

地球空间数据元数据标准初探

李 军 周成虎

(中国科学院地理研究所资源与环境信息系统国家重点实验室, 北京 100101)

摘 要 论文首先概要介绍了目前有影响的几个地球空间数据元数据标准, 然后从地球空间数据及数据集自身特征出发探讨了数据集元数据的内容, 提出了地球空间数据集的层次结构, 并给出了基本层和基础层元数据内容。

关键词 元数据标准 地球空间元数据 数据集元数据

地球空间数据元数据标准既应有元数据的共性, 也应反映地球空间数据自身的特征。同其它数据相比地球空间数据较为复杂, 它既涉及到空间位置, 也有属性数据及其之间的关联, 所以地球空间元数据标准的建立是项复杂的工作。由于种种原因, 某些数据组织或数据用户开发出来的标准很难被广泛接受, 对地球空间元数据标准的研究是地球空间元数据中最受关注和最迫切的任务。元数据标准的建立是为地球空间数据的使用和共享服务的, 是面向各类数据生产者和数

据用户的。由于现在已有的地球空间数据许多没有建立起相应的元数据, 因而地球空间元数据标准不宜太复杂, 以降低元数据制作的费用和被更多的数据使用者所接受。目前, 地球空间元数据已形成了一些区域性或部门性的标准 (见表 1), 正象数据转换标准一样, 由于人为的和客观的原因, 仍没有一个标准成为全球统一的空间数据元数据标准。

表 1 目前已使用的主要元数据标准状况

(据 David Medyckyj-scott, 1996)

Tab. 1 Current standard geo-spatial metadata
(after David Medyckyj-scott, 1996)

元数据标准名称	建立标准的组织
地学空间数据元数据内容标准	美国联邦地球空间数据委员会
GDDD 数据集描述方法	MEGRIN, 欧洲地图事务组织
CGSB 地球空间数据集描述	加拿大标准委员会
CEN 地学信息- 数据描述- 元数据	CEN/TC287
DIF	NASA
ISO 地理信息	ISO/TC211

1 数字地球空间数据元数据标准

数字地球空间元数据内容标准 (Content Standard for Digital Geospatial Metadata) 由美国联邦地球空间数据委员会组织编写并发布。该标准 1992 年 7 月开始起草, 几经修改, 1994 年 7 月 8 日, FGDC 正式确认该标准为美国国家地球空间数据元数据标准, 并于 1997 年 4 月发布其修订版 (FGDC 1994, FGDC 1997)。

1.1 标准的格式与特点

该标准格式与特点如下:

(1) 标准的内容项采用数字式章节组织结构, 即: 1, 1.1, 1.1.1.....

(2) 数据要素分为单值数据要素 (Elements) 和混合数据要素 (Compound Elements), 前者为数据的逻辑基本项; 后者表示用单个数据要素不能表示的更高层的概念, 它为元数据的抽象聚类处理提供了依据。

(3) 元数据内容中可选项与必选项并存。用特定的符号来表达内容项的必要性, 并建立用于表示元数据要素之间及混合要素中数据要素间逻辑关系的规则, 该规则由三部分组成即符合、符合左侧的标识符和符合右侧的表达式。标准中用到的符合及其含义如下:

= 取代或由什么组成

+ 和

[|] 选择, 从括号内的列项中选择一个, 列项用 | 符号分割开

m { } n 重复, 括号中的项重复 m 到 n 次

() 任选, 括起来的项为任选项

如: $a = b + c$ 表示 a 由 b 和 c 组成

$a = [b | c]$ 表示 a 由 b 或 c 组成

$3 \{b\} 5$ 表示 b 出现 3 到 5 次

$a = b + (c)$ 表示 a 由 b, 或由 b 和 c 组成

(4) 数据项丰富。该标准中数据项的总数有 219 项, 从各个方面对地球空间数据进行了全方位的说明。

1.2 元数据的内容

该标准将元数据内容分为 7 个部分, 分别是数据标识信息、数据质量信息、空间数据组织信息、数据空间参考消息、实体及属性信息、数据传播及共享信息和元数据参考信息, 每部分又由若干项组成, 标准的整体结构和各部分的内容如下:

(1) 元数据 = 数据标识信息 + 0 {数据质量信息} 1 + 0 {空间数据组织描述} 1 + 0 {数据空间参考消息} 1 + 0 {实体及属性信息} 1 + 0 {数据传播} 1 + 共享和元数据参考信息;

(2) 数据标识信息 = 引述 + 描述 + 内容的时间周期 + 数据状态 + 空间范围 + 关键词 + 访问限制 + 使用限制 + (访问点位置) + (1 {浏览图形} n) + (数据集信誉) + (安全信息) + (数据集内环境) + (1 {其它参考信息} n);

(3) 数据质量信息 = 0 {属性精度} 1 + 逻辑一致性报告 + 完整性报告 + 0 {位置精度} 1 + 继承性 + (覆盖云量);

(4) 空间数据组织描述 = 0 {间接空间参考} 1 + 0 {直接空间参考方法} ([点与矢量对象信息 | 栅格对象信息]) 1;

(5) 数据空间参考消息 = 0 {水平坐标体系定义} 1 + 0 {垂直方向上坐标体系的定义} 1;

(6) 实体及属性信息 = [1 {详细描述} n | 1 {概述} n | 1 {详细描述} n + 1 {概要描述} n];

(7) 数据传播 = 数据传播者 + 0 {信息源描述} 1 + 传播义务 + 0 {正常订购手续} n + 0 {用户订购手续} 1 + (技术要求) + (可能的时间周期);

(8) 元数据参考信息 = 元数据形成日期 + (元数据检查时间 + (元数据未来检查时间))

+ 元数据联系方法+ 元数据的标准名称+ 元数据标准版本+ 0 {元数据约定时间类型} 1+ (元数据访问限制) + (元数据使用限制) + (元数据安全信息) + 0 {元数据扩展} n_0

2 ISO/TC211 元数据标准

ISO/TC211 元数据标准由国际标准化组织 (the International Organization for Standardization) 第三工作组组织研究, 项目编号为 15046- 15。1996 年 2 月 9 日通过 1.0 版草案, 后几经修改, 于 1997 年 1 月 20 日发布 2.0 版标准 (ISO/TC211, 1997)。

2.1 元数据标准格式与特点

(1) 元数据内容项分为三种类型: 必须型 (M) 指必须给出的内容 (M 是 Mandatory 的缩写)、条件型 (C) 指在一定条件下需要给出的内容 (C 表示 Conditional)、可选型 (O) 指可有可无的内容 (O 表示 Optional);

(2) 元数据内容采用逐项逐行方式表达;

(3) 标准中给出了元数据制作、管理等规范。

2.2 元数据标准的内容

该标准中把元数据的内容分为 7 类, 每一类中又包括若干子类或具体元数据项, 以下列出该标准中较高层次的元数据内容 (内容项后的字母为数据项的类型标识):

(1) 元数据内容包括: 标识信息、数据质量信息、空间数据表达信息、空间参考信息、特征与属性信息、数据传播信息、元数据参考信息、引述信息、联系信息, 其中最后两部分内容为数据集使用的推荐参考信息;

(2) 标识信息内容包括: 引述 (M)、数据使用 (M)、概要 (M)、状态 (O)、扩展 (M)、数据领域 (O)、访问位置 (M)、图形浏览 (M)、数据集限制 (C)、其它参考信息 (O);

(3) 数据质量信息内容包括: 数据质量 (C)、世系 (C);

(4) 空间数据表达信息内容包括: 空间表达信息 (M);

(5) 空间参考信息内容包括: 空间参考系统 (C)、平面坐标系系统定义 (C)、垂直坐标系系统定义 (C);

(6) 特征与属性信息内容包括: 特征与属性描述 (M)、特征类别描述 (C)、特征类别信息 (C);

(7) 数据传播信息内容包括: 数据发布者信息 (C)、正常数据购买手续 (C)、可能的获取时间周期 (O);

(8) 元数据参考信息内容包括: 元数据时间信息 (C)、元数据访问 (M)、元数据标准信息 (M)、元数据辅助信息 (O)、元数据使用限制 (C)。

3 其它元数据标准

除了上述两个标准外, 元数据还有许多标准, 限于论文的篇幅不可能对之逐一描述, 每种元数据标准都有自己的特色, 并且在处理某些问题中占有一定的优势。下面简单介绍另一个元数据标准即 DIF。

目录交换格式- D I F (Directory Interchange Format) 由美国N A S A 组织开发。1989 年开始在N A S A 使用, 该标准是为卫星及其它遥感数据而设计的, 它提供了描述数据的元数据要素、规范所选要素的内容值, 并为元数据在不同系统之间交换提供一种数据结构。之后该标准的内容根据 FGDC 的元数据标准的内容做了些改动。

规则只有两条: * 为表示必选字段, + 为可重复的字段。
标准采用组 (group) 结构, 每组内包括相关的各种内容, 组可以嵌套。

4 基于集成应用的地球空间数据集元数据体系

4.1 地球空间元数据建立的依据

通过对地球空间数据的研究发现: 地球空间数据实质是对地学过程的抽象、概括、整体或部分分层次的表达, 并把具有同种特征的地学过程或地学过程的片段组织到一起形成的数据集, 所以, 详细全面描述数据集应从数据的本质入手 (见图 1)。

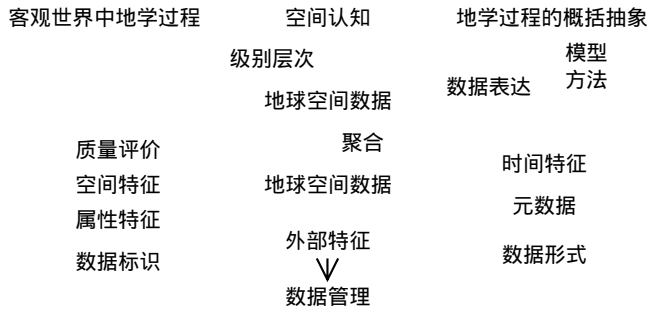


图 1 基于地球空间数据特征的元数据组成及来源示意图

Fig. 1 The components of Metadata based on the basic characteristics of geo-spatial data

地球空间数据集是对地学过程、现象、特征的认识表达的结果, 所以数据集元数据首先应包括对数据表达特征的描述; 内容是依附于形式存在的, 故元数据应包含数据集整体特征的信息。由上图知道数据集元数据至少应包含质量评价、空间特征、属性特征、时间特征、元数据、数据标识、数据管理、数据形式等内容。其中每一项又都包括许多具体的次级内容。

4.2 地球空间数据集元数据

目前, 许多国家、行业、国际组织已建立了自己的元数据标准, 但或多或少都有一些问题 (Tamas Prajczar 1996), 如已成为国家标准的 FGDC 的元数据内容标准也存在不少问题 (Timothy W. Foreman 1996), 如数据要素有 219 项, 获得如此多的元数据内容, 需要增加 15% ~ 20% 的工作时间, 编写该标准的元数据要大量的时间和资金, 并且要专业人员才能完成; 其中含有大量的专业词汇, 没有一定专业背景的用户很难使用; 标准不实用, 由于其专业性太强, 从而限制了大学科范围内统一地球空间数据的形成, 因此有必要在应用上述标准的过程中提出另一种更容易被多数用户接受和使用的多层次的元数据标准, 不同用户只用到不同层次。考虑到数据用户对数据要求的层次性和已有元数据内容标准的复杂

性, 我们认为地球空间元数据应该分为不同的层次以服务于不同的数据用户, 由于元数据的可抽象性、简洁性等特点要求, 这种结构是可以实现并且是必须的。故我们设计将地球空间元数据分为三个层次(见图 2), 第一层次为元数据基本集, 服务对象为非地学领域的人员了解和检索数据集, 它描述了数据集最基本、一般数据用户最想了解的特征; 第二层次为元数据基础集, 服务对象为有一定地学知识背景的非数据处理和管理目的的数据用户, 它描述了数据集较详细的特征, 把基本集中的元数据内容做了进一步的具体化; 第三层次为元数据详细集, 详细列出了描述数据集内容的各种要素项, 服务对象是专业数据处理和管理人员操作控制数据库。

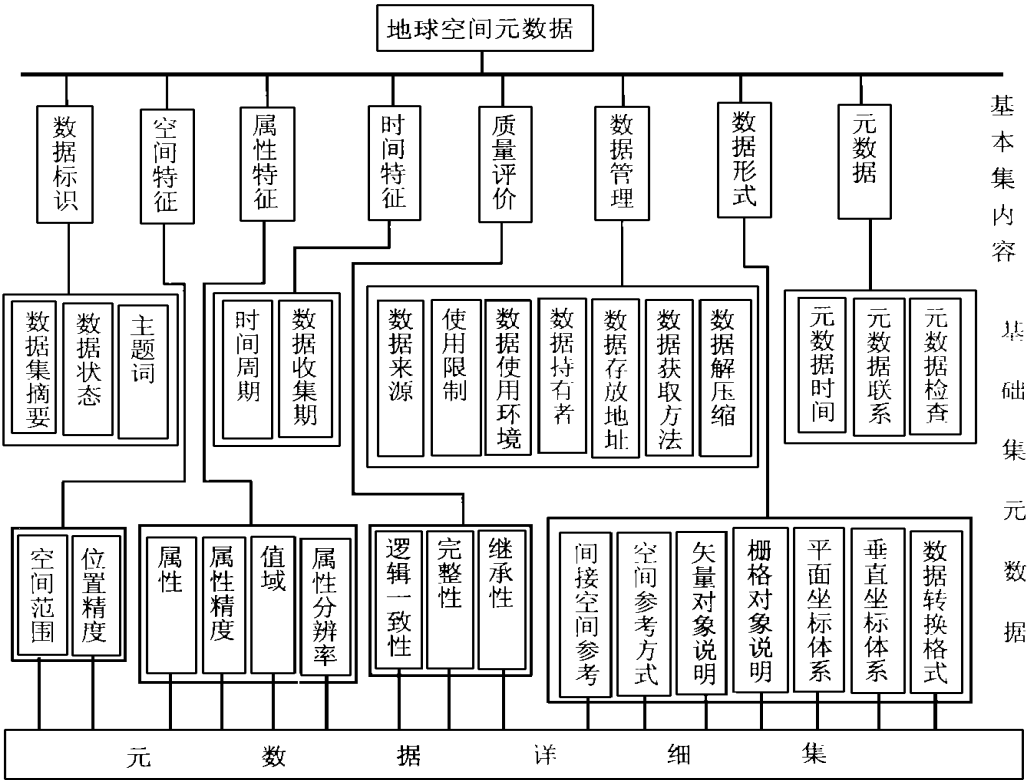


图 2 地球空间数据集元数据内容及结构

Fig. 2 The structure and contents of metadata for geo-spatial dataset

各层次的元数据内容存在着密切的联系, 对于元数据制作者来说只需要创建元数据详细集, 基本集和基础集的形成可以通过两种途径实现: 根据元数据的可抽象和归并特征由系统或专业软件自动形成上层元数据, 如: 根据基础集中空间特征和属性等的质量评价可以得出基本集中数据质量项的值; 或从下级元数据中抽取有典型意义的元数据项作为上层元数据集中对应项的内容。

4.3 数据基础集标准内容及其含义

数据集基本集元数据的内容过于简单, 一般不能满足数据用户的需求; 元数据详细集内容过于复杂, 使用起来并不方便, 从实际应用和元数据管理的角度出发, 我们认为基础集元数据已能够满足多数用户的需要, 所以这里重点介绍基础集元数据内容, 见表 2。

表 2 基础集元数据
Tab. 2 Metadata of fundamental dataset

标号	元数据项	元数据项内容描述	格式说明
11	数据集标识	数据集标识号，可以在数据库建立时添加	文本格式
12	数据集名称	在数据库中该数据集的英文名称和相应的中文名称	文本格式
13	比例尺（精度）	数据集的比例尺大小，表达的格式为用文本形式给出比例尺分母的数字，如数据的比例尺为 1 500000，则该项的表示为 500000	数字形式
14	时间域说明	数据内容表示的时间域“起始-终止日期”如：1950 年 3 月到 1980 年 3 月表达为“195003- 198003 ”	数字形式
15	时间周期	数据内容时间说明，数据表达的时间，如 1990 年 3 月表达为“199003 ”	数字形式
16	数据状态	数据集完成状态说明，如“已完成 ”	固定文本格式
21	空间最北位置	数据空间轮廓北至点的纬度坐标，度格式，如北纬 72 度 30 分表示为“72.5 ”	数字形式
22	空间最南位置	数据空间轮廓南至点的纬度坐标，度格式，如南纬 72 度 30 分表示为“- 72.5 ”	数字形式
23	空间最西位置	数据空间轮廓西至点的经度坐标，度格式，如东经 113 度 36 分表示为“113.6 ”	数字形式
24	空间最东位置	数据空间轮廓北东点的经度坐标，度格式，如西经 113 度 36 分表示为“- 113.6 ”	数字形式
31	数据来源	提供该数据集或该数据集部分构件的个人或单位，单位的中英文名称及名称缩写	文本格式
32	使用限制	用于使用该数据集的限制状况，内容有：“有限制”、“无限制 ”	固定文本格式
33	环境要求	使用数据集需要的硬软件环境，主要指使用的软件名称等	文本格式
41	空间参考方式	表达空间对象的系统，即以何种形式表达空间特征，内容有：“矢量”、“栅格”或其它（应具体说明）	固定文本格式
42	矢量对象说明	对于矢量数据，说明矢量对象的类型，类型仅限于说明是弧段、封闭多边形、环等	文本格式
43	栅格对象说明	说明栅格的类型，类型包括：“像元”、“栅格单元”、“voxel”（三维数据）等	固定文本格式
44	平面坐标体系	用于量测空间点位置的参考体系，如：经纬度	文本格式
45	数据投影类型	投影名称说明，如：等积圆锥投影	文本格式
451	椭球体	数据投影类型中采用的椭球体名称	文本格式
452	第一标准纬线	第一条标准纬线的度数	数字格式
453	第二标准纬线	第二条标准纬线的度数	数字格式
454	中央经线	中央经线的度数	数字格式
455	向北偏移量	数据坐标相对于原点向北的偏移量	数字格式

续表

标号	元数据项	元数据项内容描述	格式说明
456	向东偏移量	数据坐标相对于原点向东的偏移量	数字格式
457	坐标原点	投影的坐标其始位置坐标	数字格式
458	其它参数	其它投影参数的说明	数字或文本
46	数据单位	数据的度量单位描述	文本格式
47	垂直单位	数据表达的垂直方向上的单位描述, 如: 米、公里	文本格式
48	数据量	指定格式下数据需要磁盘空间的大小, 如: 米、英尺	数字格式
51	数据实体	数据实体表达的内容说明	文本格式
52	值域	数据集中属性值的取值范围及度量单位	数字格式
53	垂直分辨率	单位垂直距离内样点的数量, 整数表达	数字格式
54	属性分辨率	能区分的最小数据性值数量	数字格式
61	数据持有者	数据拥有单位或个人的名称	
62	数据拥有者通讯地址	数据拥有者的通讯地址	文本格式
63	数据拥有者电话	数据拥有者的联系电话	文本格式
64	网络地址	数据浏览或存放的网络地址	文本格式
65	数据获取方法	获取数据集方法说明, 如网络下载、购买光盘	文本格式
66	数据费用	获取该数据集需要的费用说明 (不同用户群分别说明)	文本格式
67	数据转换格式	数据在计算机或与计算机兼容介质上存放的格式, 内容是数据格式的名称如: “ASCII”、“B L”、“DXF”、“SDTS”等	固定文本格式
68	数据解压缩	是否使用了数据压缩格式, 压缩和解压方法说明, 使用的软件说明	文本格式
69	技术支撑	用该数据集用户应有的技术基础	文本格式
71	元数据时间	元数据形成的时间	时间字符串
72	元数据制作	元数据制作单位名称 (或个人)	文本格式
73	元数据更新	元数据更新周期与方法	文本格式
81	其它信息	上述未说明具体应用中需要的信息说明 1	文本格式
82	其它信息	上述未说明具体应用中需要的信息说明 2	文本格式

注: 固定格式文本指特定关键词文本字符串。

5 问题与说明

元数据的内容与具体应用及数据的内容有一定的关系, 已建立的地球空间数据元数据标准虽然有各种各样的问题存在, 但它们对规范化元数据有一定的参考价值。根据我国目前对数据共享需求及元数据在数据集成与共享中的独特作用, 最紧迫的任务是首先形成适合于中国地球空间数据现状的元数据标准。本文中提出的多层次的元数据标准只是一个参考标准, 对于具体的应用项目都或多或少有一些问题, 因而元数据在使用中应该有一定的修改。

根据地球空间数据基础元数据集的内容标准, 下面给出了某一植被数据集的基础集元数据内容, 见表 3。

表 3 地球空间元数据示例
Tab. 3 Examples of geo-spatial metadata

标号	元数据项	元数据内容描述
11	数据集标识	Chndata- 113
12	数据集名称	植被类型
13	比例尺 (精度)	400 000 000
14	时间域说明	1980~ 1990
15	时间周期	
16	数据状态	已完成
21	空间最北位置	+ 54.000 0 °
22	空间最南位置	+ 3.866 7 °
23	空间最西位置	+ 72.000 0 °
24	空间最东位置	+ 135.500 0 °
31	数据来源	数据源由中国地图出版社提供, 数据内容由中国科学院植物研究所编制
32	使用限制	无限制
33	环境要求	能接受 ARC/INFO 之 .E00 格式数据的软件 (如 Map Info 等)
41	空间参考方式	以矢量线组成的多边形图斑表示土壤的类型的边界投影为, 参数分别为
42	矢量对象说明	封闭多边形
43	栅格对象说明	
44	平面坐标体系	大地坐标系, 平面坐标
45	数据投影类型	等积圆锥投影
451	椭球体	Clark 1866
452	第一标准纬线	+ 25.000 0 °
453	第二标准纬线	+ 47.500 0 °
454	中央经线	+ 110 °
455	向北偏移量	0.000 0
456	向东偏移量	0.000 0
457	坐标原点	0.000 0 0.000 0 °
458	其它参数	
46	数据单位	m
47	垂直单位	
48	数据量	7.543 MB
51	数据实体	植被类型
52	值域	属性为 7 位数字编码
53	垂直分辨率	
54	属性精度	最小图斑的面积是 10 394.46 m ² , 代码为 1 104 007, 表示的植被类型为温带常绿阔叶林中的松林
61	数据持有者	中国科学院地理研究所资源与环境信息系统国家重点实验室 The state key Laboratory of Resources and Environment Information System, CAS.
62	数据拥有者通讯地址	中国, 北京, 安外大屯路 917 大楼, 100101 Building 917, Datun Road, Anwai, Beijing, 100101
63	网络地址	http: //reis.ac.cn
64	数据拥有者电话	086-10-64889764 Fax: 086-10-64889630
65	数据获取方法	网络获取或购买光盘, 或共享交换使用
66	数据费用	光盘版数据全套为 1 000 RMB (教学科研用)
67	数据转换格式	ARC/INFO "COVERAGE"、".E00"
68	数据解压缩	非压缩格式
69	技术支撑	地球空间数据概念
71	元数据时间	1997 年 9 月
72	元数据制作	中科院地理所资源与环境信息系统国家重点实验室
73	元数据更新	元数据更新周期为 6 个月
81	其它信息	1999 年更新
82	其它信息	

参 考 文 献

- 1 David Medykjy- scott, Mike Cuthbertson, Ian Newman. Discovering environmental data: metadatabases, network information resource tools and the GEN IE system. *Int. J. Geographical Information Systems*, 1996, **10**(1) 65~ 84
- 2 FGDC. Content standards for digital geo-spatial metadata. *Federal Geographic Data Committee*, 1994, 6
- 3 FGDC. Content standards for digital geo-spatial metadata. *Federal Geographic Data Committee*, 1997, 6
- 4 ISO/TC211. Geographic information- metadata. ISO Standard 15046- 15 Metadata, Version 2.0, 1997.
- 5 Tamas Prajczet. Metadata initiatives in hypertext in geographical information, from research to application through cooperation, Volume 1, Edited by M. Rumor R. Macmillan and H. F. L. Ottens, Ohmsha DS, 1996 268~ 275
- 6 Timothy W., Foreman H V., Wiggins D L P. Metadata myth: misunderstanding the implications of federal metadata standards, first IEEE metadata conference april. NOAA Auditorium, Silver Spring, Maryland, 1996, 16~ 18

OVERVIEW ON METADATA STANDARDS OF GEO-SPATIAL DATA

Li Jun Zhou Chenghu

(State Key Laboratory of Resources and Environment Information System,
Institute of Geography, CAS, Beijing 100101)

Abstract

Dozens of Metadata standards for geographic information or geo-spatial data have been found recently. The paper introduces Content Standards for Digital Geo-spatial Metadata, Geographic Information - Metadata, Directory Interchange Format, and Metadata standards of ISO/TC211. Based on those Metadata standards, the authors give one metadata standard for geo-spatial data application. The Metadata standard includes three level Metadata-sets. The first, named Basic Metadata-set, includes fundamental information of Geo-spatial dataset such as content, name, format, and it is introduced to serve ordinary data users. The second level Metadata-set, named Summary Metadata-set, is established to serve geo-spatial data managers. And the third level Metadata-set, named Detail Metadata-set, includes detailed information about Geo-spatial data, serving data specialists to manipulate and process geo-spatial data.

Key words metadata standards, geo-spatial metadata, metadata of geo-spatial dataset

作 者 简 介

李军, 男, 30 岁, 中国科学院地理研究所资源与环境信息系统国家重点实验室博士, 主要从事地球空间数据集成研究。已发表数篇相关论文, 博士毕业论文题目是《地球空间数据集成基础研究及其应用》。