

自然环境试验元数据体系研究

周俊炎^{1,2}, 王竟成^{1,2}, 赵方超^{1,2,3}, 舒畅^{1,2}, 黄伦^{1,2}

(1.西南技术工程研究所, 重庆 400039; 2.国防科技工业自然环境试验研究中心, 重庆 400039;
3.弹药贮存环境效应重点实验室, 重庆 400039)

摘要: **目的** 形成自然环境试验元数据体系, 为数据资源提供完备的描述及合理的规则规范, 为数据库设计开发、数据挖掘分析、数据共建共享等提供支持。**方法** 针对自然环境试验专业特点、数据类型以及应用需求, 借鉴其他领域的元数据标准规范, 梳理分析自然环境试验涉及的数据特征元素集合, 划分自然环境试验元数据类别, 评估元数据作用, 形成自然环境试验元数据体系。**结果** 形成了覆盖数据集描述信息、要素参数信息、数据质量信息、维护信息、限制信息、数据集分发信息、元数据参考信息、数据应用成果信息共8个类别的自然环境试验元数据体系。**结论** 自然环境试验元数据具有资源描述、数据选择、信息检索、管理溯源、解释分析、储存入库等作用, 为数据有效管理和共建共享提供了基础。本文提出的元数据体系为自然环境试验元数据标准提供了参考思路。

关键词: 自然环境试验; 元数据; 元数据确认原则; 元数据类别; 元数据作用; 元数据体系

中图分类号: TP391

文献标志码: A

文章编号: 1672-9242(2024)04-0156-09

DOI: 10.7643/issn.1672-9242.2024.04.019

Metadata System of Natural Environment Experiments

ZHOU Junyan^{1,2}, WANG Jingcheng^{1,2}, ZHAO Fangchao^{1,2,3}, SHU Chang^{1,2}, HUANG Lun^{1,2}

(1. Southwest Institute of Technology and Engineering, Chongqing 400039, China; 2. Research Center for Natural Environment Test of National Defense Science and Technology Industry, Chongqing 400039, China; 3. CSGC Key Laboratory of Ammunition Storage Environmental Effects, Chongqing 400039, China)

ABSTRACT: The work aims to form a metadata system that provides comprehensive description and reasonable rule specifications for dataset resources and support for database design and development, data mining and analysis, data co-construction and sharing, etc. Aiming at the professional characteristics, data types, and application needs of the natural environment experiment, metadata standards and specifications from other fields were drawn to sort out and analyze the collection of data feature elements involved in natural environment experiments, classify the metadata of natural environment experiments, evaluate the role of metadata, and finally form a metadata system for natural environment experiments. This natural environment experiment metadata system covered eight categories, including dataset description information, element parameter information, data quality information, maintenance information, restriction information, dataset distribution information, metadata reference information and data application achievement information. The metadata of natural environment experiments has functions such as resource description, data selection, information retrieval, management traceability, interpretation and analysis, and storage, which provides a foundation for effective data management and co-construction and sharing. The metadata system proposed provides a

收稿日期: 2024-01-21; 修订日期: 2024-03-08

Received: 2024-01-21; Revised: 2024-03-08

引文格式: 周俊炎, 王竟成, 赵方超, 等. 自然环境试验元数据体系研究[J]. 装备环境工程, 2024, 21(4): 156-164.

ZHOU Junyan, WANG Jingcheng, ZHAO Fangchao, et al. Metadata System of Natural Environment Experiments[J]. Equipment Environmental Engineering, 2024, 21(4): 156-164.

reference for the metadata standards of natural environment experiments.

KEY WORDS: natural environment experiment; metadata; principle of metadata confirmation; metadata category; role of metadata; metadata system

自然环境试验是在自然环境条件下对材料、零部件、产品等进行的环境适应性试验与研究,评价贮存性能,特别关注典型或极端环境对材料、产品的影响^[1]。自然环境试验贯穿产品的论证、设计、生产、使用全过程,通过自然环境试验评估环境对产品功能/性能的影响,针对性采取保护措施,提升产品的环境适应性水平,减轻复杂环境对产品功能/性能的影响^[2-4]。材料和产品自然环境腐蚀或老化机理、情况复杂,采用机理仿真分析或实验室试验的方法只能在一定程度上贴近真实环境,不能完全复现实际环境带来的影响,这也是自然环境试验不可替代的原因^[5-6]。几十年来,相关单位开展了不同材料、不同类别产品的自然环境试验,积攒了大量自然环境试验数据,如钢材^[7]、吸波涂层^[8]、缓冲材料^[9]等。自然环境试验数据已成为基础战略资源,但自然环境试验数据资源分散存储在各个试验场站、研究中心,数据无法高效利用、资源无法快速整合,知识无法形成合力,面向种类繁杂的自然环境试验数据信息,亟需开展自然环境试验元数据体系研究,促进自然环境试验数据资源充分发挥科研服务作用、资源共享作用、科技创新作用。

元数据是具有抽象概念的结构化信息,是数据的数据,通过元数据对数据的描述、解释作用,实现对数据资源的快速定位检索,辅助用户更快、更好地实现信息利用、资源共享^[10]。通过元数据对数据的规约作用,规范自然环境试验数据资源的数据标识描述、数据结构、组织管理、质量控制、定位查询、共建共享、应用推广等数据全生命周期活动^[11],提高内容描述的准确性,促进对自然环境试验数据资源的发现、攫取、梳理、评价,搭建自然环境试验数据体系到数据服务的桥梁,通过元数据促进数据资源的流动和积累,加快数据服务的开发和应用。

随着数据库、大数据等技术的不断进步,元数据标准研究得到了同步发展。在国际上,2019年ISO发布的ISO/IEC 19583-1: 2019 Information technology—Concepts and usage of metadata Part 1: Metadata concepts,定义了元数据的概念、功能、分类^[12]。在其他某些领域,已形成较为完备的元数据标准,如GB/T 30524—2014《科技平台 元数据注册与管理》、T/PSC 7—2022《海洋科学数据共享服务规范 核心元数据》、GB/T 37282—2019《产品标签内容核心元数据》等。在自然环境试验领域,尚未形成明确的标准规范。同时,元数据也逐渐运用到各方各面。如魏笑等^[13]基于元数据开展了指标遴选方法筛选研究。Satyendra等^[14]提出了一种高效元数据模型设计原则。张晓艳^[15]基于元数据与分布式技术,实现了管理系统

的设计与应用落地。

本文以自然环境试验元数据为研究对象,针对自然环境试验专业特点、数据类型以及应用需求,梳理分析规范化描述自然环境试验涉及的观测数据、模型数据、影响评估数据、应用数据等数据特征的元素集合,开展自然环境试验基础数据元数据类别划分、需求分析与作用评估研究,最后形成自然环境试验元数据体系。

1 自然环境试验元数据确定原则

自然环境试验具有典型性、真实性、不可替代性等特点^[16],使得自然环境试验数据具有以下特点。

1) 高维度。不同材料、产品所处的自然环境类型复杂,不同环境具有不同特点,不同材料、产品具有不同的显著环境影响因子,具有较高的维度特性。同时,并非每次自然环境试验都对所有因素进行全量测量,存在数据缺省的情况。

2) 高噪声。自然环境试验数据积累过程时间跨度较大,基本以年月为单位,容易造成数据丢失或属性值不完整,存在属性间量纲不一致等情况。这种数据稀疏的状况,进一步增加了数据噪声。

3) 多变性。在自然环境试验数据采集过程中,易受到噪声干扰,同一地点不同年份气候特征存在较大差异,导致材料或产品的自然环境试验结果不稳定,造成自然环境试验数据存在多变性特征。

4) 突变性。由于测量仪器、人员操作等影响,特别是由于长时间跨度下仪器、测量人员的改变,导致自然环境试验数据容易出现奇异值,数据具有突变性。

由于自然环境试验数据具有高维度、高噪声、多变性、突变性等特点,要求自然环境试验元数据对数据进行简明但全面的描述。通过元数据的规范作用,消除一部分自然环境试验数据资源之间的异构性,并保留不同数据之间的差异性,使数据更加清晰、便于理解。

自然环境试验元数据是关于数据的数据,用来描述数据的简要介绍、具体内容、时空范围、数据质量、管理行为、数据共享分发等有关信息,以数据规范、数据共享为目的,充分考虑自然环境试验数据的分类体系、层次结构等,又要满足自然环境试验数据管理,以及发现、定位、获取数据的需要,自然环境试验元数据确定原则如下。

1) 目标明确性。数据集中汇聚、共享互通是元数据制定的核心目标,目标贯穿核心元数据归纳、确立全流程,明确的自然环境试验元数据目标,帮助数

据收集、管理、使用者了解数据的用途和价值,确保元数据实用,若目标不明确,可能导致元数据过度描述或描述不足、不到位,不能精确满足需求。明确的目标指导一致性、描述性、精确性等元数据特征制定。

2) 一致性。确保元数据的定义和命名一致,使用统一的标准和规范描述不同类型数据和资源,便于用户理解和使用元数据。一致的元数据能降低用户混淆和误解的可能性,同时有助于维护数据质量,使数据质量问题更容易识别和纠正。元数据一致性通过规范命名、数据类型和格式、度量单位、数据分类和层次等确定。

3) 描述性。元数据提供关于数据或资源的详细描述,使数据能被理解,帮助用户了解数据集的性质、构造、内容、特性、关系等信息,包括数据的标题、摘要、关键词、主题、作者、日期、地点、版本、语言、数据类型、层次结构、创建/修改/过期时间等。元数据为使用者提供尽可能详尽和全面的信息,以便数据被准确理解、评估和利用,提高数据资源的可发现性和可用性。

4) 精确性。确保元数据的信息是准确、可靠的,元数据反映与数据资源相关的真实情况,使用者能够依赖元数据了解数据的性质和价值,避免错误和不准确性,进而保障数据的可信度和可用性。如“环境试验时间”这一字段是记录时间、投样时间还是试验时长需要精确定义。精确性确保了元数据信息与实际数据资源相符,提供可信的数据描述和管理基础。

5) 可搜索性。元数据应包含足够、完整的信息,便于用户通过搜索、查询能够精确定位,支持多条件检索等,确保数据资源更容易被发现和访问。如需要,考虑增加过滤和筛选选项、建立索引等方式,以此增加可搜索性,允许使用者根据不同数据行、特征或条件来缩小搜索范围,索引根据元数据信息组织或存储数据资源创建,便于使用者快速检索。

6) 可维护性。元数据应易于维护和更新。随着技术手段的发展,数据和资源不断变化,元数据的设计遵循“低耦合”的原则,避免元数据元素或实体关系性过强,修改其一就必须修改耦合的元素或实体,确保元数据的可维护性。避免元数据的结构过于复杂或刚性,设计时充分考虑元数据添加、删除、修改和重新排列操作的可能性。同时,通过版本控制、记录历史文档等方式增强元数据的可维护性。

7) 安全性。构建元数据时应充分考虑保护敏感、涉密信息,维护数据隐私和安全,元数据内容需覆盖数据权限,如数据密级(公开、内部、秘密、机密、绝密)。根据用户涉密等级对应数据查看使用权限,如数据提供单位、数据使用单位等。根据权责划分、共享互通权限,实现同批数据不同群体不同权限。同时,数据的不同成分密级要求可能不同,如基础信息可公开,试验结果不可公开等,通过元数据的制定为数据定制化划分数据权限,为数据的安全提供基础。

以目标明确性、一致性、描述性、精确性、可搜索性、可维护性、安全性为确定原则,构建自然环境试验元数据具体内容,推动自然环境试验数据发现、识别、确认、记录、规范、共享^[17]。

2 自然环境试验元数据类别与作用

2.1 自然环境试验元数据类别

ISO 元数据标准中将元数据分为结构元数据、描述性元数据、管理元数据 3 类。参考这 3 种类型,结合自然环境试验专业特点、数据类型以及应用需求,自然环境试验元数据覆盖信息发现、数据类型管理、储存架构、元数据更新等方面^[18]。将自然环境试验元数据划分为数据集描述信息、要素参数信息、数据质量信息、维护信息、限制信息、数据集分发信息、元数据参考信息、数据应用成果信息共 8 种类别。

1) 数据集描述信息。数据集描述信息涵盖数据集的概要介绍,是对数据集的整体描述,详细阐述了数据集的基本信息,整合了数据集的属性内容,包括数据集名称、标识、主题、类型、数据量、数据来源、更新频率、字符集、范围信息等,使用者基于数据集描述信息对数据集整体情况形成宏观理解。

2) 要素参数信息。要素是自然环境试验与观测中记录的各个物理量或各类相关信息,要素的监测、采集、测定、记录是自然环境试验元数据中最关键核心的信息。要素参数信息描述了自然环境试验数据集中要素的参数特性,用户可以通过要素参数信息了解数据文件中数据资料的内容、排列方式、数据的采集方式和处理方式等。要素参数信息依据自然环境试验专业制定,是最重要的核心数据,具体包含:监测地点、试验站信息(编号、名称、所属、位置、环境类型、场地情况等)、实验室信息(名称、所属、地址、检测类别、联系方式等)、试验场景、观测频率、产品信息(类别、品类、牌号、处理工艺、生产单位、生产日期)、投试日期、投试时长、要素(编码、名称、单位、测量时间、测量方法、统计特征)、归档日期、归档人、审核人等。

3) 数据质量信息。该部分是对数据质量的总体评价,包括数据质量状况、数据志和完整性等方面的信息。使用者基于数据集质量信息对数据质量形成定论,进而判断数据质量是否满足业务要求和用户需求。

4) 数据集分发信息。数据集分发信息描述了数据集流转、共享、获取相关内容,包括数据集的格式、使用数据方式、收费策略,权限申明等,用户基于数据集分发信息可以了解数据集对外分发的限制、形式、内容等信息。

5) 元数据参考信息。元数据参考信息描述了元数据相关的标准名称、创建时间、修改时间等内容,通过元数据参考信息,使用者可以了解数据规范化相

关信息。

6) 维护信息。维护信息是数据资源更新频率与更新范围的信息。

7) 限制信息。由于不同自然环境试验数据具有不同密级、知悉范围,限制信息用于确保数据资源和元数据的安全性,防止信息泄露和滥用,维护国家秘密、商业机密等重要利益。

8) 数据应用成果信息。数据应用成果信息用于描述数据资源典型应用领域和成果。

2.2 自然环境试验元数据需求分析

针对自然环境试验高维度、高噪声、多变性、突变性的特点,面向数据无法高效利用、资源无法快速整合、知识无法形成合力的核心矛盾,自然环境试验元数据的需求主要表现在数据定义、采集、存储、管理和应用等方面。

1) 数据定义与一致性需求。自然环境试验数据资源涉及材料、零部件、产品等多个层级,参数类型多、覆盖范围广,不同群体对专业术语具有不同理解,导致数据含义不清晰,容易发生误解,需要通过元数据规范数据定义,约束术语和数据单位,保障自然环境试验数据的一致性。

2) 数据采集与完整性需求。自然环境试验往往基于独立的项目开展,不同项目的侧重点不同,采集的数据类型、体量不同。通过元数据制定数据采集规划,可减少疏漏项,增强数据完整性。同时,元数据要求明确数据采集、获取来源,增加数据的可信度。

3) 数据存储与持续性需求。数据依托于信息系统、数据库存储,但各个数据系统之间异构性较强,数据资源无法快速交换,新数据无法高效入库,通过元数据可实现数据模型的设计和数据库的搭建,进一步实现存量数据的规范与梳理、保障增量数据的持续性入库。

4) 数据管理与合规性需求。质量控制、溯源分析是数据管理的关重环节,但由于数据自管自用,缺乏数据全流程管理,自然环境试验数据资源合规性难以保障,需要通过元数据的规范作用,对数据志、数据质量评价、数据处理信息进行要求,提升数据的合规性。

5) 数据应用与共建共享需求。由于数据资源分散存储在各个试验场站、研究中心,数据不能在数据产生单位和数据应用单位之间进行有效流通,出现数据建设冗余、数据应用层次不足的割裂局面。数字化的推进也对数据共建共享提出更高层次的要求,通过元数据的规约作用,促进数据共建共享,提升数据应用建设水平。

针对自然环境试验元数据需求,并借鉴成熟领域元数据的功能、作用,总结提炼自然环境试验元数据的作用。王凯等^[19]认为,铁路元数据有描述、检索、选择、定位、管理和评估共6个方面的作用。王志梅

等^[20]将装备合同监管信息元数据作用凝练为信息采集、管理、检索、保存、统计分析等功能。王秀琴等^[21]认为气象档案元数据支撑数据源管理、数据存储、历史数据记录与溯源。自然环境试验元数据的作用归纳如下。

1) 资源描述。描述自然环境试验数据资源内容、属性是自然环境试验元数据的基本功能,要求通过元数据的描述能完整反映资源对象的概貌。

2) 数据选择。通过对元数据的阅读,使用者不用详细浏览、理解数据资源细节,就能对数据资源对象形成基本了解和认识,进而对数据资源进行选择、取舍。

3) 信息检索。元数据的内容包含来源、类型、数据量、试验站信息、样品信息、测量要素等信息内容,帮助使用者进行资源的检索、确认。

4) 管理溯源。自然环境试验元数据规范了数据来源、数据集提供者等数据资源加工、存档、使用相关信息,还包含知悉范围、权限等限制信息,支持对数据资源进行管理与溯源。

5) 解释分析。自然环境试验元数据提供了试验条件、试验地点、检测结果的详细信息,助于科研人员正确解释分析数据,了解产品或材料自然环境试验的全流程。

6) 储存入库。元数据对自然环境试验数据进行全面、具体的描述,通过对元数据的阅读理解,有助于规划数据逻辑模型,以及进行数据的结构化表单设计和储存入库。

自然环境试验元数据是高效管理和共建共用自然环境试验数据的基础,也是该类数据资源充分发挥科研服务作用、资源共享作用、科技创新作用的重要前提之一,其作用涵盖资源描述文档的创建、阅览、分发、共享等多个方面^[22],统一规范了多类型、多来源、多要素数字资源,助于科学描述与有序管理^[23]。

3 自然环境试验元数据体系

通过对自然环境试验过程的梳理和数据资源的归纳,参照元数据确定原则,确立自然环境试验元数据体系。元数据体系由元数据集合、实体、元素组成,自然环境试验元数据体系主要由数据集描述信息、要素参数信息、数据质量信息、维护信息、限制信息、数据集分发信息、元数据参考信息、数据应用成果信息8个集合组成,实现对自然环境试验数据的定义、采集、存储、管理和应用内容的全覆盖,每个集合有其元数据实体或元数据元素,元数据层次结构如图1所示。其中,元数据元素是元数据的基本单元,是不可拆分的最小类别;元数据实体是同类元数据元素的整理集合,元数据实体可只包含元数据元素,也可以同时包含1个或多个实体,形成聚合实体;元数据集合是元数据实体和元素的整合汇聚。在后文中,元数

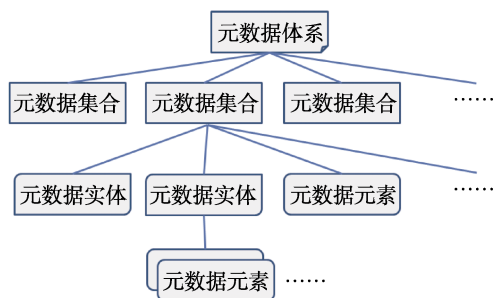


图 1 元数据层次结构
Fig.1 Metadata hierarchy diagram

据元素的类型细分为 String (字符或文本型)、Num (数值型)、Date(日期或时间型)、Bool 型(布尔型)。

1) 数据集描述信息。关于数据集的基本信息, 包含关键字、数据集提供者等实体, 包含数据集名称、标识等元素, 见表 1。

2) 要素参数信息。是数据集包含要素的参数信息, 包含试验站信息、实验室信息、测量要素等实体, 包含监测地点、观测频率等元素, 见表 2。

3) 数据质量信息。数据质量信息是总体评价数据集质量的信息, 包含数据质量评价、数据处理信息

表 1 数据集描述信息
Tab.1 Dataset description information

编号	名称	定义	类型
1.1	数据集名称	数据集的名称	String
1.2	数据集标识	根据某种规则赋予数据集的代码	String
1.3	关键字	描述数据集内容的关键字	实体
1.3.1	关键字内容	概括数据集主要内容的通用词、形式化词或短语	String
1.3.2	词典名称	关键字所属的专业关键词词典名称	String
1.4	描述	对数据集内容的文本描述	String
1.5	类型	对数据集所属类型的说明	String
1.6	数据量	对数据集的数据量说明	实体
1.6.1	记录数	数据量的条数	Num
1.6.2	存储空间占用	存储的物理空间占用量	String
1.7	数据来源	当前数据资源的源头信息	String
1.8	数据集提供者	提供数据资源的个人或单位的基本信息	实体
1.8.1	提供方名称	提供数据资源的个人或单位名称	String
1.8.2	提供方地址	提供数据资源的个人或单位地址	String
1.8.3	提供方电子邮件地址	提供数据资源的个人或单位电子邮件	String
1.8.4	提供方联系人姓名	提供数据资源的个人或单位联系人姓名	String
1.8.5	提供方电话号码	提供数据资源的个人或单位电话	Num
1.8.6	数据集贡献者	对数据集创建做出贡献的个人或组织的名称	String
1.9	数据类目	对数据资源所属类目的描述	实体
1.9.1	类目名称	相关分类与编码的类目名称	String
1.9.2	类目编码	类目名称对应的标识码	String
1.10	更新频率	描述数据集在多大的时间内更新一次	String
1.11	数据集时间	描述数据集的增改事件相关的时间	实体
1.11.1	创建时间	数据集的创建时间	Date
1.11.2	最近修改时间	数据集最新的更新时间	Date
1.11.3	数据集发布时间	数据集发布的时间	Date
1.12	语种	数据集内容对应的语言种类	String
1.13	字符集	计算机中用来编码字符的一组字符集合, 定了每个字符对应的数字编码, 支持计算机理解、存储、展示文本内容, 如 ASCII 等	String
1.14	URL	数据集相关网络服务的链接地址	String
1.15	关联数据集	与当前数据集相关的其他数据集	实体
1.15.1	关联数据集名称	对关联数据集名称的描述	String
1.15.2	关联 URL	关联数据集网络服务链接地址	String
1.16	数据集范围	数据集范围包含分类、时间和空间等内容	实体
1.16.1	学科范围	数据集内容所涉及的学科范围	String
1.16.2	时间范围	数据集内容所涉及的时间范围	String
1.16.3	空间范围	数据集内容所涉及的空间范围	String
1.17	数据格式	数据集内容所涉及的主要格式	String

表 2 要素参数信息
Tab.2 Element parameter information

编号	名称	定义	类型
2.1	监测地点	监测要素所处的地点或产品的试验地点	String
2.2	试验站信息	进行环境因素监测或产品环境试验的试验站的相关信息	实体
2.2.1	试验站编号	与试验站名称匹配的唯一编号	String
2.2.2	试验站名称	试验站的名称	String
2.2.3	所属单位	试验站所属单位	String
2.2.4	地理位置	试验站经纬度、省市划分等地理位置信息	String
2.2.5	建站时间	试验站建站时间	String
2.2.6	环境类型	试验站所处气候环境类型	String
2.2.7	场地情况	试验站场地数量、占地面积、支持试验类型等情况	String
2.2.8	联系方式	试验站固定电话或相关人员电话、电子邮件等信息	String
2.3	实验室信息	进行产品性能检测的实验室相关信息	实体
2.3.1	实验室名称	实验室名称	String
2.3.2	所属单位	实验室所属单位信息	String
2.3.3	通讯地址	实验室通讯地址	String
2.3.4	检测类别	实验室支持的检测项目和类别	String
2.3.5	联系方式	实验室固定电话或相关人员电话、电子邮件等信息	String
2.4	试验场景	观测与试验所处的场景形式	String
2.5	观测频率	要素在单位时间内进行观测的次数	String
2.6	样品信息	进行自然环境试验样品的相关信息	实体
2.6.1	样品类别	样品的材料、型号类别	String
2.6.2	样品品类	样品类别下细分的种类	String
2.6.3	样品牌号	样品品类下细分的牌号	String
2.6.4	处理工艺	样品制备时采用的处理工艺	String
2.6.5	生产单位	样品制备、加工的生产单位	String
2.6.6	生产日期	样品制备、加工的生产时间	Date
2.6.7	投试日期	样品进行试验的投试时间	Date
2.6.8	试验时长	试验进行试验的试验时长	String
2.7	测量要素	数据集中各个要素特性的详细描述	实体
2.7.1	要素编码	作为测量要素的唯一标识	String
2.7.2	要素名称	测量要素的名称描述	String
2.7.3	计量单位	测量要素的计量单位	String
2.7.4	测量时间	测量要素的时间	Date
2.7.5	测量方法	测量要素时使用的仪器、方法	String
2.7.6	统计特征	测量要求统计的特征	String
2.8	归档日期	数据集要素录入档案的日期	Date
2.9	归档人	数据集要素录入档案的人员	String
2.10	审核人	数据集要素合规审核的人员	String

等实体, 包含数据质量情况、数据志等元素, 见表 3。

4) 维护信息。是数据集资源更新频率和更新范围信息, 包含维护和更新频率、维护和更新责任单位等元素, 见表 4。

5) 限制信息。是描述密级、知悉范围相关的信息, 包含数据集密级、元数据密级等实体, 包含涉密等级、知悉范围等元素, 见表 5。

6) 数据集分发信息。描述数据集共建分享、获

取方法的信息, 包含数据格式、联系方式等实体, 包含介质名称、收费策略等元素, 见表 6。

7) 元数据参考信息。是元数据的相关标准规范信息, 包含元数据创建时间、最近修改时间、标准名称等元素, 见表 7。

8) 数据应用成果信息。是数据集资源的应用领域和典型成果, 包含数据应用领域、典型成果等元素, 见表 8。

表 3 数据质量信息
Tab.3 Data quality information

编号	名称	定义	类型
3.1	数据质量情况	数据集质量检验的情况	String
3.2	数据志	对数据采集、处理及结果记录等全链条工作的描述和记录,包括数据来源、采集方式、整理方法、处理过程等信息	String
3.3	数据质量评价信息	关于数据集中各个要素完整性、误差、评价方法等质量评价信息	实体
3.3.1	数据完整性	数据要素是否完整	Bool
3.3.2	绝对误差	测量数据与真实值的绝对误差	String
3.3.3	相对误差	测量数据与真实值的相对误差	String
3.3.4	评价方法说明	关于数据质量评价的方法描述	String
3.3.5	评价程序	数据质量评价程序步骤	String
3.3.6	评价日期	进行质量评价的时间	Date
3.3.7	评价结果	关于质量评价的结果描述	String
3.4	数据处理信息	进行数据质量控制的处理信息描述	实体
3.4.1	处理时间	进行数据处理的时间	Date
3.4.2	处理说明	数据处理的方案、流程描述	String
3.4.3	处理者	执行数据处理的人员	String
3.4.4	处理软件	实现数据处理的软件工具或源码程序等	String

表 4 维护信息
Tab.4 Maintenance information

编号	名称	定义	类型
8.1	维护和更新频率	数据维护和更新的时间间隔	String
8.2	更新范围说明	更新的数据内容、时间等范围	String
8.3	维护和更新责任单位	负责维护和更新的单位信息	String

表 5 限制信息
Tab.5 Restriction information

编号	名称	定义	类型
5.1	数据集密级	描述数据集密级、知悉范围相关的信息	String
5.1.1	涉密等级	描述数据集的涉密等级	String
5.1.2	知悉范围	描述数据集的知悉范围	String
5.2	元数据密级	描述元数据密级、知悉范围相关的信息	实体
5.2.1	涉密等级	描述元数据的涉密等级	String
5.2.2	知悉范围	描述元数据的知悉范围	String

表 6 数据集分发信息
Tab.6 Dataset distribution information

编号	名称	定义	类型
4.1	数据格式	数据格式的说明介绍	实体
4.1.1	名称	数据集、数据汇编的名称	String
4.1.2	版本	数据集的版本说明,包括增减信息等	String
4.1.2	格式说明	数据的格式说明,通常一个数据集不止包含单一类型	String
4.2	介质名称	存储、交换数据集的介质,如光盘、纸质、网络 API 等	String
4.3	技术要求	数据集的技术要求、技术说明信息	String
4.4	收费策略	付费数据的时间范围、知悉范围、定价准则等	String
4.5	权限声明	数据的权责说明	String
4.6	访问时间	数据集分发的时间,若对于网络 API 形式则是请求时间	Date
4.7	联系方式	数据集提供方的联系方式	实体
4.7.1	联系人	数据集提供方联系人	String

续表 6 数据集分发信息
Tab.6 Dataset distribution information (continued)

编号	名称	定义	类型
4.7.2	单位	数据集提供方单位	String
4.7.3	联系地址	数据集提供方联系地址	String
4.7.4	电话	数据集提供方电话	Num
4.7.5	电子邮件	数据集提供方电子邮件	String
4.7.6	传真	数据集提供方传真	String
4.7.7	联系时间	数据集提供方支持的联系时间, 通常为工作日	Date

表 7 元数据参考信息
Tab.7 Metadata reference

编号	名称	定义	类型
6.1	元数据创建时间	元数据标准规范的创建时间	Date
6.2	元数据最近修改时间	元数据最新的修改日期	Date
6.3	元数据标准名称	元数据标准的名称, 以此作为检索项	String

表 8 数据应用成果信息
Tab.8 Information on data application achievements

编号	名称	定义	类型
7.1	数据应用领域	数据资源可应用到的领域范围	String
7.2	典型成果	数据资源形成的典型成果	String

4 结语

1) 本文描述了自然环境试验数据的特点及元数据的确认原则, 阐述了元数据类别与作用, 形成了自然环境试验元数据体系, 为自然环境试验元数据标准提供了参考思路。

2) 自然环境试验元数据具有资源描述、数据选择、信息检索、管理溯源、解释分析、储存入库等作用, 为数据有效管理和共建共享提供了基础。

3) 自然环境试验元数据体系覆盖数据集描述信息、要素参数信息、数据质量信息、维护信息、限制信息、数据集分发信息、元数据参考信息、数据应用成果信息共 8 个类别, 其中要素参数信息是最核心、价值密度最高的内容。

4) 在自然环境试验专业的不断发展和数字转型的时代背景下, 自然环境试验元数据也将与时俱进、迭代优化, 如多套系统之间的元数据互操作问题^[24]、元数据质量问题^[25]、元数据管理模式问题^[26]等内容值得深入研究。

参考文献:

[1] 赵朋飞, 苏晓庆, 张生鹏. 装备自然环境试验工作管理方法探讨[J]. 装备环境工程, 2020, 17(7): 116-121.

ZHAO P F, SU X Q, ZHANG S P. Management Methods of Materiel Natural-Environment Tests[J]. Equipment Environmental Engineering, 2020, 17(7): 116-121.

[2] 苏艳, 舒德学, 杨晓然. 装备全寿命期自然环境试验工作的管理[J]. 装备环境工程, 2004, 1(3): 13-17.

SU Y, SHU D X, YANG X R. The Management of Natural Environmental Test in Total Life Cycle of Materiel[J]. Equipment Environmental Engineering, 2004, 1(3): 13-17.

[3] GJB 4239—2001, 装备环境工程通用要求[S].
GJB 4239—2001, General Requirements for Equipment Environmental Engineering[S].

[4] GJB 8893—2017, 军用装备自然环境试验方法[S].
GJB 8893—2017, Natural Environment Test Methods for Military Materiel[S].

[5] 秦晓洲, 李颖. 军用装备自然环境试验方法剪裁探讨[J]. 装备环境工程, 2011, 8(3): 54-57.

QIN X Z, LI Y. On Tailoring of Natural Environmental Test Methods of Materiel[J]. Equipment Environmental Engineering, 2011, 8(3): 54-57.

[6] 朱玉琴, 陈源, 张燕, 等. 国内外动态自然环境试验技术发展现状[J]. 装备环境工程, 2015, 12(6): 93-99.

ZHU Y Q, CHEN Y, ZHANG Y, et al. Current Status of Dynamic Natural Environmental Test Techniques at Home and Abroad[J]. Equipment Environmental Engineering, 2015, 12(6): 93-99.

[7] 覃粒, 吴德权, 胡涛, 等. 海南湿热海洋大气环境 Q235 钢腐蚀行为研究及严酷度评估[J]. 装备环境工程, 2023, 20(7): 90-97.

QIN L, WU D Q, HU T, et al. Corrosion Behaviors of Q235 Steel and Severity Evaluation for Humid and Hot Marine Atmosphere Environmental in Hainan[J]. Equipment Environmental Engineering, 2023, 20(7): 90-97.

[8] 宿兴涛, 孙敬哲, 朱晓蕾. 雷达吸波涂层海洋大气环境适用性试验研究[J]. 装备环境工程, 2023, 20(2): 17-25.

SU X T, SUN J Z, ZHU X L. Experimental Study on Applicability of Radar Absorbing Coating to Marine Atmospheric Environment[J]. Equipment Environmental Engineering, 2023, 20(2): 17-25.

[9] 胡涛, 李茜, 孙茂钧, 等. 聚氨酯缓冲材料在 2 种典型环境下的耐候性研究[J]. 装备环境工程, 2023, 20(1): 119-126.

HU T, LI Q, SUN M J, et al. Weather Resistance of Polyurethane Cushioning Materials in Two Typical Environ-

- ments[J]. *Equipment Environmental Engineering*, 2023, 20(1): 119-126.
- [10] 张朔. 基于元数据的开放政府数据整合研究[D]. 太原: 山西大学, 2021.
ZHANG S. Research on Integration of Open Government Data Based on Metadata[D]. Taiyuan: Shanxi University, 2021.
- [11] 兀伟, 王焕萍, 张浩, 等. 新型基础测绘与实景三维中国建设地理场景元数据标准化探讨[J]. *测绘标准化*, 2023, 39(3): 1-9.
WU W, WANG H P, ZHANG H, et al. On the Standardization of Geo- Scene Metadata of New Fundamental Surveying and 3D Real Scene China Construction[J]. *Standardization of Surveying and Mapping*, 2023, 39(3): 1-9.
- [12] 严红, 穆志勇, 李明哲, 等. 元数据标准化发展研究[J]. *信息技术与标准化*, 2023(9): 25-29.
YAN H, MU Z Y, LI M Z, et al. Research on the Development of Metadata Standardization[J]. *Information Technology & Standardization*, 2023(9): 25-29.
- [13] 魏笑. 基于元数据的指标遴选方法研究[D]. 北京: 军事科学院, 2021.
WEI X. Index Selection Method Research Based on Metadata[D]. Beijing: Military Science, 2021.
- [14] SATYENDRA N B, SWAMI N K, BHAILUME P V. Design Guidelines for Building User Centric Metadata Models for Data Warehouse Systems[C]// 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). Vellore: IEEE, 2020.
- [15] 张晓艳. 分布式元数据管理系统的设计与实现[D]. 成都: 电子科技大学, 2020.
ZHANG X Y. Design and Implementation of Distributed Metadata Management System[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
- [16] 宣卫芳, 朱蕾, 张伦武. 自然环境试验在型号研制生产中的应用[J]. *装备环境工程*, 2009, 6(3): 51-55.
XUAN W F, ZHU L, ZHANG L W. Applications of Natural Environmental Test on Materiel Development[J]. *Equipment Environmental Engineering*, 2009, 6(3): 51-55.
- [17] 董仁才, 王韬, 张永霖, 等. 我国城市可持续发展能力评估指标的元数据分析与管理[J]. *生态学报*, 2018, 38(11): 3775-3783.
DONG R C, WANG T, ZHANG Y L, et al. Metadata Analysis and Management for an Urban Sustainable Development Capability Evaluation Index in China[J]. *Acta Ecologica Sinica*, 2018, 38(11): 3775-3783.
- [18] 赵凯楠. 基于元数据的军事领域多源数据集成应用研究[D]. 北京: 中国电子科技集团公司电子科学研究院, 2022.
ZHAO K N. Research on Multi-Source Data Integration Application in Military Field Based on Metadata[D]. Beijing: China Electronics Technology Group Corporation Electronic Science Research Institute, 2022.
- [19] 王凯, 王健慧, 杜军, 等. 重载铁路元数据研究综述[J]. *电子世界*, 2021(2): 79-80.
WANG K, WANG J H, DU J, et al. Summary of Metadata Research on Heavy-Haul Railway[J]. *Electronics World*, 2021(2): 79-80.
- [20] 王志梅, 郭建明, 杨铭. 装备合同监管信息元数据研究[J]. *信息技术与标准化*, 2021(10): 62-65.
WANG Z M, GUO J M, YANG M. Research on Metadata of Equipment Contract Supervision Information[J]. *Information Technology & Standardization*, 2021(10): 62-65.
- [21] 王秀琴, 梁中军. 气象档案数据元数据设计与应用[J]. *气象水文海洋仪器*, 2022, 39(4): 132-134.
WANG X Q, LIANG Z J. Design and Application of Metadata in Meteorological Archival Data[J]. *Meteorological, Hydrological and Marine Instruments*, 2022, 39(4): 132-134.
- [22] 王志强, 杨青海. 科技资源元数据标准化研究[J]. *标准科学*, 2021(5): 29-33.
WANG Z Q, YANG Q H. Research on Standardization of S & T Resource Metadata[J]. *Standard Science*, 2021(5): 29-33.
- [23] 田红. 面向多源异构航天电子档案的元数据方案设计[J]. *北京档案*, 2023(10): 38-40.
TIAN H. Metadata Scheme Design for Multi-Source Heterogeneous Aerospace Electronic Archives[J]. *Beijing Archives*, 2023(10): 38-40.
- [24] 孔祥盛. 我国档案元数据互操作的现状、问题和对策研究[J]. *档案管理*, 2023(1): 45-49.
KONG X S. Research on the Status, Problems and Countermeasures of Archives Metadata Interoperability in China[J]. *Archives Management*, 2023(1): 45-49.
- [25] NOGUERAS-ISO J, LACASTA J, UREÑA-CÁMARA M A, et al. Quality of Metadata in Open Data Portals[J]. *IEEE Access*, 2021, 9: 60364-60382.
- [26] WU H, NATHELLA K, SUNWOO D, et al. Efficient Metadata Management for Irregular Data Prefetching[C]// Proceedings of the 46th International Symposium on Computer Architecture. Phoenix Arizona: ACM, 2019.