

面向对象方法在元数据研究中的应用

吴 萍, 许惠平, 陈华根

(同济大学 海洋与地球科学学院, 上海 200092)

摘要: 面向对象设计技术, 将元数据标准中的元数据子集、元数据实体和元数据元素抽象为面向对象中的对象类和属性, 将元数据管理中涉及的对象抽象为对象类, 提出了元数据标准和元数据管理的对象模型. 这种设计理念增强了元数据标准之间的兼容性、元数据管理的可扩展性及代码重用.

关键词: 元数据; 面向对象设计; 元数据管理

中图分类号: P 208

文献标识码: A

Applications of Object-Oriented for Metadata Research

WU Ping, XU Huiping, CHEN Huagen

(College of Ocean and Earth Science, Tongji University, Shanghai 200092, China)

Abstract: Multi-standard coexistence and metadata management system reconstruction are presented in metadata research. The paper presents object models of metadata standard and metadata management system based on object-oriented design. Metadata section, metadata entity and metadata element are abstracted to be classes and properties, and the objects involved in management system are abstracted to be classes. In this way, compatibility between standards is strengthened. Management system extensibility and code reuse is improved.

Key words: metadata; object-oriented design; metadata management system

随着计算机和地理信息系统 (geographic information system, GIS) 技术尤其是网络通信技术的发展, 地理空间数据呈几何级数增长. 如何有效地存储、管理和应用海量数据, 成为 GIS 专家面临的一大挑战. 作为“数字地球”六大核心技术之一的元数

据成为解决这一问题的有效途径^[1], 是近十年来 GIS 研究的热点之一.

元数据标准是元数据应用的数据基础, 元数据管理是元数据应用的实现工具. 目前, 元数据应用存在以下问题: ①由于不同的研究对象和项目需求, 产生了多种元数据标准, 而这些标准相互之间的兼容性很差, 造成了元数据标准制定的重复做工; ②现有的元数据管理是针对某一特定的元数据标准而构建, 元数据标准的不同导致元数据管理的重复构建和编程, 浪费大量的人力和物力.

因此, 本文采用面向对象技术解决上述问题, 深入分析了元数据标准的对象特性和元数据管理中涉及的对象, 提出了元数据标准和元数据管理的对象模型, 并在东海信息中心的海洋数据共享与交换系统中得以实施和验证.

1 面向对象的元数据标准

1.1 元数据概述

元数据应用存在已久, 早期的地图图例、书本目录到后来的图书馆卡片分类等, 都属于元数据. 但元数据还没有一个统一定义^[2]. 目前, 关于地理空间元数据比较流行的是国际标准化组织 (ISO) 地理信息技术委员会 (ISO/TC 211) 的定义: 元数据是关于数据的数据, 即数据的标识、覆盖范围、质量、空间和时间模式、空间参照系和数据分发等信息^[3-5].

随着网络技术飞速发展, 数据共享需求日益迫切, 元数据已经从最初的数据描述和索引方法发展成为数据表现、数据转换、数据管理和数据使用整个信息传输过程中不可或缺的工具和方法之一^[6-8].

元数据标准是各国际组织和应用部门研究的重心, 在这十多年的发展历程中, 根据不同的应用领

收稿日期: 2009-10-26

基金项目: 国家深部探测专项基金资助项目; 上海市科委重点资助项目 (09DZ1201000)

作者简介: 吴 萍 (1982—), 女, 博士生, 主要研究方向为地球信息科学. E-mail: wuping_sc@126.com

许惠平 (1965—), 男, 教授, 博士生导师, 工学博士, 主要研究方向为 3S 技术和地球物理. E-mail: xuhuiping@tongji.edu.cn

域,提出了不同的元数据标准.表 1 列出了较为常用的元数据标准.

表 1 元数据标准
Tab.1 Metadata standard

组织机构	元数据标准
美国联邦地理数据委员会 (FGDC)	数字地理空间元数据内容标准 (CSDGM)
美国宇航局和全球变化数据管理国际工作组	目录交换格式 (DIF)
国际标准化组织地球信息委员会 (ISO/TC 211)	ISO 地理信息元数据
美国联邦政府	政府信息定位服务 (GILS)
CEN/TC287	CEN(欧洲标准委员会)地理信息—数据描述—元数据
欧洲地图事务组织 (MEGRIN)	GDDB (Geographical Data Description Directory) 数据集描述方法
加拿大通用标准委员会地理信息专业委员会	CGSB (Canadian General Standards Board) 数字地理参考集的目录信息
中国	地理信息元数据

1.2 元数据标准的对象模型

为了有效地组织管理复杂的地理空间元数据,FGDC 在其元数据标准 CSDGM 中提出了元数据子集、复合元数据、元数据元素的元数据层级划分策略.同时,我国国标《地理信息元数据》基于 ISO 19115:2003 提出了与之相似的元数据子集、元数据实体和元数据元素的元数据划分策略^[9],形成了地理空间信息元数据的树状层次模型.

一方面,这种树状层次结构很好地体现了面向对象中的封装、聚合、对象和属性等概念.元数据与元数据子集、元数据子集与元数据实体之间是一种整体与部分的关系,类似于面向对象中的聚合.同时,不同元数据子集以及元数据子集下属的同一层级的不同元数据实体之间具有相对的独立性,类似于面向对象中的封装.元数据子集、元数据实体和元数据元素之间就是面向对象中的对象与属性的关系.

另一方面,地理空间信息元数据树状层次模型还体现了面向对象的泛化概念.元数据随着人们对地理空间信息描述需求的全面和深入而同步发展,该过程是人们对地理空间信息元数据普遍性认识的基础上,对特性逐渐把握的结果,反映了从一般到特殊,从抽象到具体,从粗略到精细的认识演化规律.

因此,地理空间元数据在认识和组织方面具有面向对象中对象抽象、封装、继承、聚合等特性.采用面向对象的通用建模语言 UML 绘制地理空间元数据对象模型,如图 1 所示.

此图清晰地显示了元数据中各实体之间的对象关系.元数据由元数据子集聚合而成,一个元数据标准由一个或多个元数据子集组成;元数据子集是由零到多个元数据实体聚合而成;元数据实体既可继承元数据子集产生,还可聚合或继承其他实体生成.元数据元素作为元地理空间数据的基本单元,抽象为元数据体系中的属性.基于这样的模型,利用面向对象的继承、关联等关系,易于实现元数据标准的扩展和新建等操作.

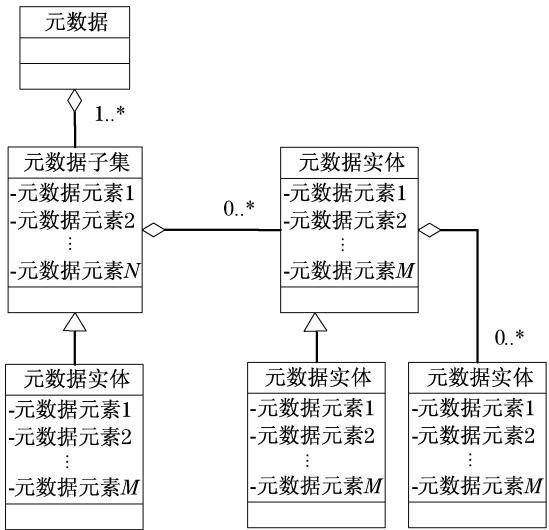


图 1 元数据标准对象关系图
Fig.1 Objects relationship of metadata standard

本文以地理信息元数据国际标准 ISO 19115 作为通用元数据标准,并建立其对象模型,以促进其他专业应用元数据的制定和互操作.

1.3 专业元数据标准的制定方法

专业元数据标准的制定是以通用元数据标准为基础,进行标准扩展、修改,制定满足专业需求的专业元数据.基于元数据标准的对象模型,通过类提取、类继承、类属性的覆盖、类或类属性的删减和类的新建等面向对象操作,实现专业元数据标准的制定.

1.3.1 类的提取

根据所采用的通用元数据标准,将相关的元数据实体和元数据子集抽象为对象类,并建立相互之间的逻辑关系(关联、聚合、继承等),将元数据元素抽象为对象类属性,构建基础元数据类.

目前,多个国际海洋组织均采用 ISO 19115 作为海洋元数据标准开发基础,如 AODC(澳大利亚海洋数据中心)、WMO(世界气象组织)采样的元数据标准,为了增强与现有元数据的互操作和交互性,本

文采用 ISO 19115 元数据标准来通用元数据标准,直接复用了其中的元数据子集类 MD_数据标识信息、MD_数据质量信息、MD_数据发布、MD_坐标系统等。

1.3.2 类的继承

类的继承是以某元数据类为基类,根据所要描述对象的特性,创建一个子类,在继承其父类所有属性的基础上,还具有其特有的属性。

海洋现象的数据获取手段主要有现场观测、数值模拟和遥感等^[10]。因此,在 MD_数据标识信息类的基础上,还需要通过继承而产生 MD_观测、MD_数值模拟和 MD_遥感类,分别描述这三种数据来源的相关平台、仪器以及数值方法的信息。

1.3.3 类属性的覆盖

类属性实际就是元数据元素,类属性的覆盖是指通过自定义类属性的最大出现次数、约束/条件、域,实现对元数据标准原型中元数据类所定义相应特征的修改,以满足特定的应用需求。这主要用于对特定值域的代码表、枚举等值的修改、替换。

如,在海洋数据中,海洋环境数据分为海洋生物、海洋水文、海洋气象、海洋物理、海洋化学、海洋地质地球、海洋遥感和海洋灾害等,因此,在设置 MD_数据标识中的专题属性时,就需要将上述数值添加到该属性的代码表中。

1.3.4 类或类属性的删减

作为国际通用标准,为了兼顾多学科需求,元数据标准具有较为广泛的定义,因此,针对具体应用时,需要对该元数据标准内容进行删减、精简,以求增强元数据的实用性。类和类属性的删减主要包括以下操作:删除无意义的元数据类;删除类中的无用属性项,即元数据元素;删除作为类属性的元数据元素的域值等。

1.3.5 类的新建

即通过面向对象中的关联(包括关联、聚类和组合)而添加的元数据类。

对于海洋现场观测数据来说,多数是通过观测平台和观测仪器进行描述的,因此,在该元数据标准中,为 MD_观测添加了新建的元数据类 MD_观测平台和 MD_观测仪器。

基于这种思想构建的元数据标准,一方面能最大限度地提高元数据标准之间的兼容性;另一方面,以元数据类作为操作单元,便于元数据研发时的分工合作和进度控制,有助于元数据的扩展和维护。

2 面向对象的元数据管理

元数据标准的制定是元数据应用的基础,而元数据管理是元数据功能实现的工具。用户通过元数据管理实现元数据录入、管理、搜索查询及数据挖掘等功能。现有的元数据管理均是基于某一元数据标准而设计,针对不同的元数据标准需要进行重新构建,故造成了较大的资源浪费。而本文的思想是构建元数据管理的对象模型,实现元数据管理构建的代码重用。

2.1 元数据管理概述

元数据管理通常包括三类用户:一般用户(browser)、专业用户(client)和管理员。一般用户有元数据查询、浏览等元数据服务功能;专业用户拥有一般用户功能,还增加了元数据录入、编辑等元数据管理和数据服务功能;管理员主要实现用户、数据库管理功能。基于这样的用户分类和用户职能分配,通用的元数据管理功能如图 2 所示。

2.2 元数据管理的对象模型

元数据管理中涉及的对象主要有数据对象、元数据对象和用户对象,结合上文所介绍的元数据管理功能,采用面向对象设计方法,建立了元数据管理的对象模型,如图 3 所示。

2.2.1 元数据项类

元数据项类由数据集元数据类和目录元数据类聚合组成,数据集元数据是数据集的完整元数据描述,而目录元数据仅是数据集元数据的子集,同时,也代表了元数据目录中目录项所显示的元数据信息。数据集元数据类同时还封装了显示、输出和获取数据的方法。同时,元数据项类还具有一元关联关系。

2.2.2 元数据容器

元数据容器是对所有元数据对象的操作管理。该类设置了元数据个数属性,记录其所含元数据对象的数量。方法主要包括:元数据管理(添加元数据、删除元数据、更新元数据)和元数据查询(关键字查询、字段组合查询和空间范围查询)。查询操作结果是元数据项对象集合。

2.2.3 数据类

数据类主要实现对数据资源的操作管理。数据类属性有数据 ID、数据集名称和数据等级。数据等级属性主要是对不同保密等级的数据设置等级,实现不同等级数据的不同获取操作(在线显示、下载和与

相关人员联系). 方法主要有: 添加数据、删除数据、更新数据、自动获取元数据、格式转换和数据显示.

2.2.4 用户信息类

用户信息是对用户权限和用户信息的管理. 用户一般分为 3 类: 大众用户、专业用户和管理员. 大众用户具有元数据的查询、浏览及数据获取等操作; 专业用户在具有大众用户的操作权限基础上, 还具有数据上传、元数据上传、更新、修改等操作权限; 管理员主要实现对大众用户和专业用户的管理和权限分配. 用户根据不同的权限配置, 实现不同的操作, 这也是系统安全的保障措施之一.

用户信息属性包括有用户名、密码和权限等级. 方法主要包括: 新建用户、删除用户、修改密码和修改用户名. 新建用户和删除用户是管理员所特有的功能.

2.2.5 目录项

目录项类主要实现对元数据目录项的构建和管理. 目录项类属性包括目录 ID、目录等级、目录类、父目录. 根目录的目录等级为 0, 往下依次增加 1. 方法主要有: 是否是叶节点、是否是根节点和获取父节点. 根目录没有父目录, 目录编目的相关字段来自于目录元数据.

2.2.6 元数据编辑器

元数据编辑器类用于根据已定义的元数据模式, 载入相应元数据, 实现元数据编辑, 或根据数据信息获取相关元数据, 提供录入界面供用户进行输入其他元数据项, 实现元数据录入, 最终生成 XML (extensible markup language, 可扩展标记语言) 文档的元数据.

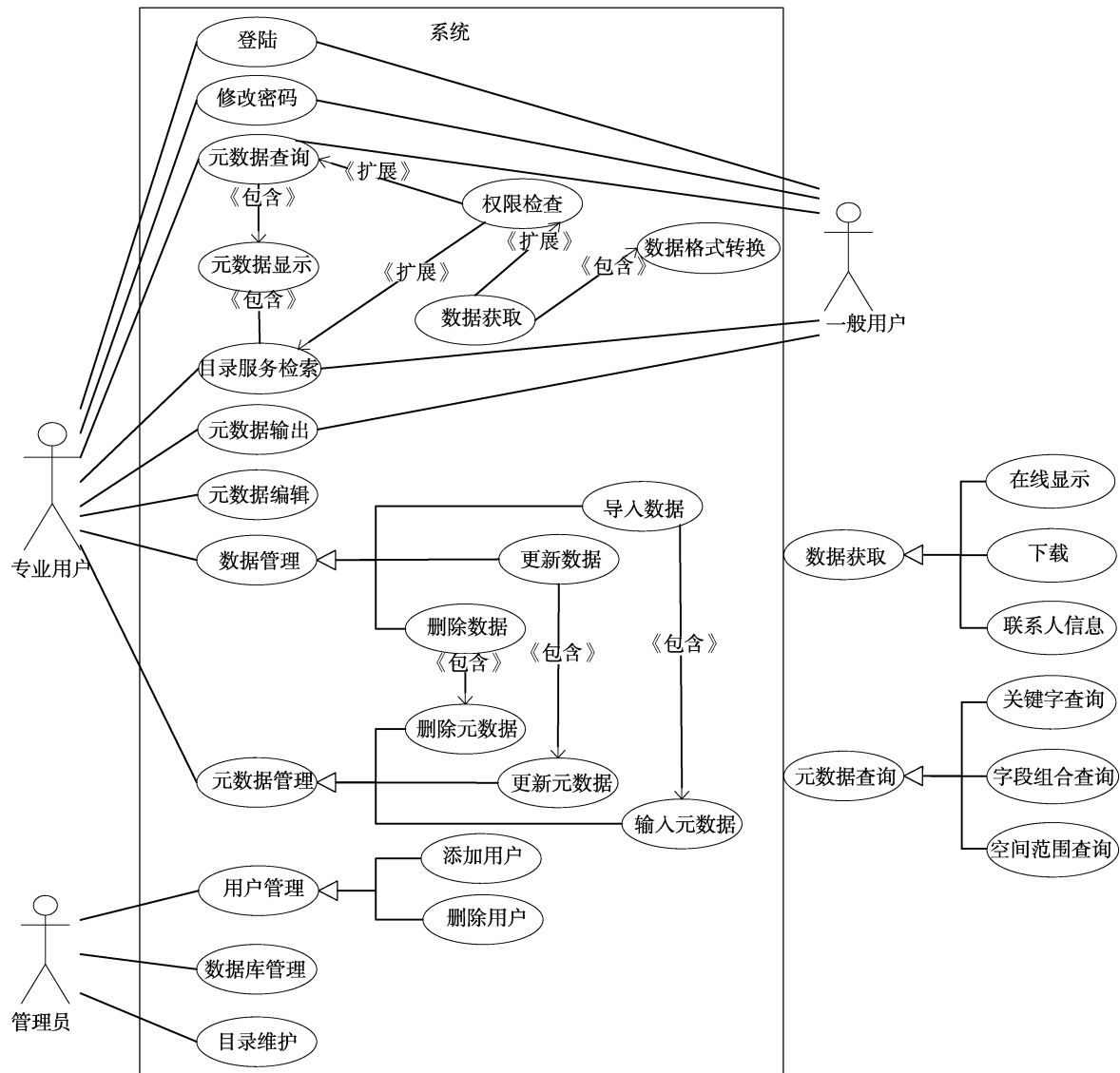


图 2 元数据管理图

Fig.2 Metadata management system

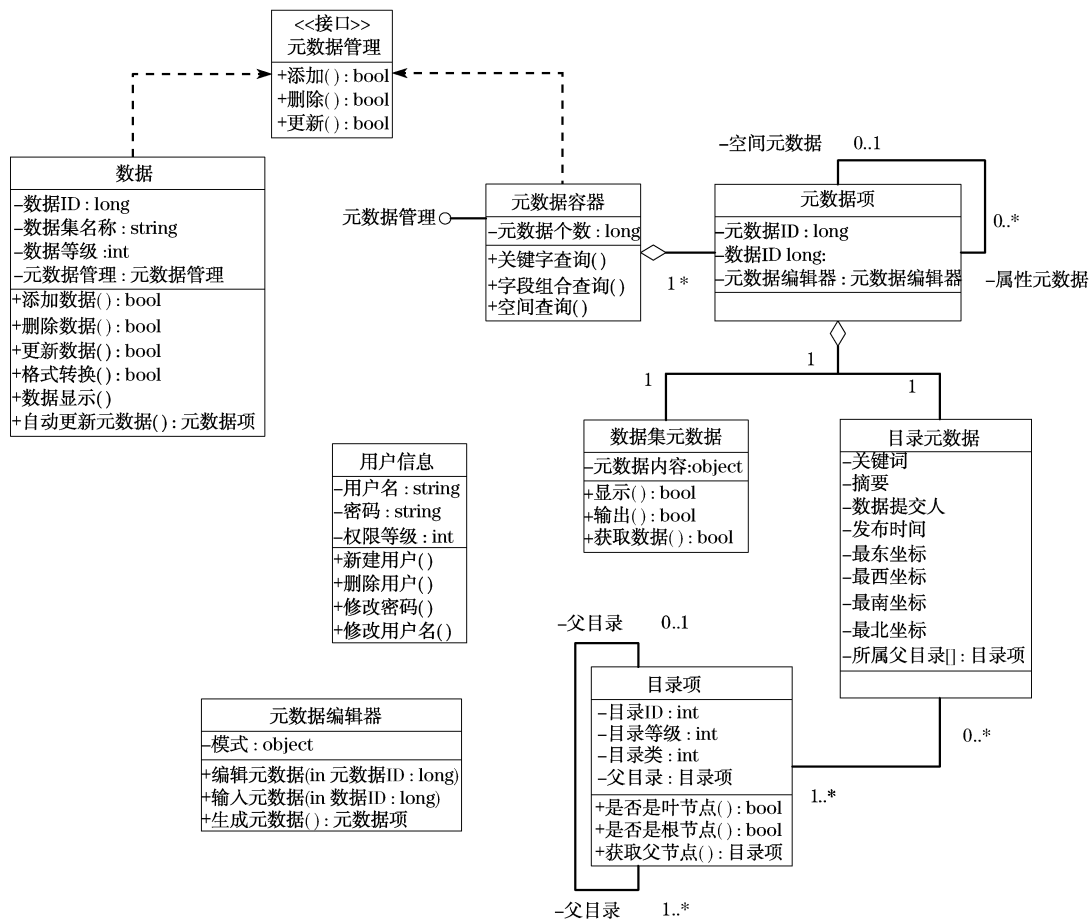


图 3 元数据管理对象图

Fig. 3 Objects of metadata management system

这种基于面向对象思想设计的元数据管理能很好地解决现有元数据管理中存在的问题. 元数据标准和元数据功能分别封装为对象的属性和行为, 通过类的重新定义即可实现代码重用, 大大节省了人力. 这个元数据管理对象模型具有很强的普适性, 可应用于各行各业的元数据管理.

2.3 实现技术

2.3.1 元数据的存储

XML 具有数据和显示分离、简单易读等特点, 已广泛用于网络数据传输和显示. 毫无疑问, 使用 XML 来表达和传输元数据具有很大优势, 且已被大家普遍接受, 目前的关键问题是如何有效存储和管理基于 XML 的元数据以方便用户查询、使用.

由于 XML 数据库还没有成熟, 目前以 XML 描述的元数据存储主要有两种方式: 直接存储为数据库字段和映射为 2 维表格. Oracle 数据库提供了 BLOB(binary large object, 二进制大对象) 字段, 可用于存储整个 XML 文档, 这种方式的优点是存储、读取速度快, 不受元数据模式影响, 但元数据查询和

处理较慢. 将元数据映射为 2 维表格存储在 RDBMS (relational database management system, 关系型数据库管理系统) 中, 数据查询处理速度快, 充分应用了数据库成熟的查询处理技术, 但其缺点是元数据模式更改会直接导致表结构的变化, 而且, 同时读取多个表格也会延缓数据读取速度. 可见, 这两种方式各有优缺点.

同时, 元数据通常被分为两级元数据, 即目录元数据和数据集元数据. 其中, 目录元数据是对整个数据的概括描述, 是数据集元数据的子集, 是整个元数据中经常需要查询使用的元数据要素; 而数据集元数据是关于整个数据的元数据, 多用于元数据的显示.

因此, 本文提出关键字段 + XML 文档的存储方式. 其中, 关键字段为用于数据查询、目录显示的元数据字段信息, 存储在 2 维表中; 而数据集元数据就存储为一个大数据段. 这样, 对元数据的查询实际就是对 2 维表的查询, 提高了查询速度. 这种思想在元数据管理的对象模型中表现为元数据项类由数据集元

数据类和目录元数据类聚合而成,其中,数据集元数据类和目录元数据类的属性就分别封装了 XML 元数据文档和关键字段.

2.3.2 元数据与数据的一体化管理

尽管元数据和数据集之间存在一一对应的关系,但在元数据管理模型中,元数据的管理和数据集的管理却是相互独立的.当数据集发生增、删、改等操作时,元数据缺乏相应的响应和更新机制,会导致元数据信息的准确性下降,直接影响元数据的质量.

依赖倒转原则是面向对象设计原则之一,其基本原则是:要依赖于抽象,不要依赖于具体,更好地实现系统的可扩展性.本文以依赖倒转原则为指导,引入了元数据管理接口,提供了元数据添加、删除和更新 3 个抽象方法,通过类之间的继承和依赖关系,实现数据和元数据的同步更新,符合面向对象的开-闭原则.

元数据和数据同步更新包括数据添加、删除和更新时元数据的相应操作,其实现流程如图 4 所示.

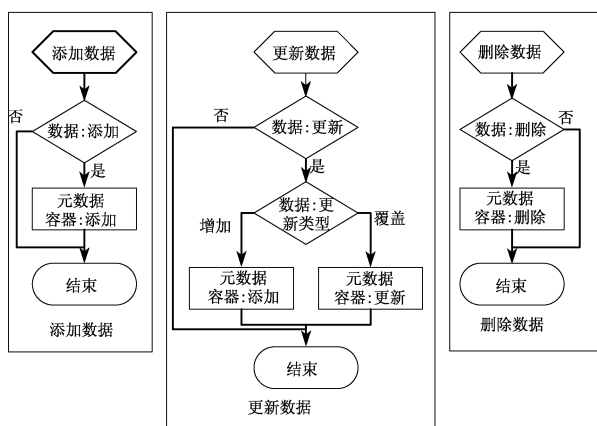


图 4 元数据和数据同步更新

Fig.4 Synchronized updating of metadata and data

2.3.3 数据与数据之间的交叉引用

元数据查询主要有 3 种方式:关键字查询、字段组合查询和空间范围查询.空间范围查询是基于元数据中所提供的空间范围信息进行查询,而在实际应用中,属性数据通常没有空间范围信息描述,因此会出现属性数据遗漏的问题.通常情况下,某些属性数据是对空间数据的其他属性描述,而这些属性数据也是有一定隐含的空间范围的,如何查询这类属性数据也是本文研究中涉及的问题.

面向对象设计中的关联关系可以很好地解决这个问题,本文通过元数据项类的一元关联关系来实现.一个空间数据的元数据对应零到多个属性数据

的元数据,一个属性数据的元数据对应零到一个空间数据的元数据.这样,就能很方便地查询到与某一元数据相关的其他元数据信息.

2.3.4 数据安全

数据安全是数据共享中首要考虑的问题.本文采用用户权限控制和数据安全等级设置来保证数据安全.为用户设置权限等级,决定用户有无获取数据的权限;根据数据安全级别的高低,将数据分为 3 种安全等级:可下载、可在线浏览和与相关联系人联系.数据获取流程如图 5 所示.

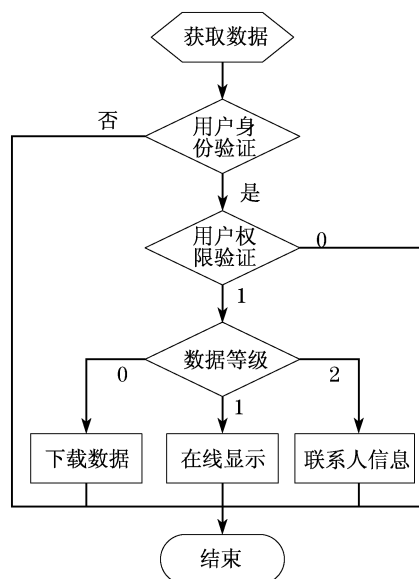


图 5 数据获取流程图

Fig.5 Flow chart of data retrieval

3 应用实例

本文所介绍的对象模型应用于东海信息中心的海洋数据共享与交换系统建设中,其整体思路为:根据本研究区描述数据和对象的特点,基于通用元数据标准制定专业元数据标准,建立元数据标准的对象模型;根据制定的元数据标准和前文的元数据管理对象模型,设计适合该应用的元数据管理对象模型,最终建立元数据管理应用系统.

该系统中所涉及的数据包括两大类:地理空间数据和非空间数据.地理空间数据包括基础地理数据、遥感影像数据等.非空间数据主要包括:海洋环境基础数据、模型参数、系统管理中的用户数据等.图 6 以影像数据为例,说明专业元数据标准的制定.虚线框内的遥感影像类是对通用元数据标准类的继承,是对通用元数据标准的扩展.

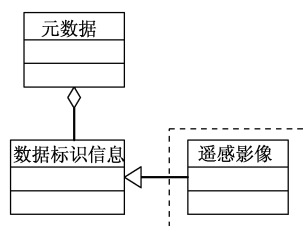


图 6 海洋元数据实例

Fig.6 Example of ocean metadata

该系统采用海洋数据分布式存储,元数据集中式存储的策略,构建了“一个中心,多个节点”的系统平台,主要实现元数据输入、管理、元数据查询、目录查询、数据获取和用户管理等功能。因此,本系统采用 B/S(browser/server,浏览器和服务端)结构的 WebGIS 架构,在服务器端采用前文的元数据管理对象模型,实现该系统的快速构建,并有利于系统的扩展、维护。

4 结语

元数据具有很强的实践性,因此,建立统一的元数据标准是不可能实现的。而如何增强元数据标准之间的互操作性,实现元数据扩展和维护是目前元数据研究的重要方向。本文基于面向对象设计技术,实现了元数据标准和元数据管理的对象化模型设计,很好地解决了元数据标准和管理系统的扩展和维护。同时,在元数据管理对象模型设计中解决了目前多数管理系统在元数据存储、元数据和数据同步更新、元数据查询等方面存在的问题。

地理信息元数据模型是地理信息数据库中对地理信息数据进行定义、管理、操作、扩展等的工具,是整个地理信息数据库的核心。它存储了地理信息元数据的基本操作,通过基本操作及其组合,可实现对地理信息元数据、地理信息数据的所有操作^[11]。ISO 提出了电子文件管理相关的元数据模型^[12]。本文目前仅是对元数据模型的初步抽象和概括,在下一步工作中,将全面地考虑元数据自身特点及其功能,建立完善的元数据模型。

参考文献:

[1] WEI Yaxing, DI Liping, ZHAO Baohua, et al. Transformation of HDF-EOS metadata from the ECS model to ISO 19115 - based XML[J]. Computers & Geosciences, 2007, 33: 238.

[2] 辜寄蓉. 基于元数据的综合数据管理与信息共享[D]. 成都: 成都理工大学地球科学学院, 2003.
GU Jirong. Metadata-based information integrated management and sharing system [D]. Chengdu: Chengdu University of Technology. College of Earth Science, 2003.

[3] 中国国家标准化管理委员会. GB/T 19710—2005 地理信息元数据[S]. 北京: 中国标准出版社, 2005.
GB/T 19710—2005 Geographic information-metadata [S]. Beijing: China Standards Press, 2005.

[4] 崔益乐, 周新忠. 面向对象方法在地理空间信息元数据标准研发中的应用研究[J]. 测绘科学, 2008, 33(1): 147.
CUI Yile, ZHOU Xinzong. Research on application of object-oriented for developing metadata standard for geospatial information [J]. Science of Surveying and Mapping, 2008, 33 (1): 147.

[5] Batcheller J K. Automating geospatial metadata generation——an integrated data management and documentation approach [J]. Computers & Geosciences, 2008, 34: 387.

[6] 冯艳杰, 朱欣焰, 胡海棠. 基于 XML 的地理空间元数据网上发布[J]. 地理空间信息, 2005, 3(4): 16.
FENG Yanjie, ZHU Xinyan, HU Haitang. Geo-spatial metadata distribution on internet based on XML [J]. Geospatial Information, 2005, 3(4): 16.

[7] 陈惠荣, 游雄. 地理空间元数据及其相关技术的探讨[J]. 测绘学院学报, 2003, 20(4): 290.
CHEN Huirong, YOU Xiong. Researches on geo-spatial metadata and related techniques[J]. Journal of Institute of Surveying and Mapping, 2003, 20(4): 290.

[8] 王国复, 涂勇, 王卷乐, 等. 科学数据共享中的元数据技术研究[J]. 中国科技资源导刊, 2008, 40(1): 30.
WANG Guofu, TU Yong, WANG Juanle, et al. Application of metadata technology in scientific data sharing service[J]. China Science & Technology Resources Review, 2008, 40(1): 30.

[9] 蒋景. 中国地理信息元数据标准研究[M]. 北京: 科学出版社, 1999.
JIANG Jing. Researches on metadata standards of Chinese geographic information[M]. Beijing: Science Press, 1999.

[10] 侍茂崇, 高郭平, 鲍献文. 海洋调查方法[M]. 青岛: 青岛海洋大学出版社, 2000.
SHI Maochong, GAO Guoping, BAO Xianwen. Method of oceanographic investigation [M]. Qingdao: Qingdao Ocean University Press, 2000.

[11] 梁大圣, 刘纪平, 梁勇, 等. 基于存储结构的三元组地理信息元数据模型[J]. 测绘通报, 2009, 2: 31.
LIANG Dasheng, LIU Jiping, LIANG Yong, et al. Storage structure based geographic metadata model described by three data element groups[J]. Bulletin of Surveying and Mapping, 2009, 2: 31.

[12] 程妍妍. 国际标准 ISO 23081 元数据模型分析[J]. 现代图书情报技术, 2008, 9: 31.
CHEN Yanyan. ISO 23081 metadata model analyses[J]. New Technology of Library And Information Service, 2008, 9: 31.