等多个角度综合考虑，较为典型的当属DC元数据和DataCite Metadata Schema（以下简称DataCite）。1995年制定的DC元数据包含15个元素，能够实现对不同学科领域科学数据的描述，并在其发展过程中引入了限定词的概念，进一步细化对科学数据元数据的描述[22]；最新的DataCite 4.4版本[23]明确10个必备元素、6个推荐元素以及8个可选元素，并使用永久性唯一标识符（DOI）辅助科学数据的检索、共享和重用等。英国考古数据服务（UK Archeology Data Service，UK ADS）和中国科学院国家科学数字图书馆均使用DC元数据对数据进行收集、描述、编目和保存等，从数据描述的基础层保障了不同学科背景和应用目

的的研究人员对数据资源的操作。

同为通用型元数据标准，DC元数据与DataCite均包含Title、Creator、Subject、Publisher、Contributor、Date、Format、Identifier、Rights、Language、Description、Type等12个元素，重合率为37.5%（见表2），涵盖对科学数据资源的描述、权限、管理等多个方面。除此之外，DC的Type元素和DataCite的Resource Type元素，是对科学数据类型的不同描述，即相同语义在两个标准中呈现为相近但不同的元素名称。

表2 典型通用元数据标准元素一览表

名 称	最初版本发布时间	最新版本发布时间	下载格式	映 射	元 素
DC元数据	1995年	2012年	PDF	UK AGMAP；DataCite Metadata Schema；PROV；DDI；MARC	Title；Creator；Identifier；Subject and Keywords；Publisher；Contributor；Date；Language；Format；Description；Type；Rights；Source；Relation；Coverage
DataCite Metadata Schema	2009年	2021年	PDF、XSD、XML	Dublin Core；IDF Metadata Kernel；OECD；DDI	Title；Creator；Identifier；Subject；Publisher；Contributor；Date；Language；Format；Description；Resource Type；Rights；Publication Year；Alternate Identifier；Related Identifier；Size；Version；Geo Location；Funding Reference

2.2.2 面向学科领域科学元数据标准的元素差异

面向学科领域的科学元数据标准受学科性质影响，元素设置既具有一致性，又存在较为明显的差异。从元素数量上看，生物科学领域最新版本的Darwin Core包含记录级元数据、发生信息、材料样品、事件、位置、地理上下文、标识、分类和补充词汇等不同类型共172项元素；地理科学领域ISO 19115包含元数据包数据字典和数据类型信息409项元素，共计13个元数据包；物理科学领域AVM包含55个元数据元素；人文社会科学领域DDI包含复合元素、简单类型、元素组和属性组共1 181个元素。由此可见，元数据元素数量设置从几十到上千个不等，而元素数量的多少与科学元数据标准所要描述对象的详细程度密切相关。

除了元素数量上的明显差异，相同语义内涵的元素表示也存在差异。以对时间的描述为例，生物科学领域Darwin Core使用Event Date元素描述事件发生的日期时间，形式为YYYY-MM-DDThh:mm；地理科学领域ISO 19115利用title元素的时间子元素描述时间，形式为DD/MM/YYYY；物理科学领域AVM使用YYYY-MM-DD形式的Date元素描述日期；人文社会科学领域MIDAS-Heritage使用DD-MM-YYYY形式描述与日期相关的Date元素。由此可见，科学元数据标准在具体的语义约束上具有明显差异。

特定学科领域的科学元数据标准存在与其他学科领域标准不同的特有描述元素项，从而更为明显地体现学科主题和研究内容特征。如生物科学领域ABCD、Darwin Core以及EML都使用Taxon元素记录描述对象所属的生物学类群；地球科学领域的ISO 19115和DIF均使用Platform元素描述支撑传感器的结构；物理科学领域的AVM和CSMD均使用Facility元素描述实验过程所使用的工具；而人文社会科学领域MIDAS-Heritage和EDA均使用Archival元素描述对象地理位置、来源和类型等相关信息。

由于相同领域不同元数据标准建立的角度和描述目的不同，其元素之间也存在差异。生物科学领域，ABCD从生物多样性角度构建用于访问和交换有关标本和观察的数据[24]，Darwin Core基于分类群通过提供标识符、标签和定义来促进生物多样性信息的共享[25]。由于标准之间描述的侧重点不同，相较于ABCD，Darwin Core在测量角度包含measurement ID、measurement Remarks、measurement Type、measurement Unit等元素，对测量分类群进行更加深入的描述。

2.3 科学元数据标准的文件格式

关联开放数据的提倡者Tim Berners Lee在2010年5月提出开放数据五星评价标准。该标准又称为数据复用的马斯洛金字塔，第一层是指基本需求的满足，而后逐层深入，第五层表示数据开放的最佳状态[26]。其中，一星标准指数据以任何格式存在于Web上；二星指数据的形式为结构化数据；三星是以非专有的数据开放格

式，如CVS；四星是指使用URI标识指向信息标的；五星是指使用数据链接其他数据并提供上下文信息。根据上述标准，JSON、RDF、LOD等格式能够使用URI表示信息，可以提供数据链接到其他数据，满足四星甚至五星的数据开放要求，是元数据文件格式中较好的实践。

调查可知元数据文件格式中以RDF、XML格式为主，部分提供JSON格式，均符合较高水平的数据开放文件格式（见表3）。元数据的文件格式随着科学数据共享、复用的需求不断丰富，例如DCAT从支持XML、RDF格式到支持各种特殊格式[27]。

表3 元数据标准的元数据文件格式一览

元数据标准名称	元数据文件格式
Dublin Core	XML、RDF
Darwin Core	XML、RDF、JSON
CERIF	XML
DCAT	XML、RDF
RDF Data Cube Vocabulary	RDF
DataCite Metadata Schema	XML
Observations and Measurements	XML、JSON
ABCD	XML
Darwin Core	RDF、XML
MIBBI	XML
ISO 19115	XML
CDMD	XML、HTML、PDF
DDI	XML、RDF、JSON、UML/XMI
SDMX	XML、JSON、CSV

2.4 科学元数据标准的映射关系

科学数据往往由组织机构通过科学数据仓储完成存储、管理、检索和共享，由于学科背景及数据属性的不同，科学数据多为异构资源，众多科学元数据标准也存在很大差异。现阶段仓储库主要分为结构异构和语义异构两种类型。前者多是由于存储结构不同导致的，后者主要是由于相同概念在不同仓储库所使用的数据结构不同导致的，也可以理解为不同仓储库使用不同的数据结构表达同一概念[28-29]。语义异构是数据共享中要重点解决的问题，而映射关系则是解决这一困难的有效工具。

科学元数据标准能够保证科学数据在所属仓储库中的一致性，通过映射关系科研人员能够实现不同仓储库科学数据的互操作。经调查，DDI使用DataCite和DC元数据两种标准进行映射，能够实现通用学科领域、生物领域和人文社会科学领域三个不同学科领域仓储库中科学数据的互操作，增强跨领域的学科交流。DDI与DataCite的相互映射更是极大地方便了科研人员使用通用元数据标准对专业学科领域科学数据的访问（见表4）。

表4 元数据标准映射关系一览

元数据标准名称	映射标准
DataCite Metadata Schema	Dublin Core；IDF Metadata Kernel；OECD；Data Documentation Initiative
Dublin Core	UK AGMAP；DataCite Metadata Schema；PROV；Data Documentation Initiative；MARC
Observations and Measurements	XML Implementation
ABCD	Darwin Core；Eurisco Descriptors（Draft）；OECD Minimum Data Set
Darwin Core	ABCD
EML	ISO 19139（XML Schema for ISO 19115）
ISA-Tab	isatools conversions to MAGE-TAB；OWL；RDF Data Cube Vocabulary
PDBx/mmCIF	PDBML/XML；RDF Data Cube Vocabulary
DIF	ISO 19115
FGDC/CSDGM	ISO 19115；Directory Interchange Format
ISO 19115	FGDC/CSDGM；Directory Interchange Format；UK AGMAP；NetCDF
Observations and Measurements	XML Implementation
SPASE Data Model	OAI
DDI	DataCite Metadata Schema；Dublin Core

3 科学元数据标准的特点

3.1 多面向以定量研究为主的学科领域

当前数据共享以自然科学领域为主，社会科学领

域也越来越重视数据共享。这与学科研究及输出结果的形式有关，自然科学多通过定量研究产生数据资源，社会科学更多使用定性研究或者定性与定量相结合的方式获取研究数据。通过上述调查得知，面向学科领域的科学元数据标准主要涉及产生大量数据的自然科学领域，如物理化学、生物科学、地球科学、农学林学等。调查所涉及的元数据标准共47种，面向学科领域的元数据标准共37种，约占78.7%，其中基于自然学科的元数据标准共30种，约占元数据标准总数的63.8%。受学科性质影响，自然科学领域研究多为量化研究，研究成果多以数据形式呈现，形成大量数据集合，且自然科学各研究领域之间差异大，对数据粒度有一定要求，因此自然科学领域科学元数据标准元素项数量大、描述维度丰富。与此类似，社会科学与人文学科领域科学元数据标准多存在于经济学、统计学等以数据为核心研究工具和产出结果、定量研究特点明显的学科领域。

3.2　不同标准所包含的元素存在明显差异

科研人员受学科知识背景以及研究手段的影响，产生不同的数据共享需求。进而对描述科学数据所使用的元数据标准元素设置进行规范和约束。通用元数据标准不局限于某一学科的科学数据存储机构库，应用范围广，可扩展性强。面向学科领域的科学元数据标准受学科性质及实际研究需要的影响，专指性强，应用面窄，元素受适用对象控制明显，与其他领域科学数据建立映射关系比较困难。生物科学、地球科学、物理科学和人文社会科学4个学科领域的科学元数据标准所包含元素差异较大。在相同学科领域，由于科学元数据标准的描述对象、涉及的研究领域、科学研究的过程和方法不同，其所包含的元素也存在明显的差异。如同为生物学领域科学元数据标准，ABCD比Genome Metadata的描述对象范围更为宽泛，Genome Metadata的描述对象更为具体，因此其元素大多数不一致。不同学科领域科学元数据描述角度的差异，使得不同学科领域数据资源在基础特征描述、上下文环境以及附加信息方面无法统一，为跨学科领域科学数据共享带来了一定阻碍。

3.3　多数标准局限于科学研究和数据管理的某一阶段

数据共享多以科研人员研究需求为导向，所以科学元数据标准的制定往往针对某一研究过程或者某一专门领域的数据存储库，具有较强的实际应用性。例如，DDC列举的源于开放档案信息系统参考模型的PREMIS数据字典，包含关于蛋白质、核酸、复杂装配体的3D结构信息的档案库和高分子晶体信息框架的PDBx/mmCIF，用于归档天文数据的SDAC，在档案和手稿存储库中使用XML对文档查找进行辅助编码的EAD等，都表现出科学元数据标准与具体科研实践相结合的特点。

除此之外，数据共享行为可以发生在数据产生到再利用全过程的任意一个或几个阶段。根据科学数据生命周期所含五个阶段[30]发现，现有的科学元数据标准很少覆盖整个科学数据生命周期，往往针对某一个或某几个阶段构建。例如，CERIF是欧盟向其成员国推荐用于记录研究活动信息的标准，Data Package是一种用于交换数据的通用包装格式，QuDEx用于数据归档和交换的定性数据交换模型等。科学数据共享不只涉及科学研究完成之后的数据共享，更加注重对数据的溯源，要求实现数据资源的完整、一致和可追溯。科学元数据标准如果不能实现对数据完整生命周期的揭示与记录，则难以支持研究过程的完整性和流畅性，也给科研人员的数据操作带来一定的困难。

3.4　通过多种方式扩大标准适用范围

数据开放共享要求科学数据拥有更高的可访问性和互操作性，因此越来越多的组织机构注重科学元数据标准的扩展和改进。DCC列举的4类面向学科领域的元数据标准中，生物科学拥有扩展标准14种，地球科学拥有18种，物理科学拥有6种，人文社会科学领域拥有4种。基于XML的ABCD模式，主要用于发布丰富的自然历史收藏标本数据，试图全面且高度结构化地支持来自各种数据库的数据[31]，并在实践中衍生出3种扩展标准，扩展DNA数据的ABCDDNA、扩展地球科学数据的ABCDEFG以及对植物标本数据存储和运输的HISPID。

此外，元数据的复用即复用一个或多个其他元数据标准中的元素来共同描述复杂资源，能够提高不同学科之间元数据的可比性、互访问性和可转换性[32]，不仅能够扩大元数据标准的使用范围，而且能够为元数据标准的互操作性提供可行基础。例如，DatA Tag Suite对DataCite、W3C HCLS Dataset Description、Common

Metadata Elements for Cataloging Biomedical Datasets等多个标准中的元素进行复用；W3C HCLS Dataset Description对Data Catalog、Dublin Core Metadata Types、PROV Ontology等标准中的元素进行复用。

4 科学元数据标准建设与应用的改进建议

4.1 满足数据共享需求

随着科学技术不断发展，各领域数据量急剧增多，新的科学数据共享平台和科学数据集不断涌现，需要应用发展成熟的元数据标准对科学数据进行管理，以提高科学数据的可访问性、互操作性和重用性。机构自定义元数据标准存在应用局限，不利于科学数据的可发现和可获取性。因此，正如哈佛-麻省理工数据中心（Harvard-MIT Data Center）以DDI数据标准为基础进行改进扩展来建设的Dataverse[33]一样，组织机构可首先明确自身科学数据描述的目标，根据现有科学元数据标准的优缺点和可扩展性选择适合自身特点和需求的科学元数据标准，以便科研人员更快更好地接纳和利用，提高元数据的认可度，推动数据共享体系的完善。

4.2 推进通用化设计与应用

通过Tenopir等[34]的调查结果得知，学科领域内丰富的元数据标准应用率较低。专指性强的元数据标准使得不同数据集之间元数据字段差异较大，共有核心字段数量不足会对数据质量和数据获取造成一定影响，也会给未来的数据整合带来不便。构建通用型元数据标准为解决标准之间的差异化提供了思路。

根据刘峰等[35]关于科学元数据标准通用化设计研究中对地学、生物、物理、空间和社会与人文等多个学科领域共22种元数据标准中的元素进行统计，得出通用元数据项33个，可以根据不同学科领域的科学研究需要对通用元数据项进行选择性应用。除此之外，还可以通过受控词表对元数据标准中元数据项所含的词语进一步规范化处理，以提高不同领域内科学数据的可比性、共享性和互操作性。与此同时，可以选择使用DC或DCAT等认可度高、流通性强的元数据标准进行映射关系的构建，使不同领域的科学数据能够更好地互通。

此外，还需要对元数据标准的文件格式进行规范，使文件格式能够更加符合开放数据五星评价标准，尽可能选择互操作性强的格式，并且尽可能支持更丰富的文件格式，以支持与其他科学数据资源在文件格式层面的可关联性，对数据共享过程进行优化，满足不同科研人员的需求，最大程度地释放科学数据的价值。

4.3 考虑科学研究的整体性

科学元数据标准的构建应当考虑数据资源的完整生命周期，以保证科学数据从研究到利用过程的连贯性和科学数据描述的一致性。根据调查，英国研究委员会中央实验室委员会（Council for the Central Laboratory of the Research Council，CCLRC）构建的CSMD是一种以研究数据为导向的模型[36]，它涉及分析物质结构并且支持跨学科通用[37]，还可以在整个科学研究流程中收集数据。收集到的数据可以根据元数据标准中的规范建立深层次、多角度的科学数据特征标识，以提高科学数据描述的准确性以及检索查询的效率。由此，同一项目中的科研人员可打破不同研究阶段的数据共享壁垒，提升研究连贯性和流畅性，形成良好的研究循环。

5 结语

元数据标准是科学数据开放共享必不可少的支撑，对科学数据的高质量描述利于提高科学数据互操作性、整体性和重用性。随着数据量的增加，科学元数据标准的开发和应用越来越受到重视。通过对国内外科学元数据标准进行调研发现，科学元数据标准多面向以定量研究为主的学科领域，且多局限于科学研究或数据管理的某一阶段，所含元素差异明显，在实际应用中多通过开发衍生标准或元素复用扩宽适用范围。语义网技术、关联技术、本体技术等技术在元数据标准领域发挥的积极影响有利于更好改进现有元数据标准在新兴领域的应用，也有助于进一步构建更为通用的元数据标准。为了顺应科学数据开放共享的趋势，推动更加透明化、科学化、互操作性强的科学数据生态体系的构建，可以从开放科学建设、数据共享实际需求、科学研究整体性等多个角度综合考虑进行科学元数据标

准的设计和实践应用。

参考文献

[1] VAN VLIJMEN H，MONS A，WAALKENS A，et al. The need of Industry to go FAIR [J]. Data Intelligence，2019，1（1）：276-284.

[2] Department for Digital，Culture，Media & Sport.National Data Strategy [EB/OL]. [2022-10-09]. https://www.gov.uk/government/publications/uk-national-data-strategy/national-data-strategy.

[3] Final NIH Policy for Data Management and Sharing [EB/OL]. [2022-10-10]. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html.

[4] 赵华. 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.

[5] BAGLEY P. Extension of Programing Language Concepts [M]. Philadelphia：University City Science Center，1968.

[6] AHRONHEIM J R. Descriptive metadata：emerging standards [J]. Journal of Academic Librarianship，1998，24（5）：395-403.

[7] KAREN C. Understanding metadata and its purpose [J]. Journal of Academic Librarianship，2005，31（2）：160-163.

[8] 孙晓菲，韩子静，曹玉霞，等. 数字时代的元数据实践 [M]. 杭州：浙江大学出版社，2013：1-2.

[9] DAVENHALL C. Digital curation reference manual-Instalment on scientific metadata [EB/OL]. [2022-10-11]. https://www.dcc.ac.uk/sites/default/files/documents/Scientific%20Metadata_2011_Final.pdf.

[10] 黄如花，邱春艳. 国内外科学数据元数据研究进展 [J]. 图书与情报，2014（6）：102-108.

[11] 李善青，郑彦宁，赵辉，等. 大数据背景下科学元数据的重要问题研究 [J]. 科技管理研究，2019，39（18）：184-188.

[12] 徐坤，蔚晓慧，毕强. 基于数据本体的科学数据语义化组织研究 [J]. 图书情报工作，2015（17）：120-126.

[13] DAVENHALL C. Scientific Metadata [EB/OL]. [2022-11-15]. http://www.dcc.ac.uk/resources/curation-reference-manual/chapters-production/scientific-metadata.

[14] DIEDERICH J，MILTON J. Creating domain specific metadata for scientific data and knowledge bases [J]. IEEE Educational Activities Department，1991，3（4）：421-434.

[15] Metadata Standards [EB/OL]. [2022-10-14]. http://www.dcc.ac.uk/resources/metadata-standards.

[16] Re3Data.org [EB/OL]. [2022-10-04]. https://www.re3data.org/.

[17] FAIRSharing [EB/OL]. [2022-10-04]. https://fairsharing.org/.

[18] DataCite Metadata Schema [EB/OL]. [2022-10-20]. https://www.dcc.ac.uk/resources/metadata-standards/datacite-metadata-schema.

[19] GONZALEZ A N，CAMPBELL J，DUNN P，et al. Data discovery with DATS：exemplar adoptions and lessons learned [J]. Journal of the American Medical Informatics Association： JAMIA，2018，25（1）：13-15.

[20] Directory Interchange Format [EB/OL]. [2022-10-20]. https://idn.ceos.org/.

[21] Dryad [EB/OL]. [2022-10-09]. https://datadryad.org.

[22] Dublin Core Metadata Element Set [EB/OL]. [2022-10-22]. http://dublincore.org/documentts/dces/.

[23] DataCite Metadata Schema 4.4 [EB/OL]. [2022-10-20]. http://schema.datacite.org/meta/kernel-4.4/.

[24] Access to Biological Collection Data（ABCD）[EB/OL]. [2022-10-20]. https://abcd.tdwg.org/.

[25] Darwin Core [EB/OL]. [2022-10-26]. https://www.tdwg.org/standards/dwc/.

[26] 5 Star Open Data [EB/OL]. [2022-10-28]. https://5stardata.info/en/.

[27] Data Catalog Vocabulary（DCAT）[EB/OL]. [2022-10-23]. https://dvcs.w3.org/hg/gld/raw-file/default/dcat/index.html.

[28] 王新，张圆圆，许苗，等. 基于异构数据集成技术的农业信息综合管理网络平台开发 [J]. 农业工程学报，2017，33（23）：211-218.

[29] 曹旻，陈盼盼. 异构材料数据集成系统方案 [J]. 计算机工程与设计，2016，37（10）：2826-2831，2843.

[30] 张贵兰，王健，潘尧，等. 科学数据共享服务模式及其演化研究 [J]. 情报理论与践，2022，45（2）：70-77.

[31] ABCD Data Schema [EB/OL]. [2022-10-20]. https://www.biocase.org/products/schema_repository/.

[32] 李丰梅，盛梅. 基于XML语言的医学期刊元数据复用 [J]. 中华医学图书情报杂志，2003（6）：36-38.

[33] Dataverse [EB/OL]. [2022-10-20]. https://dataverse.harvard.edu/.

[34] TENOPIR C，ALLARD S，DOUGLASS K，et al. Data sharing by scientists：practices and perceptions [J]. PloS ONE，2011，6（6）：1-21.

[35] 刘峰，张晓林. 科学数据元数据标准述评及其通用化设计研究[J]. 现代图书情报技术，2015（12）：3-12.

[36] SUFI S，MATTHEWS B，DAM K. An Interdisciplinary Model for the Representation of Scientific Studies and Associated Data Holdings[J]. UK e-Science All Hands Meeting，2003：103-110.

[37] Core Scientific Metadata Model[EB/OL]. [2022-10-04]. http://icatproject-contrib.github.io/CSMD/.

作者简介

邱春艳，女，1987年生，博士，副教授，研究方向：科学元数据、科学数据共享。
陈可睿，女，1999年生，硕士研究生，通信作者，研究方向：科学数据共享，E-mail：chenkr2017@163.com。

The Status Quo, Characteristics and Suggestions for Improvement of Scientific Metadata Standards

QIU ChunYan CHEN KeRui
(School of Communication, Qufu Normal University, Rizhao 276826, P. R. China)

Abstract: Through the network survey and comparative analysis, it is found that the existing scientific metadata standards are rich and diverse, involving a wide range of fields. Different metadata standards contain elements that are both similar and different. Standards support data sharing through metadata file format and mapping relationship. Furthermore, the following characteristics are summarized: the scientific metadata standards are mainly oriented to the subject field of quantitative research, and are mostly limited to a certain stage of scientific research or data management, with obvious differences in the elements they contain. In practical applications, the scope of application is expanded by developing derived standards or element reuse. According to the research, in the process of developing and constructing the scientific metadata standard, the scientific metadata standard should be designed and applied comprehensively from the perspectives of open science construction, the actual demand of data sharing and the integrity of scientific research.

Keywords: Scientific Data; Scientific Metadata; Metadata Standard; Data Sharing

（收稿日期：2022-11-21）