

《土壤科学数据元数据》国家标准编制说明

2012-11-26

《土壤科学数据元数据》国家标准编写组

1. 工作简况

1.1 项目启动

土壤科学数据广泛应用于农业、林业、生态、环保等领域，是科技创新和经济发展重要的战略资源。由于缺乏关于数据内容、质量和数据产生方法等数据特征描述的统一规范，数据难以被发现、获取并缺乏可用性，未能发挥在数据共享和跨学科应用中应有的作用。相关元数据的缺乏和不规范成为是制约土壤科学数据发现、管理和使用的主要因素之一。

中国科学院南京土壤所通过土壤质量标准化委员会（SAC/TC 404）申请《土壤科学数据元数据》国家标准立项，2011年12月获国家标准委立项批准(国家标准制定计划项目20111821-T-326)。经土壤质量标准化技术委员会批准,2012年2月《土壤科学数据元数据》编写工作组成立。编写组成员覆盖土壤科学、信息科学方面的专家，并具有土壤养分循环联网试验研究、863农田面源污染研究、耕地质量调查、土壤科学数据和土壤空间数据库建设、碳专项研究等科研项目背景。标准编制单位为：中国科学院南京土壤所，中国科学院计算机网络信息中心，中国科学院东北地理与农业生态研究所，中国科学院成都山地灾害与环境研究所和江苏省标准化研究院。

编写组主要成员参加过《生态科学数据元数据》(GB/T 20533-2006)的制定，具有元数据标准编写的经验。2012年3月21日在三亚亚太国际会议中心召开了《土壤科学数据元数据》项目启动会议，确定了项目具体工作内容、研究范围，工作计划、时间安排和任务分工。

1.2 起草阶段主要工作内容

编写组成立后，对国内外地理信息、生态、土壤、环境、气象、地质、国土资源领域相关元数据标准进行了收集和文献研究，并对国内外基于Web的主要土壤科学数据资源和元数据的应用情况进行了调研。

在前期调研和工作基础上，于2012年3月形成了《土壤科学元数据(初稿)》，并于2012年3月21日召开会议，征求了中科院南京土壤所、中科院兰州寒旱所、中科院计算机网络中心、中国科学院地理科学与资源研究所、中科院西北水保所、中国科学院南京地理与湖泊研究所、黑龙江八一农垦大学、黑龙江大学等8位专家意见，增添了与土壤学数据应用相关的数据获取场地和方法描述元信息，对于数据质量的元数据信息进行了修改，形成《土壤科学数据元数据(修改稿)》。2012年4月23日再次召开会议，邀请农业部农业技术推广中心、浙江大学、中科院亚热带农业生态所、中国农科院农业资源与农业区划所、华中农业大学、中科院沈阳应用生态所、江苏省标准化研究院等9位专家意见对修改稿征集意见。根据专家意见和土壤学科的特点，确定了《土壤科学数据元数据》标准适用范围、工作原则，修订了与土壤科学数据查询相关的学科分类、土壤分类和数据字典等内容。之后，选取一些单位依据《土壤科学数据元数据(修改稿)》试验填写各种应用类型数据集元数据应用实例。2012年7月11日，编写组在内蒙奈曼召开《土壤科学数据元数据》标准研讨会，会议邀请野外站与土壤相关的科学家参加，重点讨论了土壤分类、元数据填写中的问题及规范推广使用的问题，修订完善了本标准，于2012年11月形成了《土壤科学数据元数据(征求意见稿)》。2012年12月，编写组对文稿中不符合GB/T 1.1—2009的文字再次进行了修改。

2 标准编制原则和技术内容的确定方法

2.1 编制原则

- 1) 目的明确。标准编写的目的主要用于数据共享和数据整合,快速发现数据。
- 2) 技术先进。面向土壤科学数据,满足科学性、国际交流性和可操作性,为未来数据技术发展提供框架。
- 3) 需求主导。充分考虑土壤科学数据管理的需求以及数据资源的特点,所制定的元数据标准应当满足土壤科学数据管理、共享、交换的基本需求。
- 4) 标准兼容。按照GB/T1.1-2009《标准化工作导则 第1部分:标准的结构和编写》的要求和规定编写本标准内容。在制定标准过程中,需要从长远考虑,尽量和已有的国家以及相关标准兼容。
- 5) 可证实原则。目前标准使用主要面向用户而非计算机系统,应便于元数据标准的推广应用,必选的简单且必须能执行。资料性附录中应附有不同类型的土壤科学数据元数据典型实例。
- 6) 关键的内容又找不到权威性定义的请资深专家商议处理。

2.2 技术内容

2.2.1 国内外相关元数据标准研究进展

国际上不断发展对元数据的研究工作。从20世纪90年代初,国际标准化机构就开始了地理信息元数据标准的研究与制定工作。美国联邦地理数据委员会(Federal Geographic Data Committee, FGDC)标准化工作组起草了地理空间数据的元数据内容标准(CSDGM)草案。该标准草案元数据内容由标识信息、数据质量信息、空间数据组织信息、空间参照系信息、实体和属性信息、发行信息以及元数据参考信息等7部分组成,1994年8月FGDC通过并发布第1版CSDGM。此后,美国联邦政府内外的许多单位根据12906号总统令的要求,从1995年开始执行这一标准。生物学研究存在大量非结构化数据和物种分类信息,1999年FGDC发布了适于生物学领域的专用元数据标准(Biological data profile),增添了分类信息、ASCII文件和地质年代。1996年美国生态学会未来长期生态学数据委员会(FLED)制定了其元数据标准,并于2000年公布了用于生态学数据交换的数据置标语言(EML)。EML强调野外观测场地和研究方法的记录,美国长期生态学研究网络和国际长期生态学研究网络,都推荐参照EML2.0及其元数据管理软件,实现元数据的管理。2003年国际标准化组织地理信息委员会(ISO TC211)发布了地理信息元数据标准(ISO 19115, 2003. Geographic information/ Geomatics)。ISO/TC 211同时还制定了一系列标准,包含了信息技术、GIS技术、RS技术、GPS技术等领域最前沿的概念、模型、模式和技术方法的研究和实践,各项标准/计划既各自独立又相互关联,具有明确的结构性,形成了对地理信息及其相关应用的系统支撑和规范。在土壤信息领域,国际上相关组织开发了适合本领域应用的元数据规范。例如,美国农业部USDA提供了基于县一级土壤调查地理数据库SSURGO的元数据;密西根州立大学研究制定了针对土壤微生物研究所需科学数据的元数据;作为欧共体空间信息框架(INSPIRE)的评估和开发策略实践,欧洲土壤地理数据服务(GS soil)开发了与ISO 19115兼容的土壤元数据。

参照ISO19115,国内全国地理信息标准化技术委员会(SAC/TC210)于2005年发布地理信息元数据(GB/T 19710)。“十一五”以来,地理信息标准化技术委员会(SAC/TC210)修制定标准69项,采用国际标准13项。一大批基础性关键性地理信息国家标准的制定实施,正在减少与国际先进水平的差距。

为使国土资源信息建设有章可循,国土资源部制定了《国土资源信息核心元数据标准(TD/T1016-2003)》。该标准的描述对象是国土资源空间与非空间信息的数据集或数据集系列,参考国际标准化组织元数据标准(ISO19115),结合国土资源信息的特点及描述要求对其内容进行一定的增删形成的。国家地质调查局参考GB/T 19710/ISO 19115 地理信息元数据和TD/T 1016-2003 国土资源信息核心元数据标准,在原有《地质调查元数据内容与结构标准》基础上,发布了《地质信息元数据标准(DD2006-05)》,该标准包含了国家标准《地理信息元数据》核心

元数据的全部内容，并包含了《国土资源信息核心元数据标准》的所有元数据元素，根据实际需要进行了扩展。为加强气象资料共享，进一步促进气象资料更好地为经济建设、国防建设、社会发展和人民生活服务，国家气象局于2005年发布了气象数据集核心元数据(QX/T39-2005)。该标准借鉴WTO和ISO 19115相关的元数据标准，根据我国气象数据集制作、存储和服务的特点而制定，规定了完整描述一个气象数据集时所需要的数据项集合、各数据项语义定义和著录规则等，它提供了有关气象数据集的标识、内容、分发、数据质量、参照和限制等信息。

为促进中国的科学数据共享，科技部非常重视科学数据共享的元数据标准化工作。根据科学数据共享标准化工作的规划，科学数据共享工程开展了元数据相关标准(包括元数据标准化基本原则和方法、元数据内容、元数据检索和提取协议)的研究，参考ISO 19115, 2005年发布了《科学数据共享元数据内容标准》。2009年11月，成立了全国科技平台标准化技术委员会(SAC/TC486)，进一步加强科技资源平台标准化工作。环境保护部为规范环境信息资源元数据内容，指导环境信息资源的规范化描述，制定了环境保护信息元数据规范(2010, 征求意见稿)，规定了环境信息元数据框架、定义了核心元数据的内容，用以描述环境信息数据集的标识、内容、管理与维护信息及著录规则。其核心元数据内容与GB/T 19710-2005和《科学数据共享元数据内容标准》兼容。

基于“科学数据库”和“中国生态系统研究网络”的应用需要，中国科学院计算机网络中心牵头，中科院地理科学与资源研究所、南京土壤所、水土保持所、寒区旱区环境与工程所、东北地理与农业生态所和成都山地灾害与环境所等研究所共同参与，制定了生态科学数据元数据标准(GB/T 20533—2006)。该标准结合生态学研究的特点，其最小元数据元素集与GB/T 19710/ISO 19115兼容。近年来，在中国科学院十一五信息化建设《科学数据应用环境》项目的资助下，中国科学院南京土壤所联合中国科学院计算机网络信息中心，中国科学院东北地理与农业生态所和成都山地灾害与环境所，完成了《土壤科学数据元数据(初稿)》，这些工作的开展为土壤科学数据元数据的规范化和标准化奠定了基础。

2.2.2 国内外土壤科学数据资源和元数据应用现状

国际上影响较大的土壤科学数据资源包括：由联合国粮农组织(FAO)、国际应用系统分析研究所(IIASA)、ISRIC世界土壤信息中心、中国科学院南京土壤所和欧共体联合研究中心(JRC)为主完成的和谐世界土壤数据库(Harmonized World Soil Database)和ISRIC世界土壤信息中心和FAO共同支持的全球土壤与地体数据库(Global Soil and Terrain Database)，其中的中国部分均由中国科学院南京土壤研究所完成。HWSD中国部分基于1:100万数字化土壤图而建立，SOTER中国部分通过SRTM1遥感雷达影像生成数字高程模型提取地体数据，将中国1:100万的数字化土壤图、1:50万的数字化岩性图和地体单元图进行空间的套合叠加，从而构成中国SOTER的空间数据库。这些数据资源都可通过互联网免费查询。HWSD元数据参照ISO 19115核心元数据部分编写(<http://www.fao.org/geonetwork/srv/en/metadata.show?id=37140>)，GSOTER元数据同时按照美国联邦地理数据委员会(FGDC)元数据标准(FGDC 1998)和ISO 19115编写。FAO还提供了全球与土壤相关的地理空间数据集(<http://www.fao.org/geonetwork/srv/en/main.home>)，其元数据采用ISO 19115标准。

国内基于互联网的现存土壤科学数据资源包括：中国科学院南京土壤研究所开发的1:100万中国土壤空间数据库、中国土种数据库、养分循环长期试验数据库、土壤样品与标本数据库、土壤分类和参比数据库、中国生态系统研究网络土壤动态监测数据库；国家农业科学数据共享中心的全国数字土壤数据库，全国土壤肥料数据库，化肥参数数据库，有机肥参数数据库，农田土壤水分数据库；国家土壤肥力与肥料效益监测网的土壤养分数据集和作物产量及养分数据集；国家科技基础条件平台项目地球系统数据共享平台1:400万中国土壤专题图数据库、土壤肥力数据库；东北地理与农业生态研究所黑土资源数据库和水土流失专题数据库；中科院水利部成都山地灾害与环境研究所紫色土农田生态数据集和高山森林土壤观测数据集；中科院水利部水土保持研究所黄土高原水土保持统计数据库和黄土高原地区主要水文站水文泥沙数据集等。数据以传统的土壤

调查、观测试验数据为多，由于缺乏统一的土壤科学数据元数据标准，这些数据库分别参照支撑项目规定的元数据标准，例如中国科学院《科学数据库元数据标准》，科技部《科学数据共享元数据内容标准》，以及GB/T 20533 生态科学元数据和GB/T 19710 地理信息元数据执行。

国内农业部农业技术推广中心组织开发了县域耕地资源信息系统和县域测土施肥专家系统，以县级行政区域内耕地资源为管理对象、以土地利用现状与土壤类型的结合为管理单元，通过对资源的管理、分析和评价，指导合理科学施肥。该系统对涉及耕地资源空间数据和属性数据的数据元进行了详细地规范，但其数据库系统只在行业系统内使用，不提供跨行业共享，尚无高一层次的统一的元数据规范。

由此可见，土壤科学数据元数据标准的制定应以ISO19115为参考基准，增添与土壤科学研究相关的信息，其核心元数据应与现存国家标准兼容。元数据制定不涉及具体数据库的数据规范，没有现存的国标/行标与其冲突。

2.2.3 技术内容的确定方法

2.2.3.1 标准框架

参照GB/T 1.1-2009，正文内容分为7个章节（范围、规范性引用文件、术语和定义、符号和缩略语、约定、要求、元数据模式）。

元数据是土壤科学信息资源的规范化描述，它是按照一定规则，从土壤科学数据资源中抽取相应的特征，组成的一个特征元素集合，这种规范化描述可以准确和完备地说明土壤科学数据的各项特征。参照ISO 19100系列，土壤科学数据元数据采用UML和数据字典相结合的方式，定义描述土壤科学数据所需的概念模式，规定了关于土壤科学数据的标识、数据质量、方法、实体、空间表示、空间参照系和分发等方面信息的一组必选的、条件必选的和可选的元数据子集、元数据实体和元数据元素，并提供了对该概念模式的形式化描述。土壤科学数据元数据描述的对象是“数据集”。定义的元数据实体和元数据元素在第6、7章和规范性附录A 数据字典中阐明。土壤科学数据的元数据应包括附录D 核心元数据中必选的最少元数据元素。

标准适用于土壤科学数据集的编目和描述、数据集的组织管理以及数据交换网站的数据服务。

为保证元数据标准的可扩展性和实际可操作性，设置了规范性附录B元数据扩展和元数据专用标准。

规范性附录C抽象测试套件提供了与土壤科学数据元数据标准严格一致和一致的测试方法。

资料性附录E提供了我国的土壤分类系统及其参比说明。

资料性附录F为土壤科学数据元数据应用示例，采用纯文本格式，给出描述土壤科学数据的元数据示例，其内容仅作参考。

2.2.3.2 土壤科学数据集类型

传统的土壤科学数据包括野外调查、观测、试验等土壤科学研究活动产生的原始数据，通过地图和测量技术形成的空间数据(矢量、栅格)，和对原始数据进行加工而形成的数据。随着科学技术的发展，遥感数据已经引入到土壤科学领域之中。例如，通过遥感雷达影像生成的地面高程模型，已经成为土壤—地体数据库(Soil and Terrain Digital Databases)的重要基础。针对遥感影像和地理格网数据的应用趋势，TC211发布了《地理信息 影像和格网数据(ISO 19121)》和《地理信息 元数据 影像和格网数据的扩展(ISO 19115-2, 2009)》，提供了影像和格网数据的概念、模式和技术应用规范。因此，元数据的服务对象除了传统的土壤科学数据外，还应包括通过测量传感器、模型生成的影像和格网数据。

由此，将土壤科学数据集分类为5种表达方式：数字化文本(包括文档和电子表格)、关系数据库、具有属性和空间位置信息的矢量数据、具有属性值及其关联的格网坐标系位置的影像和格网数据以及非数字化文档资料。

ISO 19115 元数据管理的对象是数据集和数据集系列，考虑到实际应用时数据集和数据集系列很难严格区分，定义土壤科学数据元数据管理的对象是数据集。数据集可以通过空间范围区分，或者要素类型给予限制，不同方法获取的数据可能不具有可比性，宜区分为不同的数据集。数据集在物理上也可以是更大数据集较小的部分。一张非数字化的硬拷贝地图和图表均可以被认为是一个数据集。

2.2.3.3 元数据组织结构和特点

CSDGM (FGDC1998)元数据按层次结构来组织元数据记录的，包括7个主要模块和3个辅助模块。ISO19115在很大程度上参考了CSDGM的元素和结构，但ISO19115以UML来展示标准中的包、实体、属性，通过UML模型使各个实体之间的关系显性化，更便于元数据管理系统开发人员等的规划或开发工作。ISO 19115元数据由10个元数据子集和2个数据类型组成。采用国际标准的GB/T 19710在元数据结构上与ISO 19115完全相同，但在其地理信息共享领域元数据专用标准范例应用实例上进行了精简，设置了7个元数据子集(元数据信息、标识信息、限制信息、数据质量信息、参照信息、内容信息、分发信息)和2个数据类型(引用和负责单位信息、覆盖范围信息)。GB/T 20533考虑生态学研究的特点，除了标识信息、数据质量信息、空间表示信息、空间参照系信息、实体信息、元数据参考信息和分发信息7个基本元数据子集外，增添了场地、方法和项目元数据子集。

土壤科学数据元数据包括8个元数据子集，元数据信息、标识信息和数据质量信息是必选子集，方法信息、实体信息、空间参照信息、空间表示信息和分发信息是可选子集，覆盖范围和责任者是两个公用信息子集。每个子集由若干个实体(UML类)和元素(UML类属性)构成。公共信息子集覆盖范围信息和责任者信息由其他子集调用，不单独使用。

土壤科学数据元数据与相关标准组成结构的对比见表1。其中，括号中(M)表示必选，(O)表示可选，(C)表示条件必选。

表1 土壤科学数据元数据与相关标准组成结构的对比

土壤科学数据元数据	GB/T 20533-2006	GB/T 19710/ISO 19115	CSDGM(FGDC 1998)
元数据信息(M)	元数据参考信息(O)	元数据信息(M)	元数据参考信息(O)
标识信息(M)(包括场地和维护信息)	标识信息(包括维护 and 限制信息)	标识信息	标识信息(包括维护和限制信息)
数据质量信息(M)	数据质量信息(O)	数据质量信息(O)	数据质量信息(O)
方法信息(O)	方法信息(O)		
空间表示信息(C, 矢量、格网数据必选)	空间表示信息(O) (空间数据的表示)	空间表示信息(O) (空间数据的表示)	空间数据组织(O) (空间数据的表示与数据内容)
空间参照信息(C, 矢量、格网数据必选)	空间参照信息(O)	时间和空间参照信息(O)	空间参照信息(O)
实体信息(O)	实体信息(O)	内容信息(O)	实体与属性信息(O)
分发信息(O)	分发信息(O)	分发信息(O)	分发信息(O)
	场地信息(O) 项目信息(O)	维护信息(O) 限制信息(O)	
		图示表达信息 应用模式信息	

3. 主要试验分析、综述报告，技术经济论证，预期的经济效果

3.1 与现存标准的兼容性

土壤数据具有时间、空间属性。在时间尺度，土壤性质随着自然条件和人类活动影响而改变。

在空间尺度,土壤具有水平分布特征和垂直分布特征,单个土体中的不同层次,理化性质也不同。为了解土壤性质随空间尺度的变化,常通过区域调查和土壤制图(或数字化制图)的方式表征随环境变化土壤空间分布的模式。另一方面,同植物分类一样,对不同类型土壤的定量和规范化分类描述是数据解译、模拟、对比的基础。因此,土壤数据同时具有时间、空间和分类特征。

标准编制的原则是尽量引用现有的国家标准和权威性资料,核心元数据与现存的国家标准兼容。GB/T19710/ISO 19115《地理信息 元数据》中的关于数据集内容、覆盖范围、数据格式、质量、数据分发等核心元数据适合土壤科学数据的应用。

土壤科学数据集的核心元数据与GB/T19710/ISO 19115对比,为了快速查询数据,增添了必选的“关键词”,并将“覆盖范围”设为必选,增添了条件必选的“土壤分类”。考虑到数据质量的重要性,将GB/T 19710中可选的“数据志”改为必选的“数据质量说明”。将GB/T 19710中可选的“空间参照系”改为对于矢量和格网空间数据集条件必选的“空间参照系”。考虑到数据共享时涉及到知识产权,设置了可选的“拷贝权限”。本标准主要面向国内使用,将语种和字符集改为可选元素。见表2。

表2 土壤科学数据集核心元数据与GB/T 19710对比

土壤科学数据集 核心元数据	GB/T 19710/ISO 19115 核心元数据	土壤科学数据集 核心元数据	GB/T 19710 核心元数据
标题(M) (元数据>标识.标题)	数据集名称(M) (MD_元数据>MD_数据标识.引用>CI_引用.名称)	比例尺分母或分辨率(C) (元数据>空间表示>矢量空间表示.比例尺分母)或(元数据>空间表示>格网空间表示.分辨率)	数据集空间分辨率(O) (MD_元数据>MD_数据标识.空间分辨率>MD.分辨率.等效比例尺分母或MD_分辨率.采样间隔)
摘要(M) (元数据>标识信息.摘要)	数据集摘要说明(M) (MD_元数据>MD_数据标识.摘要)	数据集地理位置(M) (元数据>标识.数据覆盖范围>覆盖范围>空间范围.地理标识符)	数据集地理位置(C) (MD_元数据>MD_数据标识.覆盖范围>EX_覆盖范围>EX_地理覆盖范围>EX_地理区域描述或地理边界矩形)
负责方(M) (元数据>标识.负责方)	数据集负责单位(O) (MD_元数据>MD_数据标识.联系方>CI_引用.日期)	数据时间覆盖范围(M) (元数据>标识.数据覆盖范围>覆盖范围>时间范围.数据起始时间和数据结束时间)	数据集覆盖范围补充信息(垂向和时间的)(O) (MD_元数据>MD_数据标识.覆盖范围>EX_覆盖范围>EX_时间覆盖范围或EX_垂向覆盖范围)
语种(O) (元数据>标识.语种)	数据集采用的语种(M) (MD_元数据>MD_数据标识.语种)	土壤分类(C) (元数据>标识.覆盖范围>土壤分类.中国土壤发生分类土类)	
字符集(O) (元数据>标识.字符集)	数据集采用的字符集(C) (MD_元数据>MD_数据标识.字符集)	数据质量说明(M) (元数据>数据质量>质量报告.数据质量说明)	数据志(O) (MD_元数据>DQ_数据质量.数据志>LI_数据志)

表2 土壤科学数据集核心元数据与GB/T 19710对比 (续)

土壤科学数据集 核心元数据	GB/T 19710/ISO 19115 核心元数据	土壤科学数据集 核心元数据	GB/T 19710 核心元数据
学科分类(M) (元数据>标识>分类, 学科分类)	数据集专题分类(M) (MD_元数据>MD_数据标识, 专题类型)	空间参照系(C) (元数据>空间参照系)	参照系(0) (MD_元数据>MD_参照系)
关键词(M) (元数据>标识, 关键词)		拷贝权限(0) (元数据>分发, 拷贝权限)	
最后更新日期(0) (元数据>标识>维护, 最后更新日期)	数据集引用日期(M) (MD_元数据>MD_数据标识, 引用>CI_引用, 日期)	元数据创建日期(M) (元数据, 元数据创建日期)	元数据创建日期(M) (MD_元数据, 创建日期)
类型(M) (元数据>标识, 类型)	空间表示类型(0) (MD_元数据>MD_数据标识, 空间表示类型)	元数据责任者(M) (元数据, 元数据责任者)	元数据联系方(M) (MD_元数据, 联系>CI_负责单位)
数据格式(0) (元数据>分发, 数据格式)	分发格式(0) (MD_元数据>MD_分发>MD_格式, 名称和MD_格式, 版本)	元数据标准名称(M) (元数据, 元数据标准名称)	元数据标准名称(0) (MD_元数据, 元数据标准名称)
在线信息(0) (元数据>分发, 在线信息)	在线资源(0) (MD_元数据>MD_分发>MD_数字传输选项, 在线>CI_在线资源)	元数据标准版本(0) (元数据, 元数据标准版本)	元数据标准版本(0) (MD_元数据, 元数据标准版本)

土壤科学数据集的核心元数据与GB/T 20533—2006对比, 增添了必选的“数据集类型”和“数据质量说明”。为了快速查询数据, 将“覆盖范围”设为必选, 增添了条件必选的“土壤分类”, 及对于矢量和格网空间数据集条件必选的“比例尺分母或分辨率”及“空间参照系”, 考虑到本标准主要面向国内使用, 将语种改为可选元素。见表3。

为了与现存国内外标准兼容, 数据字典中的中文名称、英文名称、英文缩写和定义, 尽量与GB/T 19710/ISO 19115或ISO 19115 (勘误表,2006)一致。GB/T 19710中没有的内容参考GB/T 20533或FGDC 1998。

表3 土壤科学数据集核心元数据与GB/T 20533对比

土壤科学数据集 核心元数据	GB/T 20533 核心元数据	土壤科学数据集 核心元数据	GB/T 20533 核心元数据
标题(M) (元数据>标识, 标题)	题名(M) (元数据>标识, 题名)	比例尺分母或分辨率(C) (元数据>空间表示>矢量空间表示, 比例尺分母)或(元数据>空间表示>格网空间表示, 分辨率)	

表3 土壤科学数据集核心元数据与GB/T 20533对比(续)

土壤科学数据集 核心元数据	GB/T 20533 核心元数据	土壤科学数据集 核心元数据	GB/T 20533 核心元数据
摘要(M) (元数据>标识信息.摘要)	摘要(O) (元数据>标识信息.摘要)	覆盖范围(M) (元数据>标识.数据覆盖范围>覆盖范围>空间范围.地理标识符及覆盖范围>时间范围.数据起始时间和数据结束时间)	覆盖范围(O) (元数据>标识.数据覆盖范围>覆盖范围>空间范围或时间覆盖范围)
关键词(M) (元数据>标识.关键词)	关键词说明(M) (元数据>标识.关键词说明)	土壤分类(C) (元数据>标识.覆盖范围>土壤分类.中国土壤发生分类土类)	
类型(M) (元数据>标识.类型)			方法(O) (元数据>方法>方法文件或研究方法或数据志)
负责方(M) (元数据>标识.负责方)	创建者(O) (元数据>标识.创建者)	数据质量说明(M) (元数据>数据质量>质量报告.数据质量说明)	
最后更新日期(O) (元数据>标识>维护.最后更新日期)	日期(M) (元数据>标识.日期)	空间参照系(C) (元数据>空间参照系)	
语种(O) (元数据>标识.语种)	语种(C) (元数据>标识.语种)	拷贝权限(O) (元数据>分发.拷贝权限)	知识产权声明(O) (元数据>标识>限制.知识产权声明)
字符集(O) (元数据>标识.字符集)		元数据创建日期(M) (元数据.元数据创建日期)	元数据日期(M) (元数据>元数据参考.元数据日期)
学科分类(M) (元数据>标识>分类.学科分类)	学科分类(M) (元数据>标识>分类.学科分类)	元数据责任者(M) (元数据.元数据责任者)	元数据联系方(M) (元数据>元数据参考.元数据联系方)
数据格式(O) (元数据>分发.数据格式)	分发格式(O) (元数据>分发.分发格式)	元数据标准名称(M) (元数据.元数据标准名称)	元数据标准(O) (元数据>元数据参考.元数据标准)
在线信息(O) (元数据>分发.在线信息)		元数据标准版本(O) (元数据.元数据标准版本)	

3.2 依据土壤学科特点，增添土壤分类信息。

受复杂自然条件的影响和深刻人为因素的影响，土壤千差万别类型繁多。土壤分类是土壤信息载体、标准化的基础，是土壤科学交流的媒介，也是土壤科学数据共享的比不可少的因素。国际上还没有一个统一的土壤分类，只是美国土壤系统分类(ST)和WRB(World reference base for

soil resources)分类在国际土壤学术交流和合作研究中得到广泛应用，成为当今世界的主流土壤分类系统。目前在国际上进行学术交流、在国际杂志上发表论文通常都使用这两个分类系统的名称。自从上世纪30年代我国开展土壤研究工作以来，数以万计的各种土壤数据都是长期应用中国土壤发生分类积累起来的，但这个分类与国际上广泛应用的美国系统分类的学术思想大不一样。由于国际上不了解中国土壤发生分类，中国科学家在国际上进行学术交流，开展国际合作和数据共享都遇到了很大的困难。在美国土壤系统分类的影响下，从1984年起南京土壤所牵头开展了以诊断层和诊断特性为基础的中国土壤系统分类研究，2001年分别出版了《中国土壤系统分类检索（第三版）》和《中国土壤系统分类检索（英文版）》。《中国土壤系统分类研究》获得2005年国家自然科学二等奖。这个分类系统使用日益普及，在国内外的影响正在不断扩大。

土壤科学数据应用主要是国内共享和国际共享，元数据表达的标准化、可比性是衡量元数据标准质量的指标。考虑到我国土壤分类处于土壤发生分类与系统分类并存阶段，进行两者的参比对充分利用已有资料，加强国际交流具有重要意义。当数据集有土壤类型属性时，定义国内普遍应用的土壤发生分类（GB/T 17296-2009 中国土壤分类与代码）为必选信息，与国际主流分类系统接轨中国土壤系统分类作为用户填写的可选信息。此外，标准中还设置了可选的其他分类系统信息，以满足国际主流土壤分类系统的填写需要。用户可以填写ST或WRB，或其他分类系统（如地方土壤分类系统）。

关于土壤分类的内容有11位专家提出了意见，其中建议两种分类系统共存，并对应到国际主流土壤分类的意见和建议有7条，涉及土壤分类概念的有2条，建议只涉及土壤发生分类的有1条，建议标准不涉及两种分类系统参比的1条。根据专家的意见，增添了关于土壤分类概念与参比的资料性附录，供使用者填写参考。

3.3 数据的可用性

数据共享的目的是为了更多的人使用数据，元数据应提供关于土壤科学数据集数据质量信息。数据质量作为必选的元数据子集，包括必选的“数据质量报告”和可选的“数据志”。考虑到科学试验数据对于质量评估的困难，将GB/T 19710/ISO 19115中的“数据质量报告”实体中的“评价结果”改为“质量说明”。参考 GB/T 21337/ISO 19113 地理信息质量原则和编写组成员对土壤长期观测数据质量评估的经验，增添数据质量评估指标为准确度、精密度、完整性和一致性。

从数据使用的角度考虑，对于野外长期定位观测和科学试验，侧重于在特定场地用特定的研究方法(或特殊的方法)观测土壤性质的变化，需说明数据产生的背景，包括场地、研究目的和观测试验方法。环境参数对土壤科学研究成果的获取是重要的。根据美国密西根大学对105位来自15个国家的土壤微生物学家调查，综合评判元数据中环境参数获取的重要性和困难性，得到对于土壤科学数据元数据重要的前5个参数：场地描述、采样描述、天气、土壤分类和其他土壤分析指标。对于野外长期定位试验和定位观测，数据是不可重复的，扰动可能表征着人类活动或气候变化对土壤参数的影响，而这些扰动因素恰恰容易被数据的管理者忽视。根据中国生态系统研究网络土壤分中心对野外台站土壤观测数据质量控制的研究，在911条关于观测场地或试验设计或采样的记录中，数据缺失的出现比率为10.4%。因此，参考GB/T 20533《生态科学数据元数据》，在标识信息中增添了关于场地的元数据，并增添方法信息子集。方法信息包括必选的“方法步骤说明”，可选的“试验设计”信息、“观测（采样）”信息、“样品分析”信息和“仪器平台”信息。

对于描述土壤性质空间变异分布情况的区域调查、数字化土壤图件及影像数据，使用者需了解其空间表示方法、空间数据集的内容和用于制图的空间参照系。GB/T 19710《地理信息元数据》关于空间表示、内容和空间参照系的部分内容适合土壤科学元数据的应用。根据土壤科学空间信息应用的进展，参考ISO 19115-2地理信息 元数据 第2部分：影像和格网数据的扩展，在空间表示信息子集中对矢量和格网数据的表示和内容进行了说明，当格网类型为“影像”时，对影像数据进行说明。对于空间参照系，根据土壤专业使用需要，提供常用的大地基准面列表供选择。土

壤科学数据的时间参照大都基于公历，故没有专门设置时间参照系。

3.4 可操作性

标准提供了若干个元数据应用示例，包括：单个长期试验地点组成盐亭站养分循环长期定位试验数据集元数据；单个图层组成黑龙江省黑土资源分布图元数据；侧重数据质量评估并由多个关系表格组成的中国生态系统研究网络土壤养分长期监测数据集元数据；多个GIS图层组成的1:400万土壤有效态微量元素空间分布数据集元数据。这些应用示例基本覆盖了土壤科学数据的范畴，为实际应用提供了范例。

对于在土壤科学研究中只用于科研人员一般了解，不作为进一步查询的数据元素，尽量简化结构，采用一般性的文本描述。在上述元数据示例填写过程中，发现原实体信息元数据子集设计过于繁杂，参照ISO19115，对该部分进行简化。

3.5 预期效果

通过本标准的编制，确定描述土壤科学数据所需的元数据的数据项集合、各数据项语义定义和著录规则等，从而为土壤科学数据的管理、发现和获取提供一种实际而简便的方法。使用者能够通过元数据快速发现所需的土壤科学数据，可以详细、深入地了解数据的格式、质量和获取方法等各方面细节；土壤科学数据的生产者可以利用元数据进行数据维护。

4. 与现行法律、法规和强制性国家标准的关系

科学数据资源的公开与共用，是信息时代发展的必然和科学技术自身发展特点所决定的。标准编制中依据中国科学技术蓝皮书第4号《信息技术发展政策》、《著作权法实施条例》、《计算机软件保护条例》、《计算机信息系统安全保护条例》、《国务院关于进一步加强对知识产权保护工作的决定》、《计算机信息网络国际联网安全保护管理办法》及国家有关科学数据共享和保护知识产权方面的法律法规等。

标准与已颁布实施的GB/T 19710/ISO 19115 地理信息元数据，GB/T 20533 生态科学元数据，GB/T 17296 中国土壤分类与代码，GB/T 7408数据元和交换格式 信息交换日期和时间表示法，GB/T 16831地理位置的纬度、经度和高程的标准表示法，GB/T 13745学科分类与代码，GB/T 21337-2008/ISO 19113 地理信息 质量原则，以及ISO 19115-2 Geographic information—Metadata—Part 2: Extensions for imagery and gridded data等标准相协调。

5. 重大分歧意见的处理经过和依据

来自农业部、科研院所、大专院校和野外台站的20位专家及工作组成员就标准的范围、框架结构、与现存国家标准的关系、土壤分类和如何推广提出了52条意见，其中采纳37条，部分采纳11条，未采纳4条。

《土壤科学数据元数据》国家标准编写组

