

面向网格服务流程的空间数据共享标准研究

曾 怡¹, 李国庆²

(1. 北京林业大学信息学院, 北京 100083; 2. 中国科学院对地观测与数字地球科学中心, 北京 100086)

摘要:研究网格环境中空间信息的表达规范和数据资源描述机制,是建立空间数据共享与协同服务的第一步,也是数据资源整合和互操作的前提。目前多数研究集中在相关标准的内容组成和具体实施方法上,缺乏针对数据服务具体环节和协同的具体环节处理过程的标准映射关系研究。通过面向网格环境中的数据服务体系结构,基于操作流程和处理环节对相关标准进行分析,总结空间资源描述环节和相应标准规范的映射关系,研究建立分布异构空间数据共享的资源描述机制和元数据策略。以空间信息网格平台的应用实验为基础,详细介绍了空间数据资源网格化描述模型。

关键词:空间数据标准;数据网格;数据服务

中图分类号:TP 751 **文献标志码:**A **文章编号:**1004-0323(2011)05-0698-07

引用格式:Zeng Yi, Li Guoqing. Research of Service-oriented Data Sharing Standard in Grid Environment[J]. Remote Sensing Technology and Application, 2011, 26(5): 698-704. [曾怡, 李国庆. 面向网格服务流程的空间数据共享标准研究[J]. 遥感技术与应用, 2011, 26(5): 698-704.]

1 引言

在多源异构空间信息查询检索和管理协同的各个环节,常常伴随着复杂的空间数据描述。无论是提交的数据请求、返回结果,或是数据操作进程中的中间状态参数,都需要考虑如何赋予数据完整、准确的说明。虽然数据标准封装了数据的差异,但是数据结构本身的属性还是要被完整地保存下来,例如卫星参数、成像参数、参考系统、数据类型等要素。因此,研究网格环境中空间信息的表达规范和数据资源的描述机制,是建立空间数据服务共享和协同服务的第一步,也是数据资源整合和互操作的前提^[1]。只有遵从数据资源描述标准化体系的空间数据服务系统,才能具备良好的可扩展能力。

目前,大多数对于空间数据描述规范的研究都集中在具体标准的内容组成和实施方式上,并没有针对网格数据服务中查询检索和管理协同的具体环节和处理过程,研究这些标准的映射关系。本文结合网格中空间数据服务的具体需求,针对数据服务具体流程和操作环节的特点,在充分参考各种主流空间元数据国际标准的基础上,对其进行必要的摘

录、补充和改进,建立标准化和通用性的数据资源网格化描述模型。

本文从空间信息元数据规范和语义关系特性两方面,研究空间数据描述的表达规范和相关标准,探讨构建多层次的空间数据资源描述机制和标准体系。针对具体的操作流程和处理环节,面向网格环境中的数据服务结构体系,研究建立支持分布异构数据共享的空间资源描述标准和元数据策略,详细介绍了空间数据资源网格化描述模型,并在空间信息网格(Spatial Information Grid, SIG)试验床上进行了示范验证。

2 空间元数据标准规范分析与比较

空间数据资源的描述机制与两个因素密切相关:空间元数据规范和空间数据语义关系特性。元数据策略是空间信息查询和交互的基础,而语义分析是智能检索和精确匹配的关键,两者都关系到空间数据服务的构建和适应能力。

2.1 空间数据核心元数据标准

地理空间元数据是关于数据、操作数据的进程以及应用程序结构和意义的描述信息。随着信息技

术的发展,元数据已经由一种数据描述与索引方法扩展到包括数据发现、数据转换、数据管理和数据使用的整个信息处理流程中不可缺少的强有力工具和方法^[2]。元数据是提高地理空间信息利用率的重要

条件,同时与空间数据服务的质量和适应能力密切相关。国际上参与空间元数据标准制定的机构很多,相关研究也层出不穷地发展^[3]。表 1 列出一些常见的空间元数据基本标准。

表 1 空间数据元数据相关标准
Table 1 Main standards of spatial metadata

标准简称	详细名称	制定单位
CSDGM	地理空间数据集元数据内容标准(Content Standard for Digital Geospatial Metadata,CSDGM)	美国联邦地球数据空间数据委员会(FGDC)
GDDD	GDDD 数据集描述方法	欧洲地图事务组织(MEGRIN)
GCMD&-DIF	全球变化主目录(Global Change Master Directory,GCMD)和地球科学数据集目录交换格式(Directory Interchange Format,DIF)	美国航空航天局(NASA)
CGSB	地球空间数据集描述(Canadian General Standards Board,CGSB)	加拿大标准委员会(CSC)
DCMS	都柏林核心元数据标准(Dublin Core Metadata Set,DCMS)	美国俄亥俄州都柏林董事会
ANZLIC Metadata Guidelines	澳大利亚新西兰土地与地理数据核心元数据(Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand)	澳大利亚新西兰土地信息委员会(ANZLIC)
ISO/TC211	ISO/TC211 元数据相关标准	国际标准化组织地理信息/地球信息技术委员会
OGC® Standards and Specifications	开放地理数据互操作规范(OpenGIS Consortium,OGC)	开放地理信息系统协会
NFGIS 元数据标准	国家基础地理信息系统(National Foundation Geographic Information,NFGIS)元数据标准	国家基础地理信息中心
SDBCM	中国科学院科学数据库核心元数据标准(Scientific DataBase Core Metadata,SDBCM)	中国科学院
NSII 元数据标准	国家空间信息基础设施(National Spatial Information Infrastructure,NSII)元数据标准	国家信息中心

目前,比较常用和有影响力的空间元数据标准,是美国联邦地球数据空间数据委员会 FGDC 制定的 CSDGM(Content Standard for Digital Geospatial Metadata),以及在此基础上由国际标准化组织制定的 ISO/TC211。它们对空间数据的各方面内容都做了详细的规定和描述。如果所有的空间数据源都遵照上述标准进行定义,那么对空间数据的检索和交换将是一件非常简单的事情。

然而遗憾的是,由于分布在网格中的空间元数据通常由不同的数据组织所有,其元数据内容和组织形式由于种种原因难以达到一致。同时由于空间数据自身的复杂性,造成不同应用领域的空间数据难以制定统一的元数据标准。因此,在网格环境中,实现异构元数据标准的多源分布式空间数据共享,就必须建立合理的空间信息元数据机制和交换策略。

我们所做的研究工作主要在 CSDGM 和 ISO/TC211 这两个标准上开展。在 SIG 数据服务节点

的数据查询、获取和交换处理中,有多个环节涉及到 ISO19115 标准,因此,它是 SIG 空间数据网格化描述机制最重要的参考标准之一。

2.2 空间数据查询标准分析

查询标准是数据网格资源描述规范的关键。关于空间数据的查询标准,比较受到认可的是 Z39.50 协议 OGC 组织制定 OGC Filter 规范。

Z39.50 协议是网络信息检索的标准协议,最初的规格由 ANSI(美国规格协会)制定。它具有丰富的信息资源共享接口规范,良好的互操作性和严谨的体系结构^[4]。它以标准化查询句法以及标准化格式返回查询结果,包括客户端和服务端软件,通过建立客户端和服务端连接,传送格式化查询,返回查询结果。

相对而言,OGC Filter 更适合于定义空间数据查询条件描述^[5]。OGC Filter 编码执行规范于 2005 年推出,OGC 制定它的目的是将其作为系统

中的查询谓词,用来描述 OGC CQL(Common Catalogue Query Language)的 XML 编码。OGC Filter 规范定义了表示 OGC 某些行为和操作的结构描述,在该描述中,能够根据地理空间数据的属性值识别 OGC 实体对象或某一个实例的子集。当前许多基于 WEB 的 OGC 规范都将采用 XML 格式的 Filter 编码作为地理数据表达标准,尤其是 WFS、Gazetteer 和 WRS(Web Registry Service)服务。使用 XML 工具可以容易地将基于 OGC Filter 规范的数据源访问语言进行验证、解析或是转换成 SQL 查询语句。例如,可以转换成 SQL 中 SELECT 句式的 WHERE 语句,来获取存储在基于 SQL 关系数据库中的数据。类似地,它也可以转换成 XPath 或 XPointer 来获取 XML 文档中的数据。

对于数据服务中查询标准设计的关键,是必须具备松散耦合的元素定义,使得它所能支持和表达的条件信息描述,能够充分适应各类异构数据源,执行查询条件的匹配;而操作符关系的定义,则必须尽可能多地满足数据间的比较和模式匹配。对比上述两个标准,由于 OGC Filter 规范更适合描述空间数据的请求信息,结合网格环境中的数据服务需求,本文在 OGC Filter 的基础上,制定了具有异构数据适应能力的 SIG 空间数据查询规范,具体组成如图 1 所示。

1. 全文以<request>为根元素;
 2. <request>中分为 3 部分:
 <target_list>:需要返回的属性列名,最终解析为 select 子句;
 <order_list>:规定要素以什么规则排序,最终解析为 order 子句;
 <filter>:查询条件,最终解析为 where 子句;
 3. <filter>中由若干并列或嵌套关系的<condition>元素组成:
 并列的<condition>之间关系由其父<condition>属性规定<relation>为 and 或者 or;
 嵌套:每层嵌套由括号表示;
 整个<filter>中的条件描述解析可使用树的遍历来进行解析,操作为递归+while 循环;
 4. 最底层<condition>中由 3 个元素构成:
 <operator>:操作符,例如:<, >, =, like
 <attr>:条件字段
 <value>:值

图 1 SIG 数据查询标准的 XML 定义

Fig. 1 XML definition of SIG data query standard

目前,SIG 数据服务中建立的查询元数据标准,可以满足大多数情况下对于栅格数据的操作和矢量数据文件级别的处理。

2.3 语义技术对于数据服务的改善

在空间信息服务领域引入语义和本体技术,其主要目的是为了实现在多源空间数据服务过程中的智能化和精确化,解决目前因为语义关系而造成的空间数据服务的无法精确选择、空间数据服务中的过多依赖人工交互、关键词查询中的匹配错误等问题。本体在空间数据服务的主要作用可以归纳为以下几点:

(1) 利于异构数据的集成。本体是关于特定领域的公共概念知识。通过构建本体,以及建立本体与数据,本体与本体的语义联系,可以解决空间数据在结构上、语法上、系统上、语义上的异构问题。从而屏蔽数据的异构性,为用户提供一个集成的、语义一致的空间数据服务^[6]。例如,以本体技术为核心,通过对地理空间信息服务领域的概念建模,可以建立空间实体、几何属性的标准定义,空间属性数据的标准定义,元数据标准的定义等,形式化地描述领域概念以及概念间的关系,使计算机能够理解这些描述信息,从而实现空间信息共享和语义互操作。

(2) 提供基于语义的数据检索技术。通过描述数据的语义元数据,可以实现对数据的语义检索。目前,基于语义的数据检索主要体现在 3 个方面^[7],包括基于基本概念的数据检索,即通过根据本体定义的基本概念,来检索符合该概念的语义数据;基于语义的复杂限制查询,即通过本体定义的概念,关系构建的复杂限制查询,来检索符合条件的数据;第三,通过推理获得新的语义关系,即通过已有的数据间的语义关系,以及语义关系的推理规则来获得新的数据间的语义关系。

(3) 基于本体的数据抽取技术。根据本体定义的领域知识,对无序的数据进行组织,向用户提供一个有特定语义知识结构的数据集合。运用本体来指导信息抽取,实现按照领域知识对数据进行抽取、组织的方法。按照领域知识抽取出来的数据,根据特定语义模板的定义,被组织成为有意义的文字信息。

建立空间数据本体的关键,首先需要对比国际上主要的空间数据本体化标准工作。通过比较和分析,在概念本体层面来对主要国际标准工作进行支持。同时,该工作确保了本研究形成的空间数据本体具有和其他国际空间数据本体研究项目一致的底层基础,可以进行实现数据互操作。这里,我们重点研究 GCMD 本体、SWEET 本体和基于 ISO/TC211 标准的地理信息本体,相关对比分析如表 2 所示。

表 2 地理本体对比分析

Table 2 Comparative analysis of geographic ontology

地理本体	优点	缺点
NASA GCMD	开发项目,鼓励第三方的维护和管理	受控关键词缺乏丰富的语义,并列式主题划分结构不适和构建分层本体
Drexel ISO19100	形成地理本体的合理分类	本体完整性存在一些问题
SWEET	相对全面和稳定的地理信息本体标准	/

目前,SWEET 本体是一套相对全面和稳定的地理信息本体标准。SIG 中基于空间信息本体的数据表达,正是在 SWEET 本体的基础上开展。

3 数据资源描述体系及元数据策略

本节讨论网格环境中的空间数据服务将会在哪些层面涉及到数据资源描述机制,以及如何选取相应的标准和合理的策略。

3.1 服务环节与相关标准映射关系分析

前面提到,很多标准化组织都在致力于空间数据描述标准的相关研究,不断地推出和完善,但是由于标准的制定和发展是一件很艰难的工作,无法出现一个能让所有空间数据业务运行系统都遵循的元数据标准^[8]。即使有了某个这样的标准,人们在具体组织数据时,由于有着不同的应用背景,在最终实

现的内容和表达上并不能确保完全的一致。

因此,研究工作的核心,应该是在现有标准的基础上,根据对空间数据资源描述标准的相关分析,结合网格中数据服务模型实现时的具体需求,对其进行必要的摘录、补充和改进。尤其是针对空间数据服务所涵盖的诸如信息查询检索和管理协同的具体功能,研究每个环节与相应数据标准和表达规范的映射关系,建立支持分布异构数据共享的空间信息资源描述标准和元数据策略。

首先从空间数据服务体系结构的视角,分析在哪些层面和服务环节会涉及到数据资源的描述和相关标准,如图 2 所示。

将数据描述机制集中在数据在线处理规范、数据交换标准和协议、数据服务发布规范 3 个方面,包括数据资源注册、数据查询和获取操作标准和

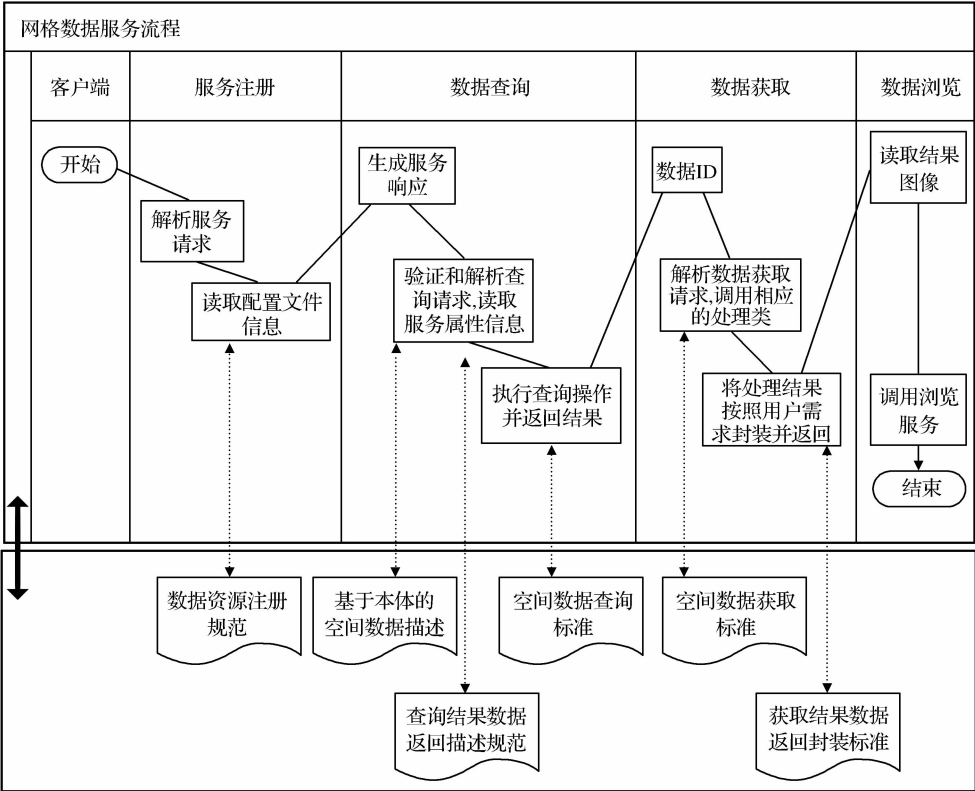


图 2 空间数据服务流程与相关标准映射关系

Fig. 2 Data service processes and related standards

返回结果表达规范等。研究空间数据网格服务的核心是支持多源、异构、分布性的空间信息,只有在每个层面都选取标准化和适应网格环境的资源描述方法,才能确保数据服务模型对于国际标准进行支持的底

层基础,确保建立在异构数据源之上的信息互操作。以空间信息网格 SIG 平台的经验为基础,总结空间资源描述环节和相应标准规范的映射关系如表 3 所示:

表 3 SIG 数据服务节点资源描述标准
Table 3 Resource description standards of SIG data service node

类别	名称	主要国际标准	适用网格数据服务标准	SIG 数据服务节点参照标准
数据在线处理规范	空间元数据基本标准	CSDGM、GCMD、DCMS、SDBCM、GDDD、ISO/TC211	CSDGM、ISO/TC211	ISO/TC211
	本体构建标准	GCMD、SWEET、ISO/TC211	SWEET、ISO/TC211	SWEET
	数据查询标准	Z39.50、OGC Filter	Z39.50、OGCFilter	OGC Filter
	数据获取标准	OGC WCS	OGC WCS	OGC WCS
数据交换标准和协议	查询结果返回标准	OGC WCS	WCS Coverage Description	SIG QUERY
	获取结果返回标准	OGC WCS	WCS Coverage Description	WCS Coverage Description
数据服务发布规范	数据服务加载	WSDL	WSDL	WSDL
	数据服务发布	UDDI	UDDI	UDDI

以上针对空间数据服务流程的特点,从 3 个方面分析空间数据描述规范及相关标准在具体服务环节的映射关系,包括在空间信息领域主要国际标准,适合构建网格服务的数据标准,及 SIG 数据服务节点资源描述主要参照的标准。

3.2 空间数据资源网格化描述模型

空间数据资源的网格化描述模型,实际就是研究如何建立一个统一的空间数据描述规范结构,用于约束和定义基于网格环境的数据服务以何种方式表达和管理空间信息元数据。基于该模型,可以建立网格中的空间数据注册、搜索、获取等操作。空间数据资源的网格化描述模型对于数据服务的应用起着至关重要的作用,因为只有功能完善的资源描述模型才能充分地表达各种空间元数据信息,才能增加网格数据服务的标准化和通用性,使得满足分布式多源异构空间数据源的集成共享需求。

表 3 阐述了遵照空间数据标准扩展方法,选定的 SIG 数据服务主要参照标准。进一步,对于网格环境中数据操作和交换的一系列环节,需要重点考虑一些特殊数据属性。具体分析如下:

① 对于服务的注册,数据源的运行结构、提供组织和对外发布方式是需要提交和说明的必选项;② 对于数据的查询,传感器、经纬度和时间范围信息是用户密切关注的查询条件描述参数,该参数规范必须具有异构适应能力;③ 对于数据的获取,具

体按需处理方式和获取条件描述是该环节的重点;④ 对于每个请求任务的返回结果,需要包括数据项标识、访问引用标识和返回图像质量信息等;⑤ 而对于改善数据服务质量的空间数据本体构建,在分析比较确定合适的空间信息本体后,本体的概念定义、属性定义和关系定义是重点需要规范的。

基于上述分析,通过结合数据服务各个环节的需求与特点,在对相关标准进行相应补充和改进的基础上,我们研究并提出了 SIG 空间数据网格描述模型,形成面向网格服务流程的数据资源描述规范系列。具体组成和关系如图 3 所示。

这样,对于每个具体的数据服务,所需要的工作实际上可以看做是该模型与具体空间数据集之间的转换方法。而对于进一步建立面向地学计算和应用的数据服务,该模型是所有应用网格服务的基础。

基于空间数据描述标准的元数据表现形式对数据的搜索、条件匹配以及用户获取条件都有影响。因此规定适当的数据表现形式是实现空间数据网格服务的重要环节,必须满足空间数据资源网格化描述标准的特点。首先,空间元数据标准所支持的类型复杂,由简单类型的数据如 Int、String,也有复杂的结构型类型。其次,描述标准中一些数据项相互间存在一定的逻辑关系,如 A 和 B 必须同时存在或者 A 和 B 只能有一个存在。第三,数据标准有关键性的概念,一些项是必选项,另一些项

是可选项。

基于上述考虑,论文选用 XML 作为数据资源网格化描述标准的表现形式,配合 XMLSchema 文件一起使用,能有效地对数据标准进行验证(图 4)。

用 XML 的大纲文件表示元数据标准,用户的查询条件和查询结果是符合该大纲文件的 XML 流。不同处理环节的数据标准用不同的大纲文件表示,这样,对数据标准的转换就是对 XML 数据的操作。

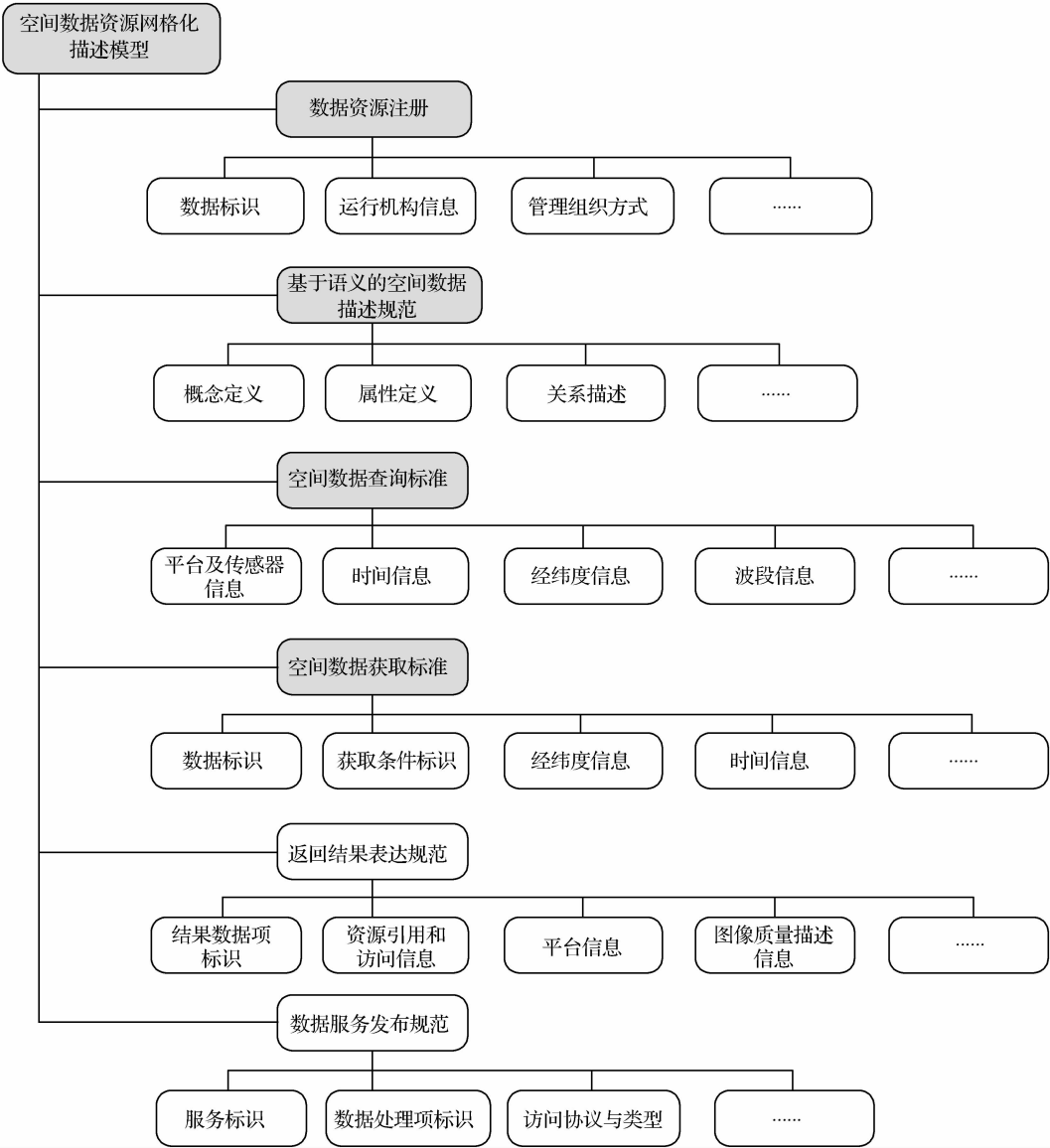


图 3 空间数据资源网格化描述模型

Fig. 3 Description model of spatial data grid resources

```
<xs:complexType name="BOUND">
  <xs:sequence>
    <xs:element name="WEST_LON" type="xs:double" default="73"/>
    <xs:element name="EAST_LON" type="xs:double" default="140"/>
    <xs:element name="NORTH_LAT" type="xs:double" default="60"/>
    <xs:element name="SOUTH_LAT" type="xs:double" default="15"/>
  </xs:sequence>
</xs:complexType>
```

图 4 基于 XML Schema 的数据校验

Fig. 4 Data validation based on XML Schema

4 结 语

上述网格空间数据资源描述模型,经过不断地改进和补充,已经形成较稳定的结构,运用在 SIG 的空间数据管理、交换和使用上。

本文针对数据共享与管理协同的不同环节,分别采用相应的标准与规范,并结合各环节的需求与特点,进行了相应的补充和改进,形成面向网格服务流程的不同层次数据资源描述规范系列。借助于该系列化规范与标准,实现了网格中不同数据服务环节下空间数据的不同描述,从而为网格化空间数据服务在不同层次的互操作提供了良好的底层基础。与其他方法相比,该方法有效结合网格中的数据操作特点,进行空间数据标准的选择和扩展,在资源表述层面建立分级描述,为数据共享和服务协同奠定基础。

参考文献(References):

[1] Zhang Li, Gong Jianya. Research and Implementation of the Management of Geospatial Metadata[J]. Journal of Wuhan Technical University of Surveying and Mapping, 2000, 25(5): 127-131. [张立, 龚建雅. 地理空间元数据管理的研究与实现[J]. 武汉大学学报(信息科学版), 2000, 25(5): 127-131.]

[2] Wang Juanle, You Songcai, Xie Chuanjie. Study on the Applica-

tion of Metadata Techniques in the Geosciences Data Clearing-house[J]. Geomatics World, 2005, 4(2): 36-40. [王卷乐, 游松财, 谢传节. 元数据技术在地学数据共享网格中的应用探讨[J]. 地理信息世界, 2005, 4(2): 36-40.]

[3] Feng Xiangyun, Xiao Long. A Comparative Study of Commonly Used Metadata Formats in Abroad[J]. Journal of Academic Libraries, 2001, 19(4): 9-17. [冯项云, 肖珑. 国外常用元数据标准比较研究[J]. 大学图书馆学报, 2001, 19(4): 9-17.]

[4] He Yi. The Condition Application and Problem of Z39. 50[J]. Information Science, 2002, 9: 944-945. [贺宜. Z39. 50 的现状、应用及问题初探[J]. 情报科学, 2002, 9: 944-945.]

[5] Miao Lizhi, Zhang Shuliang, Wu Lan, et al. GML Spatial Data Query based on Filter Encoding[J]. Computer Engineering, 2008, (3): 59-61. [苗立志, 张书亮, 伍蓝, 等. 基于 Filter Encoding 的 GML 空间数据查询[J]. 计算机工程, 2008, (3): 59-61.]

[6] Li Hongwei. Research of Geoinformation Service based on Ontology[D]. Information Engineering University, 2007. [李宏伟. 基于 Ontology 的地理信息服务研究[D]. 解放军信息工程大学, 2007.]

[7] Shum S B, Domingue E M J. An Ontology-based Digital Library Server for Research Documents and Discourse[J]. International Journal on Digital Libraries, 2000.

[8] Jiang Zuoqin, Liu Ruomei, Yao Yanmin, et al. Geographic Information Standards Reference Models[J]. Land and Resources Informatization, 2003, (3): 11-17. [姜作勤, 刘若梅, 姚艳敏, 等. 地理信息标准参考模型综述[J]. 国土资源信息化, 2003, (3): 11-17.]

Research of Service-oriented Data Sharing Standard
in Grid Environment

Zeng Yi¹, Li Guoqing²

(1. Information School, Beijing Forestry University, Beijing 100083, China;

2. Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing 100086, China)

Abstract: The research of spatial data specification and description mechanism is very important for spatial information sharing and services collaboration, especially in grid environment. Most of the studies focus on the content composition and implementation of specific standards, but not aims the procedure in data integration. This paper based on the data service architecture in grid platform, and analysis the relevant data specification and standards. We study the resource description method and metadata strategy, and summarize a standards series of spatial resources description for data grid service. The paper gives the application in Spatial Information Grid platform for case study, and discusses the result in detail.

Key words: Spatial data standard; Data grid; Data service