

通用型科学元数据标准研究

浦燕妮, 刘琪, 耿睿

(北京师范大学政府管理学院, 北京 100875)

摘要: 科学数据在科学研究第四范式的研究时代被视为具有重要的科研价值, 科学元数据在数据管理生命周期中的重要作用也逐渐受到重视。大量的科学数据亟需专门且具有统一标准的元数据进行组织和管理。国外的通用型科学元数据标准已经发展较为成熟, 而国内的相关标准还处于建设之中。本文分析国外较为成熟的元数据标准的建设目的、功能构成, 以及主要标准元素等内容, 为国内建设通用型科学元数据标准提供可参考的建议。

关键词: 科学元数据; 元数据标准; 通用型元数据标准

中图分类号: G201

DOI: 10.3772/j.issn.1673-2286.2016.12.006

1 绪论

自数据密集型科学出现以来, 科学数据成为继科学文献之后重要的研究资源, 对科学数据的有效管理成为促进和保障科学研究和交流活动的重要内容, 并由此引发科学数据管理的研究和实践热潮。

科学数据来源于对科学研究活动的事实记录(包括数值、文本记录、图像和声音等形式), 是被科学团体所共同接受的、对研究结果有用的数据^[1], 是科学界普遍认同的能对研究成果进行验证的必要材料^[2]。科学数据既包括在科学研究过程中产生的、能存储在计算机中的任何数据, 也包括能转换成数字形式的非数字形式数据^[3]。如传感器读取的数据、遥感勘测数据、调研结果、神经图像、实验数据及来自测试模型的仿真数据等^[4]。作为描述科学数据的工具, 科学元数据(Scientific Metadata)在数据发现、数据使用和数据重用利用及存储等关键环节有重要作用, 元数据标准为科学数据的存储、整序、组织、索引以及检索等一系列关键问题提供解决方案。

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^[5]。在很大程度上, 科学元数据基本是对与研究相关的所有

信息的描述, 这些描述以重用和解释数据为目的^[6]。数据用户群体的学科背景往往是小而专, 其来自某一个特定的学科或者研究领域。因此, 制定科学元数据标准通常需要特定的专业知识^[7], 这导致科学元数据多以描述对象所处的专业领域进行划分。目前, 应用比较广泛的划分方式是将其划分为学科领域元数据标准和通用型元数据标准^[3,5]。例如, 英国数字监护中心收集6种不同学科的元数据标准。据统计, 77%的科学元数据标准是特定学科领域元数据标准或指向统计、行为描述等特定领域的元数据^[8]。尽管存在丰富的元数据标准, 科研人员的使用情况却不乐观。Tenopir等的调查显示, 56%的研究人员未使用过元数据标准, 22%的研究者仅用过实验室的元数据标准^[9]。由此可见, 学科领域内复杂的元数据标准体系虽然丰富, 但实用性不强。随着科学数据管理的不断推进, 不同标准的使用也会造成数据交换和统一管理的困难。因此, 实现跨学科或者不受学科限制使用的通用型元数据标准备受重视。对比单一学科的元数据标准, 通用型元数据标准可以涵盖多个学科的数据特点, 覆盖学科范围广、适应性强, 有利于进行数据交换。

目前, 国外通用型科学元数据标准已经从标准制定发展到实际应用阶段。英国、美国等国家的多个数据管

理机构已开始使用自身研制的通用型元数据标准。在国内,研究者对通用型科学元数据标准在理论研究和技术应用等方面进行研究^[10-11]。胡芳^[12]、赵华等^[13]分析国外典型科学元数据标准以及科学数据仓储中应用的元数据方案,为国内元数据标准建设提供建议;刘峰等在对6种典型的学科类型元数据标准进行概述、分析和统计的基础上,进一步构建通用型元数据规范^[14];常颖聪等尝试以植物学基因表达实验为例进行科学实验数据元数据模型研究^[15];徐坤等通过研究科学数据的元数据模型,构建科学数据本体,将数据组织置于知识环境中^[3];周波提出针对高校科学数据的通用元数据设计原则^[16]。不过,以上研究基本是基于国外成熟学科元数据标准,进一步明确科学元数据标准建设的规范和功能,尚未对通用型元数据标准进行直接地分析探讨。本文重点对国外通用型科学元数据标准进行研究,为国内相应的元数据标准建设提供建议。

2 通用型科学元数据标准的内容

通用型科学元数据标准在适用范围和内容元素等方面与学科元数据标准存在差异。在适用范围上,通用型科学元数据标准不受学科领域限制,可以对不同学科的科学数据进行描述和组织,可以被众多综合性的数据仓储机构和数据检索平台所使用,不局限于特定学科的数据存储库。

在元素设置上,通用型科学元数据标准建设的目标是可扩展性、弹性、模块化和可移植性等,满足数据的记录、交换、检索和存档等功能需求。Huang等的调查显示,研究人员使用元数据标准侧重于其综合性、充分性、简易性和抽象性^[17],而诸多学科元数据元素以及大量的、多样化的标准和广泛而差异的命名规则,在一定程度上阻碍元数据标准的推广使用^[18]。Boyd基于分类和组织内容提出科学元数据标准内容的三个层次:第一层次是元数据标准的元素集合,如Dublin Core、CKAN等;第二层次是反应科学环境元数据(Contextual Metadata),涉及基本实体有project、person、organization unit、result、publication、patent、product;第三层次是描述性元数据,提供更细粒度的、详细的信息^[19]。这些详细的信息应涵盖数据状态、数据管理权限、地理、空间信息引用和版本说明等^[20]。例如,英国中央实验室研究理事会研发的核心科学元数据模型(Core Scientific MetaData Model)将

组成元素分为主题、研究描述、数据获取条件、数据描述、数据定位和相关资源等六个方面,从而方便以统一的形式实现科学数据的组织和获取^[8]。

3 通用型科学元数据标准现状

3.1 国外通用型元数据标准

在科学数据管理周期中,元数据标准处于核心地位。国外诸多科学数据管理模型,如ICPSR社会科学数据存档生命周期模型、DataONE科学数据生命周期管理模型,以及英国数据存档数据生命周期管理模型等都十分重视数据归档和阶段管理规范化^[21]。元数据标准是每个数据管理机构必须重视的规范化文件之一。2012年,NISO/DCMI联合研讨会更是以“科学研究数据管理元数据”为主题,对元数据标准进行系统地研讨^[22]。

目前,国外科学元数据标准十分丰富,不仅研制学科性质的元数据标准,而且在制定和使用通用型科学元数据标准。本研究调查了英国数字监护中心(DCC)^[23]、Research Data Alliance(RDA)^[24]以及都柏林核心元数据^[25]等科学元数据使用和管理平台,对目前国外通用型元数据标准进行初步统计,相关情况如表1所示。

除此之外,还有一些关于编码、控制语言等方面科学数据描述的标准。

(1) *Data Catalog Vocabulary (DCAT)*。利用DCAT可以描述数据目录中的数据集。统一的描述使得应用程序更容易从多个目录中获取元数据。该标准有利于将分散出版的目录集成在网站中,从而进行联合数据集搜索。聚集的DCAT元数据可以作为一个数字保存清单。

(2) *RDF Data Cube Vocabulary*。该标准为在网络发布多维数据(如统计数据)提供了一种方式。通过该标准和W3C的RDF(资源描述框架)标准,可以将数据链接到相关数据集和概念实体。模型支撑数据立方体(Data Cube)的词汇和立方体模型,并与SDMX(*Statistical Data and Metadata Exchange*)的数据模型兼容。

(3) *Data Package*是一个数据交换的通用包装格式。虽然支持任意的元数据,但该格式定义了数据包作为一个整体必备的、推荐的和可选的字段,以及所包含的相关资源。

表 1 国外通用型元数据标准情况调查表

标准名称	制定者	应用情况举例
<i>CERIF (version 1.6)</i>	euroCRIS	欧盟成员国可使用
<i>DataCite Metadata Schema</i>	DataCite Consortium	Centre for Digital Music Research Data Repository; Environmental Information Data Centre (EIDC)
<i>Dublin Core Metadata Element Set</i>	Z39.85 Maintenance Agency Advisory Group	ePrints Soton; National Science Digital Library Data Repository
<i>Data Documentation Initiative (DDI)</i>	DDI Alliance	Inter-University Consortium for Political and Social Research (ICPSR); UK Data Archive (UKDA)
<i>Open Archives Initiative Object Reuse and Exchange (OAI-ORE)</i>	Digital Library Federation; National Science Foundation	Resource Sync; Protocol for Metadata Harvesting (OAI-PMH)
<i>Observations and Measurements</i>	Open Geospatial Consortium	Infrastructure for Spatial Information in Europe (INSPIRE); Ocean Data Interoperability Platform (ODIP)
<i>PREMIS Data Dictionary</i>	PREMIS Maintenance Activity	The National Archives (UK); University of California
<i>PROV</i>	W3C Semantic Web Activity	StatJR eBook System; OECD Linked Data

在科学数据元数据标准发展过程中, 诸多数据存储机构也参与到标准的制定中。一些大规模的数据存储服务提供者认为目前很多元数据标准并不能满足数据库的使用, 所以按需制定了元数据描述要求和规则^[26]。这些存储机构很多都是专业领域内涵盖具体学科数据的数据库, 有大量的专业数据, 所以其制定的元数据更专注于特定的数据(如基因数据、GIS数据等)。

国外通用型元数据标准适用于综合性的科学数据管理平台, 基于效率以及管理的角度, 其重要程度应得到足够的重视。但是, 迄今为止美国和ISO没有发布通用的元数据标准。国际元数据标准中, 仅有都柏林核心元素集(15个描述元素)在科学数据管理中被应用, 而且这一标准并不可以完全覆盖科学数据管理周期的所有要求。元数据管理者更多地依赖专业领域内的元数据标准或者是利用第三方数据库服务(如datacite进行数据发现)^[27]。通用型元数据标准并没有得到统一的制定和使用, 仅处于机构和联盟内部标准阶段。

3.2 国内科学元数据标准

随着科学数据管理的兴起, 国内开始逐渐重视元数据标准发展问题。我国自2002年开始实施国家科学数据共享工程, 并制定数据共享工程的核心元数据标

准; 2005年, 科技部提出建立国家科学数据共享工程技术标准, 并发布《国家科学数据共享工程核心元数据内容》和《科学数据共享元数据标准(试行稿)》。《科学数据共享元数据标准(试行稿)》提供了科学数据共享元数据内容标准框架, 定义科学数据共享核心元数据、公共元数据和参考元数据^[28]。中国科学院也先后发布《中国科学院科学数据库核心元数据标准》《人地系统主题数据库元数据标准》《土壤科学数据库元数据标准》等一系列学科数据库元数据标准。2011年, “国家科技基础条件平台建设基础科学数据共享网项目”发布《数据集核心元数据标准》。

目前, 我国主要存在三项有关科学数据元数据的国家标准: GB/T 20533—2006《生态科学数据元数据》、GB/T 19710—2005《地理信息元数据》和GB/T 26499.3—2011《机械 科学数据 第3部分: 元数据》。另外, 部分行业的元数据标准也陆续推出。

国内元数据标准建设侧重于单个学科或部分行业, 科学元数据发展呈现不均衡状态, 通用型科学元数据标准还在研制中, 社会科学领域的科学元数据标准也尚未发展成型, 元数据标准发展较为缓慢。在元数据标准应用阶段, 国内科学数据管理平台应用的元数据标准基本是国外已有的元数据标准或平台建设软件提供的元数据。

3.3 典型通用型科学元数据标准

3.3.1 《都柏林核心元数据元素集》

作为目前唯一的通用型元数据国际标准,《都柏林核心元数据元素集》(*Dublin Core Metadata Element Set, DC*)被广泛认可,其15个DC元素得到广泛传播^[29]。我国与其对应的标准是GB/T 25100—2010。很多科学元数据标准以DC为基础进行扩展,例如:*Dryad Metadata Application Profile*^[30](用于描述同行评议的科学和医学文献等多学科数据)、*ANZLIC Metadata Profile*(方便有效地访问信息资源,特别是地理或空间数据)和*AGLS Metadata Standard*^[31](澳大利亚政府定位服务元数据标准,适用于网络资源的可视性和可用性)等。

科学数据和科研过程极大地影响科学元数据的功能。科学元数据在很大程度上是对与具体研究相关的所有信息的描述,包括对数据资源语境[相关性、质量、限制(权利、成本)]变化的记录以及相关软件和计算机资源的参数数据,最终可以根据元数据记录创建虚拟的研究环境。相对于其他科学元数据标准,DC元数据缺少必备的元素对科学数据产生过程中的信息进行全

面地记录和描述,所以DC中的15个必备元素需要一定地扩展和补充^[32]。

3.3.2 《数据文件倡议》

《数据文件倡议》(*Data Documentation Initiative, DDI*)是一个应用广泛的项目,很多科学数据管理机构(如英国数据服务中心、英国数据档案、美国ICPSR等)都采用该标准进行数据管理。*DDI*可以描述社会、经济、行为和健康科学领域内的观测方法得到的数据,并进行文档发现和互操作。*DDI*是描述社会和行为科学数据文档的标准,该标准提供内容交换和信息保存的格式标准进而可以促进文档间的共同操作(见表2)。目前,*DDI*有两个扩展版本。

(1) *DDI Codebook (DDI version 2)*是比较简单的版本,用于记录简单的调查类型数据,可以进行交换或归档。*DDI Codebook*可以创建目录记录,分为两种: *study-level descriptions*和*codebook-type variable-level descriptions*。这样的记录层次降低调查类的数据记录粒度,使得问卷中的问题可以被描述和重利用。

表 2 DDI 元素

研究层级元素 (Study-level DDI elements)		变量层级元素 (Variable-level DDI elements)
Study ID (Number, DOI)	Date (s) of Collection	Variable Name and ID
Title, Alternate Title	Mode of Collection	Variable Label
Author/Primary Investigator	Universe	Question Text
Bibliographic Citation	Sampling	Descriptive Variable Text
Funding Information	Unit of Analysis	Category Labels and Values (responses)
Abstract	Response Rates	Category Statistics (frequencies)
Keywords/Topic Classification	Weighting Information	Summary Statistics
Series Information	Data Type	Variable Format
Geographic Coverage	Extent of Processing	Notes
Temporal Coverage	Access Conditions/Restrictions	
	Version History	
	Time Method	

(2) *DDI Lifecycle (DDI version 3)*是比较丰富的版本,可用于描述处于数据生命周期任一阶段的数据,涵盖从概念产生、发布和数据重用。它是模块化和可扩展的。*DDI Lifecycle*可以对复杂数据进行层次性和纵向性的描述,创建元数据驱动的调查设计(如问卷设计);

*DDI Lifecycle*的描述可以使整个数据生命周期的任意阶段数据得到重用,如可变编码阶段、数据收集工具使用阶段等;同时,*DDI Lifecycle*创建问题/概念/变量存储、组织/个人的登记表等。*DDI*标准采用XML组织,由国际数据文档倡议联盟组织(DDIA)负责开发^[33]。

3.3.3 《科学数据共享核心元数据标准》

科技部为满足科学数据共享工程对元数据内容的基本需求制定《科学数据共享核心元数据标准》(见表3),其目的是方便科学数据共享数据集的管理,提高数据库建库质量,促进数据加工的规范化、标准化,实现数据交流与共享。该标准主要采用或参考ISO 19115《地理信息——元数据》和《都柏林核心元数据元素集》。该标准定义三个层次的元数据集合:科学数据共享核心元数据(22个元素)、公共元数据(156个元数据元素和实体)和参考元数据(388个元素)。三个层次的元数据集合在内容上是包含关系,又同时分别满足整个工程、各个领域以及具体应用层面的需求。核心元数据面对不同领域、不同类型的数据资源展开统一描述,而公共元数据和参考元数据对于制定领域元数据标准和领域专用元数据标准起着很重要的作用。

表3 《科学数据共享核心元数据标准》

元数据标识符(M)	数据集格式名称(M)
元数据语种(C)	数据集格式版本(M)
元数据字符集(C)	关键词说明(O)
元数据联系方(M)	数据集访问限制(O)
元数据创建日期(M)	数据集使用限制(O)
元数据标准名称(O)	数据集安全限制分级(M)
元数据标准版本(O)	数据集语种(M)
数据集名称(M)	数据集字符集(C)
数据集日期(M)	数据集分类(C)
数据集摘要(M)	数据志说明(C)
数据集负责方(O)	数据集在线资源链接地址(M)

3.3.4 《中国科学院科学数据库核心元数据标准》

2005年,中国科学院计算机网络信息中心和科学数据库中心联合推出《中国科学院科学数据库核心元数据标准》。该标准参考都柏林核心元数据标准、RSLP资源集合描述标准、生态学元数据语言、数字地理空间元数据内容标准。《中国科学院科学数据库核心元数据标准》包括《数据集元数据》和《服务元数据》两个主要部分。

《数据集元数据》是一个面向数据集层次科学数据资源的元数据标准,定义一组通用的元数据模块和元

素,包括数据集描述信息、数据质量信息、数据集分发信息、元数据参考信息、服务参考信息以及结构描述信息等六个主要复合元素模块,还包括范围信息和联系信息辅助模块。用户在遵循扩展机制的基础上,基于此标准开发满足特定学科或主题领域的元数据应用方案。

《服务元数据》是一个面向科学数据库系统中各种服务的元数据标准,定义一个通用的服务描述模型,并基于此模型为所需要的五类服务分别定义描述元素。这五类服务是数据集连接服务、中间件服务、网格服务、WWW服务和FTP服务。通用描述模型抽象了所有服务的描述模式,任何新的服务均可依据该模型来定义其元数据元素。

3.3.5 《数据集核心元数据标准》

2011年3月,国家科技基础条件平台建设基础科学数据共享网项目发布《数据集核心元数据标准》。该标准规定各种需求层次的元数据应用所需要的最小元数据元素(简称“核心元数据”),以及对元数据进行扩展和制定元数据应用方案的规则和方法,以满足各学科领域的特殊需求。该标准适用于科学数据库的编目和描述、组织管理,也可用于数据资源的交换、集成和服务^[34]。

核心元数据面向数据集层次的科学数据资源,主要分为3个模块:描述信息、元数据参考信息和联系信息。

(1) 描述信息。主要包括数据集的名称、简介、创建日期、创建者等基本信息,着重描述数据资源的基本属性,且与DC、*Content Standard for Digital Geospatial Metadata*等国际主要的元数据标准的相应元素基本一一对应(共22个标准元素)。

(2) 元数据参考信息。提供数据集依据的元数据标准名称和版本、有关元数据创建日期和元数据创建及维护者的联系方式等方面的重要信息。数据集用户可以了解到与元数据记录的建立有关的信息,将有助于元数据维护人员进行元数据的修改与维护。

(3) 联系信息。作为该标准的辅助模块,供其他模块的特定元素在需要的时候进行引用。联系信息分为六个方面:联系人姓名(与数据集有关的联系人员名称)、单位、地址、传真、电话以及电子邮件地址。

此外,《数据集核心元数据标准》还提供扩展和应用说明,为元数据的互操作性奠定基础。

4 对通用型科学元数据标准建设的建议

我国的科学数据管理尚处于起步阶段,目前国内外尚未形成比较完备的管理标准,特别是通用型元数据标准,标准建设不完善。虽然有对科技平台资源进行描述的标准,但是还不能完全满足科学数据的管理需求。

国外科学数据管理发展迅速,众多的数据管理机构推出科学元数据标准,再加上已存在的学科领域的元数据标准,最终形成复杂多样的科学元数据标准。管理者自主制定的标准导致数据互操作性以及共享性受到很大影响。随着数据的流动和开放数据共享的发展,标准的差异将会带来更繁杂的数据整合工作。大量的科学数据的产生必然需要高效规范的管理。

国内相关标准还在研制中,不断出现的数据仓储和管理平台迫切需要比较成熟的元数据标准进行管理。目前国内的管理者都是各自选择国外的元数据标准,国外的标准大多也还处于行业和机构标准的水平,这不利于数据交换和统一管理。科学数据的管理将会面临大储量、多学科、规范性等问题。为达到数据分享和统一管理为目的,通用型元数据标准的使用是解决这些问题的一个有效方法。因此,国内相关标准创建可以借鉴国外已有经验,缩短探索过程。

(1) 整合已有科学元数据标准元素,创建统一的元素集合。国内标准在制定时,多采用国际应用范围较广的元数据标准。国外也是如此,如: *Dataverse Network* 符合 *DDI Codebook*、*DC*、*CSDGM*、*MARC LOCKSS*、*OAI* 等标准创建; *Dryad* 的元数据服务也同时符合 *DC*、*Darwin Core*、*Bibliographic Ontology*、*METS/MODS* 等元数据标准。*Willis* 等通过对 16 个元数据标准、4 400 多个独特的元素进行统计分析,发现描述类元素数量较大而且与 *DC* 元素存在重复^[35]。所以对比国内外的标准发现,在元素设计上,很多标准可以继续整合提取,创建涵盖科学数据管理周期的元素集合。

(2) 扩展元数据描述层次,涵盖数据生命周期。国内目前的元数据标准设计主要有 3 类信息:描述数据集、元数据参考信息和数据责任方信息。描述对象集中于数据集,以及对数据集层面的关联信息,缺少数据集层次以下的数据描述和组织。科学元数据描述范围应该涵盖数据产生到引用的全过程,如 *DDI* 涵盖从项目到变量的多层次数据,建立较深层的数据索引。

(3) 以科学数据共享元数据为蓝本进行标准制定,建立通用型元数据标准。科学数据共享元数据在

覆盖信息范围、元素设计层次以及建立实体关联等方面领先于国内其他的元数据标准;而且在元素设计层次上,考虑到核心元数据集合与各领域内元数据集合的关联,在框架结构上起到通用型标准的作用,这也是国外标准未实现的功能。因此,可以此标准为蓝本进行进一步的设计和扩展。在数据标识、名称的权限控制,关联数据,本体和词汇,以及数据引用标准等方面进行规范^[36],特别是扩展描述性元数据支持资源发现和保存的功能。这也是科学元数据与传统元数据标准的不同之处。将不同学科的数据通过统一元数据标准进行关联,加强数据索引的深度和广度,为下一步的科学发现和合作做好准备。

大多数的科学元数据标准为特定学科的元数据标准或特定领域的元数据标准,通用型元数据标准建设不足,广泛认可的国际标准缺乏。国内正处于标准建设期,可从建设初期制定国家标准,避免标准发展的混乱,将会极大地促进科学数据管理的发展。

参考文献

- [1] OECD.OECD Principles and Guidelines for access to research data from public funding[EB/OL].(2007-04)[2016-09-09].<http://www.oecd.org/science/sci-tech/oecdprinciplesandguidelinesforaccesstoresearchdatafrompublicfunding.htm>.
- [2] NIH.NIH Grants Policy Statement[EB/OL].(2011-10-20)[2016-09-09].http://grants.nih.gov/grants/policy/nihgps_2011/nihgps_ch2.htm.
- [3] 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.
- [4] 李晓辉.图书馆科研数据管理与服务模式探讨[J].中国图书馆学报,2011(5):46-52.
- [5] DAVENHALL C.Scientific Metadata[EB/OL].[2016-09-09].<http://www.dcc.ac.uk/resources/curation-reference-manual/chapters-production/scientific-metadata>.
- [6] Australian National Data Service.Metadata[EB/OL].[2016-09-09].<http://www.ands.org.au/working-with-data/metada>.
- [7] Australian National Data Service.Metadata: Working level[EB/OL].[2016-09-09].<http://www.ands.org.au/guides/metadata-working#4.4>.
- [8] MATTHEWS B,SUFI S,FLANNERY D,et al.Using a core scientific metadata model in large-scale facilities[J].International Journal of Digital Curation,2010,5(1):106-118.
- [9] TENOPIR C,ALLARD S,DOUGLASS K,et al.Data sharing by scientists: practices and perceptions[J].PLoS ONE,2011,6(6):1-21.

- [10] 周波,钱鹏.我国科学数据元数据研究综述[J].图书馆学研究,2013(2):7-10.
- [11] 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.
- [12] 胡芳.国外典型科学数据仓储实施的元数据方案及启示[J].图书与情报,2015(1):117-121.
- [13] 赵华,王健.国内外科学数据元数据标准及内容分析[J].情报探索,2015(2):21-24,30.
- [14] 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.
- [15] 常颖聪,何琳.科学实验数据元数据模型构建研究——以植物学基因表达实验为例[J].图书情报工作,2015(13):117-125.
- [16] 周波.高校科学数据元数据方案初探[J].图书馆学研究,2012(1):45-49,53.
- [17] HUANG H,QIN J.Understanding metadata functional requirements in genome curation work[J].Proceedings of the American Society for Information Science and Technology,2013,50(1):1-4.
- [18] QIN J,LI K.How portable are the metadata standards for scientific data?A proposal for a metadata infrastructure[EB/OL].[2016-09-07].
http://jianqin.metadatec.org/wp-content/uploads/2013/08/DC2013-metadata-portability.pdf.
- [19] BOYD D.CERIF tutorial and UK data surgery[EB/OL].[2012-02-16][2016-09-07].http://data.blogs.ilrt.org/2012/02/16/cerif-tutorial-and-uk-data-surgery/.
- [20] Metadata for managing scientific research data NISO/DCMI webinar[EB/OL].[2016-09-07].http://www.slideshare.net/BaltimoreNISO/metadata-for-managing-scientific-research-data.
- [21] 丁宁,马浩琴.国外高校科学数据生命周期管理模型比较研究及借鉴[J].图书情报工作,2013(6):18-22.
- [22] Metadata for managing scientific research data[EB/OL].[2016-08-10].
http://www.niso.org/news/events/2012/dcmi/scientific_data/#about.
- [23] Digital Curation Centre[EB/OL].[2016-08-10].http://www.dcc.ac.uk.
- [24] RDA:Metadata directory[EB/OL].[2016-08-14].https://www.rd-alliance.org/.
- [25] Dublin Core Metadata. Announcements[EB/OL].[2016-08-10].http://dublincore.org/.
- [26] Repository-Developed metadata schemas[EB/OL].[2016-08-10].http://rd-alliance.github.io/metadata-directory/standards/repository-developed-metadata-schemas.html.
- [27] Metadata for research data management[EB/OL].[2016-08-10].http://hangingtogether.org/?p=5616.
- [28] 吴丹.近五年国内外元数据研究进展[M]//陈传夫.图书馆学研究进展.武汉:武汉大学出版社,2010.
- [29] Dublin Core Metadata. Metadata basics[EB/OL].[2016-08-10].http://dublincore.org/metadata-basics/.
- [30] Dryad Metadata Application Profile(Schema)[EB/OL].[2016-08-10].
http://wiki.datadryad.org/Metadata_Profile.
- [31] AGLS Metadata standard[EB/OL].[2016-08-10].http://www.agls.gov.au/.
- [32] DRYAD.Metadata profile[EB/OL].[2016-08-10].https://rd-alliance.org/metadata-principles-and-their-use.html.
- [33] DDI-Community[EB/OL].[2016-09-10].http://www.ddi-alliance.org/alliance.
- [34] 数据集核心元数据标准[EB/OL].[2012-04-13][2016-08-10].http://wenku.baidu.com/link?url=uhRqBPeyax2F_8Ba8lARkyn3bKts-Dz60PgfH8CXdl2jvrqGvCuaksFFfDCRGkb8GtNABBYmToBEXwqOt_BsAeEMIdNRxku8SnCIR2cWub_.
- [35] 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.
- [36] Scientific data application profile scoping study[EB/OL].[2016-09-10].
http://www.ukoln.ac.uk/projects/sdapss/.

作者简介

浦燕妮, 女, 1991年生, 硕士研究生, 研究方向: 科研数据管理, E-mail: puyanni@163.com。

刘琪, 女, 1992年生, 硕士研究生, 研究方向: 信息检索、科研数据管理, E-mail: bnliuqi@126.com。

耿骞, 男, 1965年生, 教授, 研究方向: 信息检索、网络信息管理、管理信息系统, E-mail: gengqian@bnu.edu.cn。

Research on General Scientific Metadata Standards

PU YanNi, LIU Qi, GENG Qian
(School of Government, Beijing Normal University, Beijing 100875, China)

Abstract: Scientific data is regarded as an important scientific research value in the fourth paradigm of science time. The important role of scientific metadata has been paid more and more attention in the life cycle of data management. Especially in the field of social science, a large number of scientific data is in urgent need of specialized and unified metadata standards for the organization and management. Foreign comprehensive scientific metadata standards have developed more mature, while domestic metadata standards are still in the construction. This paper analyzes the construction purpose, function and main standard elements of the foreign metadata standards, and provides a reference for the domestic construction of comprehensive scientific metadata standards.

Keywords: Scientific Metadata; Metadata Standards; General Metadata Standards

(收稿日期: 2016-10-11)