

文章编号: 0254 - 5357(2012)02 - 0370 - 06

## 基于 XML 的岩芯信息元数据标准研究与实践

蒋夕平<sup>1</sup>, 吴凤凰<sup>1</sup>, 吴芳<sup>1</sup>, 修连存<sup>2</sup>

(1. 南京农业大学, 江苏 南京 210095; 2. 南京地质矿产研究所, 江苏 南京 210016)

**摘要:** 目前岩芯信息表述尚没有统一的格式, 严重阻碍了岩芯信息资源的利用, 造成了资源浪费。为提高岩芯信息利用率, 方便用户和其他软件共享岩芯信息, 必须规范其组织、存储和交换数据格式, 并提供相应的技术实现手段通过软件自动保存、解析岩芯信息。文章参照国内外相关资料及技术标准, 创建了岩芯信息元数据标准, 应用 XML 技术生成与元数据标准对应的 XSD 文件, 并在应用程序中成功使用, 当用户交互岩芯信息时, 生成符合 XSD 格式要求的 XML 数据文件传送信息, 不同用户可以正确解析岩芯信息。通过应用岩芯信息元数据标准, 规范了数字化岩芯信息组织、存储和交换, 采用 XML 技术由软件自动实现, 大大提高了用户工作效率, 方便了岩芯信息的共享, 充分发挥了岩芯信息的价值。

**关键词:** 岩芯信息; 元数据标准; XML Schema

中图分类号: P62; G202; TP311.132 文献标识码: B

## Metadata Standardization and Application for Core Information Based on XML

JIANG Xi-ping<sup>1</sup>, WU Feng-huang<sup>1</sup>, WU Fang<sup>1</sup>, XIU Lian-cun<sup>2</sup>

(1. College of Science, Nanjing Agricultural University, Nanjing 210095, China;  
2. Nanjing Institute of Geology and Mineral Resource, Nanjing 210016, China)

**Abstract:** Due to the lack of a unified presentation format, the resource for core information is clearly under-utilized at present. To improve the utilization of core information, as well as facilitating the sharing of information among users and other softwares, the format should be standardized for organization, storage and data exchange. Relative technologies are being developed for auto-storage, interpretation and analysis of core information via software. Referring to the relative reference materials and technique standards worldwide, the authors created the standards of metadata for core information. XML technology can be used to create the XSD regarding to respective metadata standards, which has been implemented successfully in application programs. Using the XML data file generated by corresponding XSD, different users can interpret the core information correctly during interaction. With metadata standardization for core information, the organization, storage and data exchange format can be normalized automatically via software using XML technology. Results show an improvement in the efficiency of users, facilitated information by sharing, and enhanced the utilization of core information.

**Key words:** core information; metadata standardization; XML Schema

收稿日期: 2011 - 05 - 19; 接受日期: 2011 - 10 - 11

基金项目: 中国地质大调查项目(1212010916030)

作者简介: 蒋夕平, 硕士, 讲师, 主要从事计算机仿真与数据建模研究。E-mail: jiangxp@njau.edu.cn。

岩芯是地质找矿和地质研究工作中关键、直观、常用的实物,它的作用是各种间接的探测方法所不能替代的。通过数字化岩芯扫描仪扫描岩芯实物,获取岩芯图像和光谱数据,利用光谱学原理由矿物的光谱特性研究矿物晶体结构和成分,采用数据处理软件对光谱数据进行处理分析,得到岩芯矿物组分和其他数据处理结果,对矿物、岩石光谱特性的系统研究奠定了当今高光谱遥感技术发展的基础<sup>[1-3]</sup>。岩芯实物及相关资料在保存、查询和调阅过程中存在诸多不便,为更方便地利用岩芯资源,必须将岩芯实物及相关资料数字化,并将数字化后的岩芯图像、光谱、矿物成分等相关信息统一存储,这样可以极大地丰富岩芯信息内涵,使其更具科学价值,同时也方便了岩芯信息的利用,国内多家科研单位研究开发了岩芯图像管理系统<sup>[4-7]</sup>,但是这些系统对岩芯信息描述存在较大差异,国外也有类似研究。如澳大利亚联邦科工组织(CSIRO)建立了矿产资源信息查询网站(<http://nvcl.csiro.au>),为社会相关部门提供服务。

随着信息技术的快速发展和社会信息化建设的不断深入,岩芯信息资源的数量及其复杂程度的同步增长使信息搜集和检索技术越来越难以满足大部分用户的信息需求,由于岩芯信息产生的多目的性和无序性,使岩芯信息资源在空间分布上显现出复杂的格局。为了存储、管理和共享岩芯信息,各系统和部门大都根据自身需要创建了岩芯信息管理系统,但这些系统未能从全局方位分析岩芯信息管理系统的需求关系,也没有从全社会共享的思维设计开发,从而阻碍了岩芯数据的应用和共享,使许多珍贵的岩芯数据资源难以有效地用于解决地质、矿产、环境及农业等领域的问题。由于标准难以统一、数据异构等原因,给信息资源的检索、共享、利用带来各种困难。只有通过标准化、格式化的数据模型描述数据才能确保数据交换和数据重复的最小化,因此研究用于数据搜索、获取和分析的标准化元数据就显得非常重要<sup>[9]</sup>。结构化的元数据不仅能够描述数据信息,还可以规范信息的数据结构,作为岩芯数据生产者,急切需要一套有效的数据管理和维护方法,而岩芯数据用户同样也要求能够从生产者获得快捷、安全、有效、全面的服务,以便从海量的岩芯资源中快速、准确地发现、访问、获取和使用所需的数据<sup>[10]</sup>。甘露等<sup>[8]</sup>对地质矿产分析测试元数据标准进行了有益的探索。目前,元数据已经从简单的数据描述或索引,发展成为用于管理数据、共享数

据、发现数据和使用数据的一种重要工具<sup>[11-12]</sup>,因此通过将岩芯资料数字化,并且对岩芯信息特征进行概括和抽取,创建岩芯信息元数据标准,再采用人和机器都可理解的技术处理岩芯信息元数据标准,就可以方便地开发应用程序,实现软件对岩芯信息的存储、处理和交换,并最终向用户展示岩芯信息。

本文参照国内外相关资料及技术标准,通过创建岩芯信息元数据标准,生成符合XSD格式要求的XML数据文件传送信息,规范数字化岩芯信息的组织、存储与交换,方便不同用户正确解析岩芯信息,从而充分发挥岩芯信息的价值。

## 1 岩芯信息元数据标准创建

### 1.1 岩芯信息数据特点

数字化的岩芯信息,除了具有地学数据的共性外,还具有其自身的特点,参照曹代勇等对煤地质学元数据标准的研究<sup>[13]</sup>,岩芯信息元数据标准须体现岩芯信息的以下特点。

(1)岩芯信息的内容具有多样性。岩芯信息内容具有广泛性、综合性的特点,可以划分为空间数据、属性数据和知识数据三大类。空间数据除了基础空间信息之外,还包括矿区分布图、钻孔分布图等;属性数据包括岩芯的各种特征信息、岩芯图像、岩芯光谱等;知识数据包括岩芯矿物成分、化学组成等。

(2)岩芯信息的来源具有多样性。岩芯特征信息来源于地质大调查,岩芯图像来源于岩芯扫描仪,岩芯光谱来源于岩芯高光谱测量仪,岩芯知识数据来源分析软件,不同来源的岩芯信息具有较大差异性。

(3)岩芯信息的数据类型具有多样性。岩芯信息可分为非空间数据类型和空间数据类型。非空间数据类型又有文本型(如地质勘查报告等)、数值型(如光谱数据)和图像文件等,空间数据类型则包括矢量地图等多种数据形式。

(4)岩芯信息的表现形式具有多样性。岩芯信息表现时,应根据不同数据类型的需求,选择适当的表现形式。例如用虚拟现实的手法表现地质空间数据,用表格表现组分信息,用曲线表现光谱,用图像显示岩芯外貌。

岩芯信息的以上特点,使得创建岩芯信息元数据标准比较复杂,科学、实用的岩芯信息元数据标准,既要符合国际通用元数据标准的建设规范,又必须实现岩芯信息的个性化要求。

### 1.2 岩芯信息元数据标准模型

元数据一般为树形结构,从层次上可分为元数

据子集、元数据实体和元数据元素。元数据元素是元数据基本的信息单元,不可再分割;元数据实体是同类信息资源元数据元素的集合;元数据子集则是相互关联的信息资源元数据实体和元数据元素的集合<sup>[14-15]</sup>。图1为元数据层次结构图。

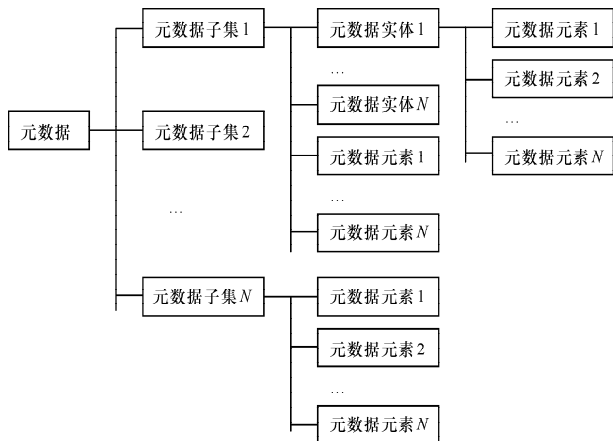


图1 元数据层次结构

Fig. 1 Hierarchical structure of metadata

针对岩芯信息数据特征,参照《国土资源信息核心元数据标准》<sup>[16]</sup>和中国地质调查局地质调查技术标准《地质信息元数据标准》<sup>[17]</sup>,制订了岩芯信息元数据标准,它由5个元数据子集构成,其中元数据信息、标识信息、数据质量信息、内容信息是必选元数据子集,分发信息是可选元数据子集。岩芯信息的具体数据包含在内容信息子集中。图2为岩芯信息元数据标准 UML 模型图,从图2可以看到,岩芯通过岩芯图像展示,岩芯图像属于某个矿区的某个钻孔,某个钻孔的岩芯图像又按照岩芯盘、岩芯格组织,不同深度的岩芯有相应的岩芯光谱、化学组分、矿物成分信息。

### 1.3 岩芯信息元数据标准字典

按照图2岩芯信息元数据标准 UML 模型图,创建岩芯信息元数据标准字典,字典详细说明岩芯信息元数据标准中每个元素、实体和子集的属性,即定义它们的名称、描述、约束条件、出现次数、数据类型、值域<sup>[11]</sup>等。岩芯信息元数据标准字典给出了岩芯信息元数据标准的详细信息,根据岩芯信息元数据标准 UML 模型图和岩芯信息元数据标准字典,可以创建岩芯信息元数据标准 XML Schema (eXtensivie Markup Language Schema, 扩展标记语言架构),也可以创建岩芯信息管理系统的数据库。

## 2 岩芯信息元数据标准实现技术

### 2.1 XML 技术应用

创建岩芯信息元数据标准,能够为岩芯数据的采集者提供数据组织、存储的数据格式,并提供岩芯信息质量控制信息,可以满足岩芯信息用户对岩芯信息的浏览、检索和查询的需要,并为岩芯信息用户提供不同岩芯信息数据格式转换平台。要真正发挥岩芯信息元数据标准的作用,必须寻求一种先进、科学的实现技术,XML 是一种简单的数据存储语言,使用一系列简单的标记描述数据,具有结构严谨、计算机易“理解”且被广泛采用的优点,已经成为描述元数据标准的首选模式语言<sup>[18-19]</sup>。国际上有一定影响力的元数据标准(如 Dublin Core、EML 和 FGDC 等),几乎都是采用 XML 作为其模式语言。

### 2.2 岩芯信息元数据标准模式(XSD)

根据岩芯信息元数据标准 UML 模型图和岩芯信息元数据标准字典,生成岩芯信息元数据标准 XSD, XSD 通过元素声明来规定 XML 文档使用的元素名称、元素内容及元素的数据类型等属性,每个 XSD 的根元素都是 `xsd:schema`,其他元素声明必须包含在这个根元素内部。

### 2.3 岩芯信息数据文件

依据岩芯信息元数据标准 XSD,可以生成包含岩芯信息的 XML 数据文件,无论是岩芯数据生产者还是岩芯数据用户,使用和交换岩芯信息时,都必须将岩芯信息组织成符合岩芯信息元数据标准 XSD 的 XML 数据文件,这样不同的岩芯信息用户都可以正确理解岩芯数据的意义,其他软件也能够正确解析 XML 数据文件包含的岩芯信息。

## 3 岩芯信息元数据标准应用

数字化的岩芯信息保存在数据库或其他临时数据文件中,当用户进行岩芯信息采集、存储、查询和交换时,需要按照岩芯信息元数据标准 XSD 的要求生成中间 XML 数据文件,同时应用程序数据库表结构设计也须符合岩芯信息元数据标准模型。

### 3.1 岩芯数据采集与使用

岩芯信息的内容具有多样性,各种数据之间存在着内在联系,组织、存储岩芯信息需要遵循岩芯信息元数据标准,只有这样应用程序才能准确、快速处理这些数据,不同的岩芯信息使用者才能无歧义地解析这些数据。例如,岩芯数据采集者测量了某段岩芯的光谱数据,在保存岩芯光谱数据时,必须将这些光谱数据组织成符合岩芯信息元数据标准 XSD

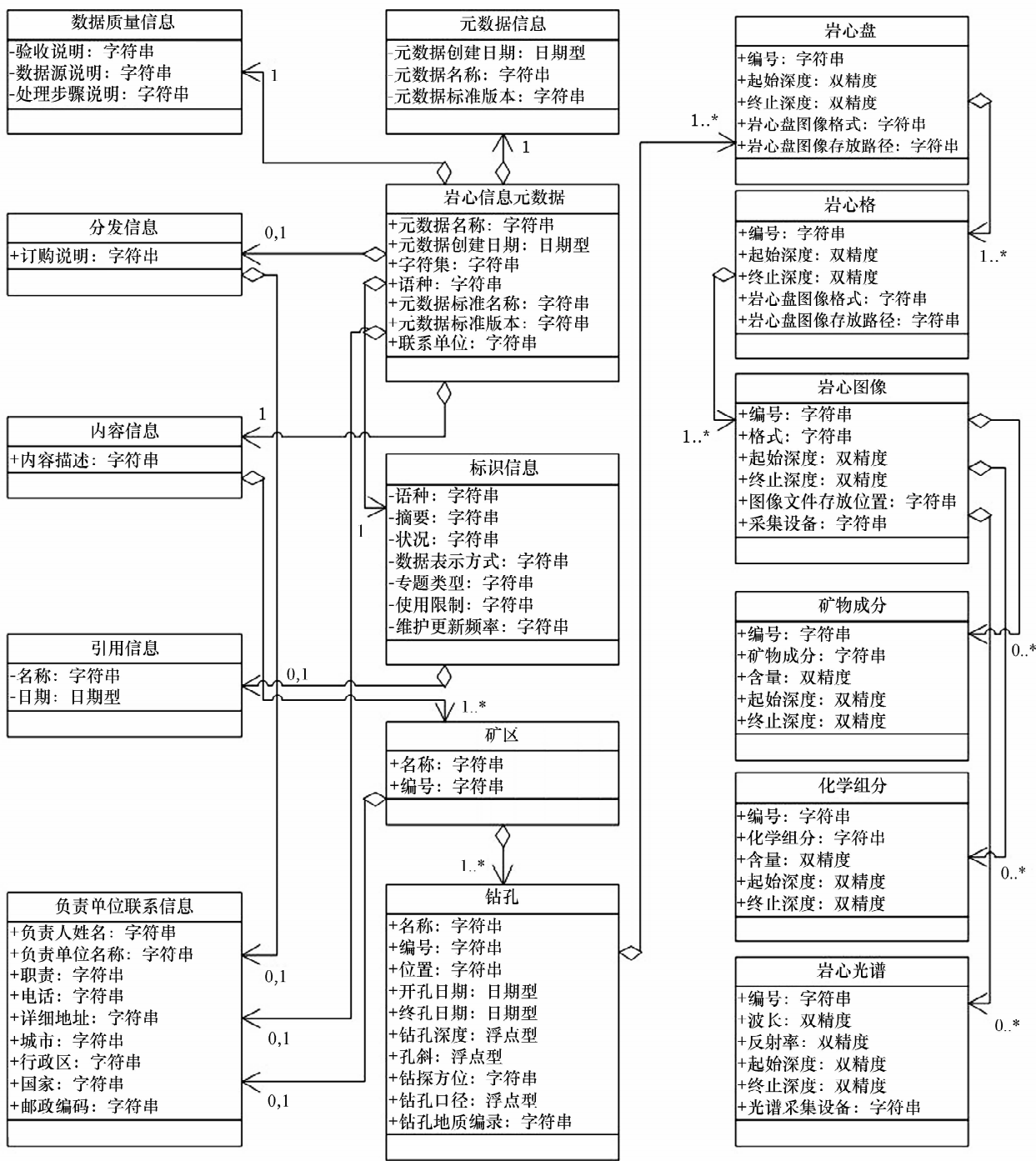


图 2 岩芯信息元数据标准 UML 模型图

Fig. 2 UML model for the standard of metadata for core information

的 XML 数据文件,这些 XML 数据文件可以作为岩芯数据的临时存储载体。当需要将岩芯光谱数据归档时,岩芯数据采集者只要提交这些 XML 数据文件,应用程序将根据岩芯信息元数据标准 XSD 自动解析 XML 数据文件中的岩芯光谱数据并存储到数据库中;当研究人员之间需要交换岩芯数据时,首先查询到需要的岩芯数据,再生成符合岩芯信息元数

据标准 XSD 的 XML 数据文件并发送给使用者,使用者收到该 XML 数据文件后,再根据岩芯信息元数据标准 XSD,准确解读其中的岩芯信息,这一切都可以借助于应用程序完成。

### 3.2 数据库设计

设计岩芯信息管理系统数据库时,必须为岩芯信息元数据标准中的每个复合元素创建相应的数据

表,数据表的字段就是复合元素所包含的简单元素;如果复合元素中引用了其他复合元素,那么需要为被引用的复合元素创建数据表,同时根据两个复合元素之间的关系创建两张数据表之间的关联。

例如岩芯格是一个复合元素,它包含了简单元素如起始深度、终止深度、岩芯格图像存放路径等,这些简单元素都是岩芯格数据表的字段,同时岩芯格还包含了岩芯图像这一复合元素,必须为岩芯图像再创建一张数据表,岩芯图像表是岩芯格表的子表,两表通过内外键关联。

### 3.3 应用程序架构

图3为全国岩芯资源信息查询系统软件架构示意图,软件基于.NET平台,采用了面向对象的编程技术<sup>[20]</sup>。图4为全国岩芯资源信息查询系统软件界面。当用户访问岩芯信息时,首先通过应用程序向Web服务器发出请求,Web服务器将用户访问参数发送给应用服务器,应用服务器分解接收的参数并分别传送给数据库访问模块和岩芯图像访问模块,这两个功能模块分别访问数据库中的岩芯信息和岩芯图像服务器上的岩芯图像;当应用服务器接收到返回的岩芯信息和岩芯图像后,将这些信息回传给Web服务器,Web服务器再将信息发送到客户端,客户端、Web服务器和应用服务器之间传递的中间数据均需要满

足岩芯信息元数据标准 XSD 对数据格式的要求。另外,应用程序通过 Web Service 向其他软件开放岩芯信息访问接口,其他软件提供岩芯信息访问参数,应用程序返回访问结果,两个软件之间交换的 XML 数据文件,也必须符合岩芯信息元数据标准 XSD 的要求,这样大大方便了岩芯信息交流,提高了岩芯信息的共享利用率,充分发挥岩芯信息的价值。

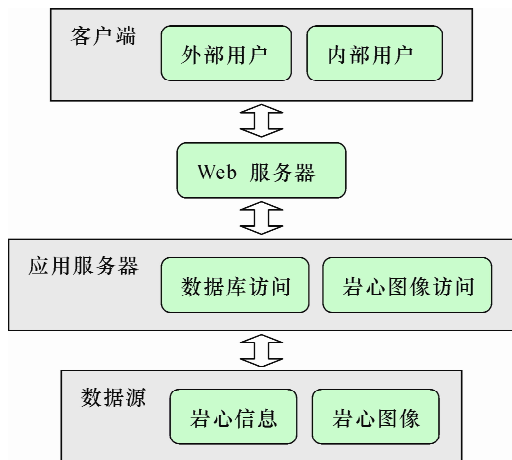


图3 全国岩芯资源信息查询系统软件架构

Fig. 3 Software architecture of inqurement for national core information

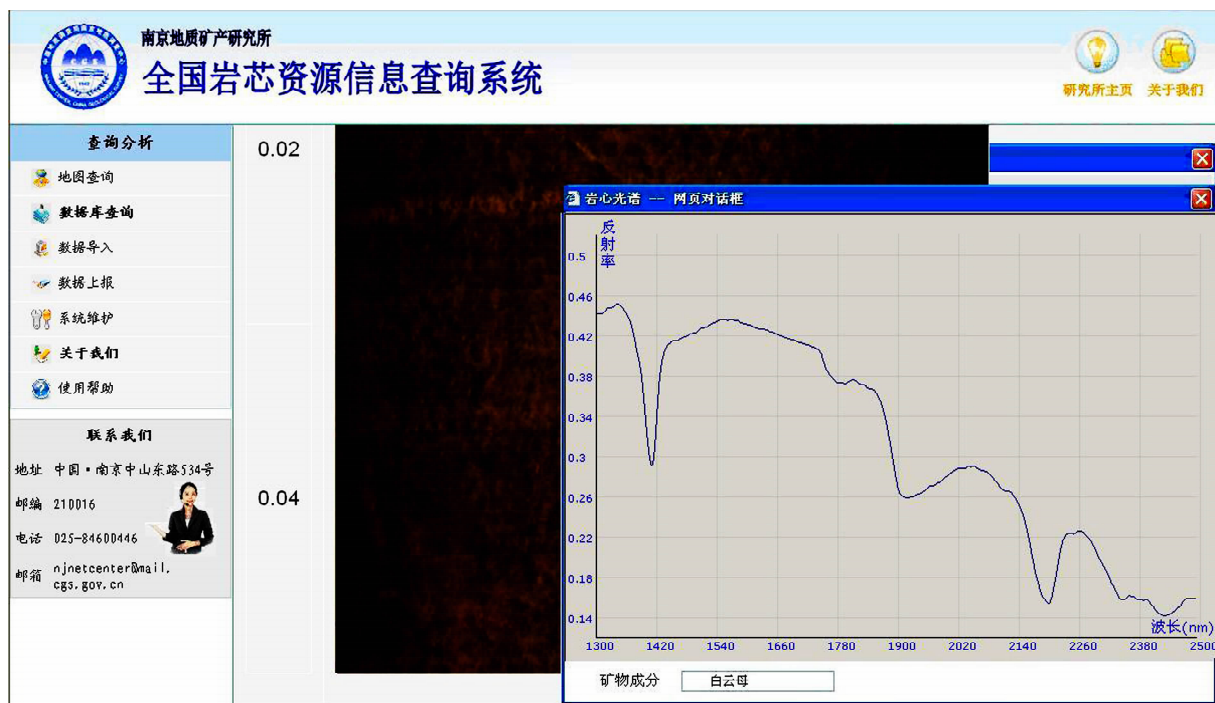


图4 全国岩芯资源信息查询系统软件界面

Fig. 4 Interface of inqurement for national core information

## 4 结语

根据岩芯信息数据特点,参照相关元数据标准,应用XML技术创建规范、科学的岩芯信息元数据标准XSD,并将它应用于岩芯信息查询软件中,用于规范岩芯信息的组织、存储和交换,大大方便了岩芯信息的共享,大大提升了岩芯信息的使用价值,对岩芯信息数据标准化具有一定的参考价值。

## 5 参考文献

- [1] 燕守勋. 矿物与岩石的可见-近红外光谱特性综述[J]. 遥感技术与应用, 2003, 18(4): 191-201.
- [2] 胥燕辉, 田庆久. 磁铁石英岩型铁矿岩芯光谱编录研究[J]. 国土资源遥感, 2005(3): 70-73.
- [3] 郑志忠, 陈春霞, 修连存. 便携式野外现场近红外地物光谱仪研究与测试[J]. 现代科学仪器, 2008(2): 25-28.
- [4] 阮玉媛. 宏观岩芯图像采集仪、图像库与分析系统[J]. 录井技术, 2002, 13(3): 61-62.
- [5] 王亚, 吴晓红, 罗代升, 何小海. 岩芯图像数据库管理系统的规划设计与实现[J]. 电脑知识与技术, 2004(20): 29-33.
- [6] 郑胜利. 岩芯图像数据库应用软件的设计[J]. 石油仪器, 2006(6): 72-73.
- [7] 韩志强, 吴晓红, 何小海. 基于Web的岩芯图文信息系统设计与实现[J]. 电脑知识与技术, 2009, 5(22): 6201-6203.
- [8] 甘露, 吴晓军, 吴淑琪, 邓赛文. 地质矿产分析测试元数据标准研究[J]. 现代科学仪器, 2006(2): 64-66.
- [9] Liu S P, Yang Y, Xie G T. Supporting ontology-based dynamic property and classification in websphere metadata server[J]. *Lecture Notes in Computer Science*, 2008, 5318(1): 861-862.
- [10] Bermudez L, Piasecki M. Metadata community profiles for the semantic web[J]. *Geoinformatica*, 2006, 10(2): 159-176.
- [11] 曹彦荣, 吴洪桥, 毕建涛, 黄裕霞, 何建邦. 国家资源与环境数据库元数据管理研究[J]. 地球信息科学, 2002(2): 6-10.
- [12] 赵改善, 曹邦功. 元数据: 勘探开发数据管理的一种新工具[J]. 石油物探, 2002, 41(2): 236-242.
- [13] 曹代勇, 郝伟. 煤地质学元数据标准化研究[J]. 煤炭学报, 2008, 33(3): 285-288.
- [14] 李振富, 张俊星, 李毅. 分布式信息资源元数据标准研究[J]. 现代电子技术, 2010(11): 147-150.
- [15] 杨德婷, 阎保平. 科学数据库元数据标准体系设计[J]. 微电子学与计算机, 2003(7): 1-4.
- [16] TD/T 1016—2003, 国土资源信息核心元数据标准[S].
- [17] 中国地质调查局. DD 2006-05, 地质信息元数据标准[S].
- [18] 刘飞, 黎建辉, 阎保平. XML和RDF在科学数据库元数据标准建设中的应用[J]. 微电子学与计算机, 2004, 21(7): 127-131.
- [19] 周骋, 刘修国, 张剑波. 基于XML的元数据技术在GIS中的应用[J]. 地理空间信息, 2006, 4(2): 28-30.
- [20] Nagel C, Evjen B, Glynn J. C# 2005 & .NET 3.0 高级编程[M]. 5版. 北京: 清华大学出版社, 2007.