

多源分布式森林资源监测数据集成研究

杨志高, 吴雨光*

(中南林业科技大学 理学院, 湖南 长沙 410004)

摘要:森林资源数据分布广泛且来源多样,通过对其进行共享研究,可以为林业信息的共享和林业信息资源利用效率的提高,提供高质量的数据服务。通过对森林资源数据使用情况进行分析,采用数据集成设计方法,基于 ROA 框架,使用 REST 方式,提出了一个关于多源分布式森林资源数据集成系统设计方案,用以满足各林业部门对森林资源数据集成整合使用的要求。

关键词:森林资源;数据集成;ROA;REST;QGIS

中图分类号: S758.4 **文献标志码:** A **文章编号:** 1001-7461(2011)02-0171-09

Integration Solution Research for Multi-Source Forest Resources Data

YANG Zhi-gao, WU Yu-guang*

(School of Sciences, Central South University of Forestry and Technology, Changsha, Hunan 410004, China)

Abstract: Forest resource data are widely distributed and diversified. Research of sharing of forest resources data can provide high-quality data services for sharing forestry information and efficiency in the use of forestry information resources. The storage and usage of multi-source forest resource data were analyzed. Adopting data integration designing ideas, and based on ROA and REST, a system designing solution on the integration of multi-source distributed forest resources data was established to meet the data integration requirements of forest department at various levels.

Key words: forest resources; data integration; ROA; REST; QGIS

林业信息化建设是现代林业建设的重要组成部分。长期以来,林业信息化建设中存在诸多问题,例如:缺乏总体规划布局,大部分林业业务应用系统自行建设,自成体系,形成了许多信息孤岛;林业信息化标准缺乏全面性、系统性和一致性。目前,虽然建立了一些资源信息中心和元数据发布中心,但都主要采用文字描述的方式对元数据进行阐述,对于大规模林业数据共享依然不够便利^[1]。森林资源数据是林业信息资源的主要组成部分,存在地域分布广、数据来源多样、标准不统一和异质性大等特点。特别是采用不同 GIS 系统大量建立森林方面的空间数据,如:ESRI 公司的 Arc/Info Coverage、Arc-Shape Files、E00 格式数据、MapInfo 的 MIF 格式数据、Intergraph 的 dgn 格式数据,更是加大了共享的难度。大量的异质数据,使得数据相互整合的成

本增加,人力、物力浪费在不必要的资料处理中^[2]。有利的是当前林业行业中对于小班、林相、森林基本统计数据等这些林业基本信息已形成了比较统一的元数据标准。在数据传输和数据处理方面,国际上有着许多开放的标准可以采用,如 WMS、WFS 和 GML 等。同时,大量符合开放标准的开源软件的开发与应用,如 Geoserver^[3]、QGIS(Quantum GIS Mimas)^[4]等,也为数据共享提供了完备的、能快速获取到且费用较低的技术手段。

本研究提出的多源分布式森林资源数据集成系统设计方案,采用数据集成设计方法,目的是为多源分布式森林资源监测数据提供一个比较完善的数据共享方法,以期在提高整合效率的同时,兼顾各方面数据要求的全面性。

收稿日期:2010-03-23 修回日期:2010-05-15

基金项目:国家林业局林业科技支撑计划森林资源综合监测技术体系研究(2006BAD23B0204-3)。

作者简介:杨志高,男,副教授,主要研究方向林业信息工程和 3S。

* 通讯作者:吴雨光,男,硕士生,主要研究方向为林业信息工程和 3S。E-mail:cekm@163.com

1 系统方案设计

1.1 系统设计思路

应用集成模型是一种用来集成软件的特定方法和结构,它为各方法和结构以及各种要求与限制提供了一系列不同的选择,并根据实际要求有着不同的着力点,如:实现集成的简单性;对于不同配置集成的可重用性;可用集成方法的广泛度;对于执行集成的过程中要求的专门技术。集成模式主要分为:表示集成;功能集成;数据集成^[5]。其中数据集成是当前国际上公认的数据整合、数据同步、数据迁移、数据质量的解决方案。在本系统设计中采用数据集成模式,跳过显示界面和业务逻辑模块,直接进入应用软件的数据结构或数据库来创建新的集成,达到系统数据整合的设计要求(图 1)。



图 1 林业数据集成示意图

Fig. 1 System data flow diagram

数据集成按照数据源所处的位置的不同可以分为 3 类,第 1 类是中心式,采用一个中心数据源,其他数据源均汇总到此,数据使用者使用时,只需查询这个中心数据源就可以获取到需要的数据。第 2 类是分布式,数据源分散在各地,数据使用者需要数据时,查询各个数据源获取数据。第 3 类是中心分布式,中心分布式可以表示为:数据源小区域集中,大区域数据源分布、集中数据源元数据,数据源分布或上述 2 种方式结合。根据森林资源数据的特性,采取第 3 类方式符合森林资源数据目前的状况。在数据集成过程中碰到的问题主要涉及到 3 个方面。

1.1.1 属性问题 属性值主要是用来描述实体数据的特性,但是属性值的不兼容,例如数据类型、数据长度、数据精度、是否为空等问题,都有可能会导致异质数据整合失败。解决属性值不兼容可以采用数据元数据来处理,例如采用数据元数据来定义长度、单位或是数据解释等。

1.1.2 数据格式问题 采取不同软件建立的不同数据格式文件,造成了用户在数据整合时的问题。其解决办法是采用将原始数据格式转换成公认的标准数据格式,公认的标准数据格式必须能完整的描述原始数据,否则在转换过程中会造成数据丢失,从而影响到整合的数据质量。

1.1.3 语义问题 对于相同的对象,使用不同的语

法或语句描述,导致语义差异。解决语义问题的方法,是建立应用领域的知识本体,建立异质数据库之间的对应关系或对相同对象标准规范化。

根据上述思路,本设计通过建立基于 GML (Geography Markup Language) 的 FML (Forestry Markup Language) 数据描述规范化语言来统一属性问题和数据格式问题。对于语义问题,在 FML 中对各类森林对象采用标准规范化进行处理。对于森林数据的数据传输流通和整合处理机制采用基于遵循 REST (Representational State Transfer) 设计原则的 ROA (Resource-Oriented Architecture) 进行设计。ROA 是一种把实际问题转换成 REST 式 Web 服务的方式,能够容易地把一个个数据源划分为一个个 REST 式资源 (RESTful resources)^[6],它使用 URI、HTTP 和 XML 具有跟其他 Web 应用一样的工作方式,有利于依据 FML 规范转换的数据进行传输和整合处理。

1.2 系统框架设计

依据 1.1 中的设计思路,系统设计成 3 个部分:元数据应用服务、遗留区域和代理域(图 2)。

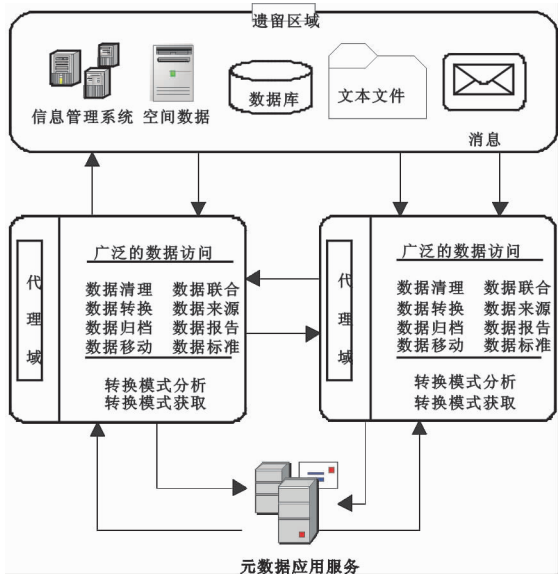


图 2 系统框架图

Fig. 2 System frame diagram

1.2.1 元数据应用服务 (Metadata Application Server, MetaAS) 元数据应用服务起着数据发布,标准化和数据中心的作用。主要目的有:存储依据 FML 建立的应用模式,FML 的应用模式基于 GML 应用模式建立,用于对数据的定义和转换,称为转换模式;提供标准化的对象命名,坐标转换和单位;数据集元数据发布。

1.2.2 遗留区域 以各种数据格式存放异质森林原始数据的区域,如数据库、文本文件和空间信息等。遗留区域可以是一个单独的物理存在,也可以

是由多个分散遗留区域组合成一个虚拟遗留区域。无论是物理的或是虚拟的,在系统中遗留区域与代理域为一对一关系。

1.2.3 代理域 代理域是一个综合区域(图 2)。代理域可以分为 3 个部分:广泛的数据访问、数据处理和转换模式处理 3 个部分。

广泛的数据访问表示代理域拥有多种数据访问接口,可以无缝访问和处理地理数据,例如:数据库中地理数据、shp 格式数据、mapinfo 格式数据等。转换模式处理指的是从元数据应用服务中获取相应的转换模式文件信息。数据处理表示将原始数据通过转换模式转换成基于 FML 的数据描述文件或获取基于 FML 的数据描述的文件将转换成遗留区域数据。并且依据 FML 应用模式,转换遗留数据,形成一个规范化的数据区域。这样一方面可以减少从遗留区域转换成符合 FML 描述标准数据的开支,另一方面可提高数据的整合共享效率和能力。

2 系统原型实现

2.1 FML (Forestry markup Language) 的构建

由 OGC(open geospatial consortium)推动,被

ISO/TC211 所采用为 ISO 19136 标准的地理标记语言 GML^[7],是目前广泛应用的地理数据流通格式,可以很方便的描述空间和非空间数据。在本设计中采用 GML Schema 来定义林业领域的应用模式,透过一组概念性的 Schema Language 来定义数据的内容及属性,根据设计的应用模式将森林空间数据或非空间数据转换为 GML 标准格式数据,以去除数据异质性。因为是一个为林业领域空间数据所设计的 Application Schema,所以称为林业标记语言(forestry markup language,FML)。同时针对 FML 元数据部分引入了 ISO:ISO9115 规范和扩展版本 ISO9115-2 规范^[8],这两个版本已经分别被 OGC 组织实现为 GMD 和 GMX 标准。

遵循 GML 应用模式的构建方法,在 FML 中采用 fml 为命名空间名。FML 数据描述结构从 gml:AbstractGMLType 集成而来,它包含了属性数据(AttributeData),元数据(Metadata),要素容器(gml:FeatureCollection)。同时引入了单位定义(unit definitions),坐标系统定义(reference system definitions),标准编码列表(CodeList)(图 3)。

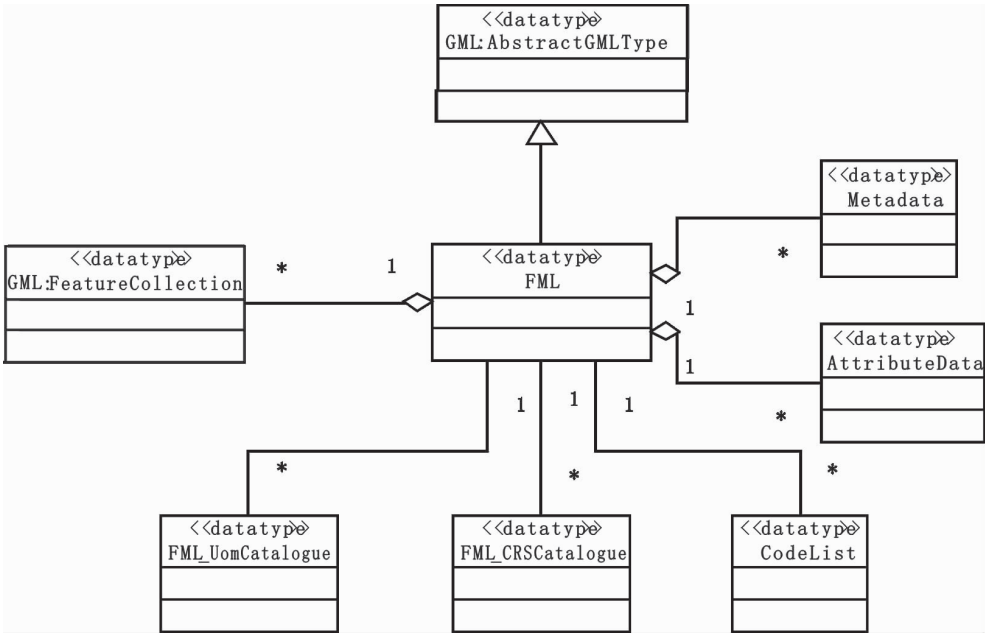


图 3 FML 数据描述构架图

Fig.3 Data description framework chart

2.1.1 单位定义 (unit definitions) FML_Uom-Catalogue 的目的是将在林业行业中经常使用的单位语义标准化,如:hm²、m² 等,从而使数据使用者在整合共享数据时,不会因单位的不统一,而导致数据上的误差。FML_Uomcatalogue 概念模型,主要由 3 部分组成,BaseUnit 包含基本单位,DerivedUnit 包含导出单位,组合单位,ConventionalUnit

包含单位之间的转换(图 4、图 5)。

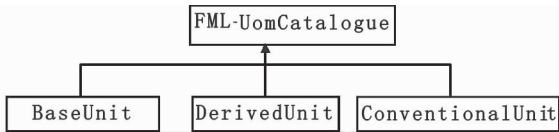


图 4 FML-UomCatalogue 概念模型

Fig.4 FML-UomCatalogue conceptual model



图 5 FML-Uomcatalogue 的 BaseUnit 单元定义

Fig. 5 Definition of baseunit FML-UomCatalogue unit

2.1.2 坐标参考系统定义(Reference System Definitions) 坐标参考系统定义主要针对空间数据在建立空间数据时采用不同 GIS 软件建立,因不同的 GIS 软件对于坐标参考系统的命名的方式是多样的,导致坐标参考系统定义语义的不一致和转换障碍。欧洲石油测量组织(European Petroleum Survey Group, EPSG)^[9]依照 OGC 坐标参考系统标准规范定义了全球各地的 CRS,并为各大 GIS 厂商和各类 GIS 软件所支持。通过采用 EPSG 标准和 ISO/OGC 定义的 CRS 坐标参考系统模型,形成 FML 中对于我国主要坐标参考系统的标准的描述,如:Beijing 1954, Xian 1980 等。我国的 2000 国家大地坐标系(CGCS2000),在 EPSG 中被命名为

China System 2000,但尚未给予唯一编号。EPSG 是一个可以扩展的标准,在 FML 中被定义为 EPSG:90001,并建立了 CGCS2000 的坐标描述(图 6、图 7)。

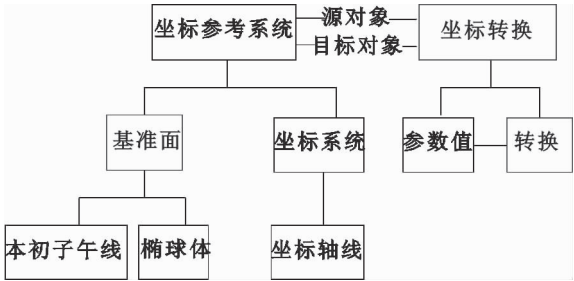


图 6 ISO/OGC 的坐标参考系统模型

Fig. 6 ISO/OGC coordinate reference system model



图 7 CGSC2000 坐标参考系统定义

Fig. 7 CGSC2000 coordinate reference system definition

2.1.3 标准编码列表 (CodeList) FML 中定义的 CodeList 继承于 GML Dictionary 结构 (图 8), GML:Dictionary 是 GML 为定义常量或标准数据提供的一个数据描述字典,根据目前已经建立最权

威的林业科学标准公用数据,如森林树种的分类、名称;地类的划分类别;林种名称等,基于 Dictionary 建立 FML 公用字典,用来规范林业领域中共用数据在使用时语义不一致性问题。



图 8 Codelist 简单实例

Fig. 8 Simple case of code list

2.1.4 元数据描述 (MetaData) 数据集元数据是数据描述文件的一部分,ISO9115 中提出了一个阶层式元数据资料架构。在标准中主要分成 3 个阶层 6 个部分,3 个层级包括数据集系列 (Dataset-Series)、数据集 (Dataset) 与要素 (Feature)。林业数据存在着同一组织建立的同一类数据,这类数据往往数据采集方式、供应者信息、数据精度、数据比例尺都一样。这样可以把要素级这一层,提升到数据集级层。FML 主要采用 gmd 中的 gmd: MD-Metadata 中的 6 个子元素 (表 1)。

表 1 FML 引用 MD-Metadata 子元素		
Table 1 FML reference MD-Metadata subelemen		
元素	必选	说明
gmd:contact	是	数据发布者信息
gmd:dateStamp	是	数据源发布时间
gmd:metadataStandardName	是	FML 应用模式
gmd:metadataStandardVersion	是	版本号
gmd:referenceSystemInfo	否	坐标参考系统
gmd:identificationInfo	是	数据精度、数据源比例尺和数据最小范围

2.1.5 属性数据 (AttributeData) 和要素容器 (gml:FeatureCollection) 在 FML 应用模式中最基本的数据单元是要素。根据 GML 定义要素类型,可以容易的用要素来表示空间数据和非空间数据 (图 9)。

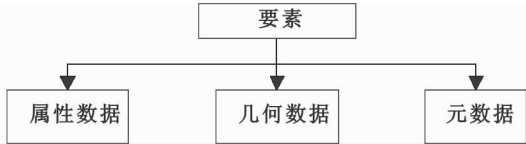


图 9 要素结构

Fig. 9 Feature structure

在实际操作中,数据库中一条数据,可以看成是一个要素,一个要素集可以看成是一个数据库。而要素容器就是一个要素数据集,它包含着要素。基于要素容器构建了 FML 数据描述的基本架构 (图 10、图 11)。

2.2 元数据应用服务

2.2.1 元数据应用服务组成部分 MetaAS 由 4 部分组成 (图 12)。

Postgresql:对象关系型数据库管理系统 (OR-DBMS),作为 MetaAS 的数据中心。

PostGIS: PostgreSQL 的空间引擎,它赋予了 Postgresql 空间处理能力,包括空间对象、空间索引、空间操作函数和空间操作,并遵循 OpenGIS 的规范。通过 PostGIS,在 Postgresql 中建立了 1 : 1 400 万中国空间地理信息数据库。在数据使用者使用 MetaAS 时,提供更为详细的地理辅助信息。

GeoServer:地理信息系统服务器,提供 WMS、WFS 和 WCS 数据服务,允许用户对要素数据进行更新、删除、插入操作。

Apache:对外发布 WEB 服务器。进行元数据

以及数据资源的发布和操作,通过 Openlayer 对于地理数据的可视化展示和操作。

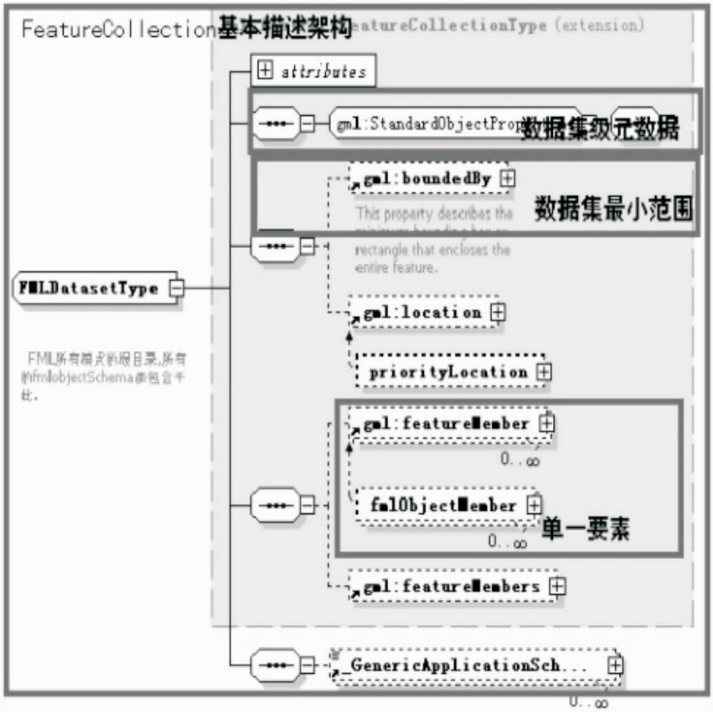


图 10 FML 结构

Fig. 10 FML structure

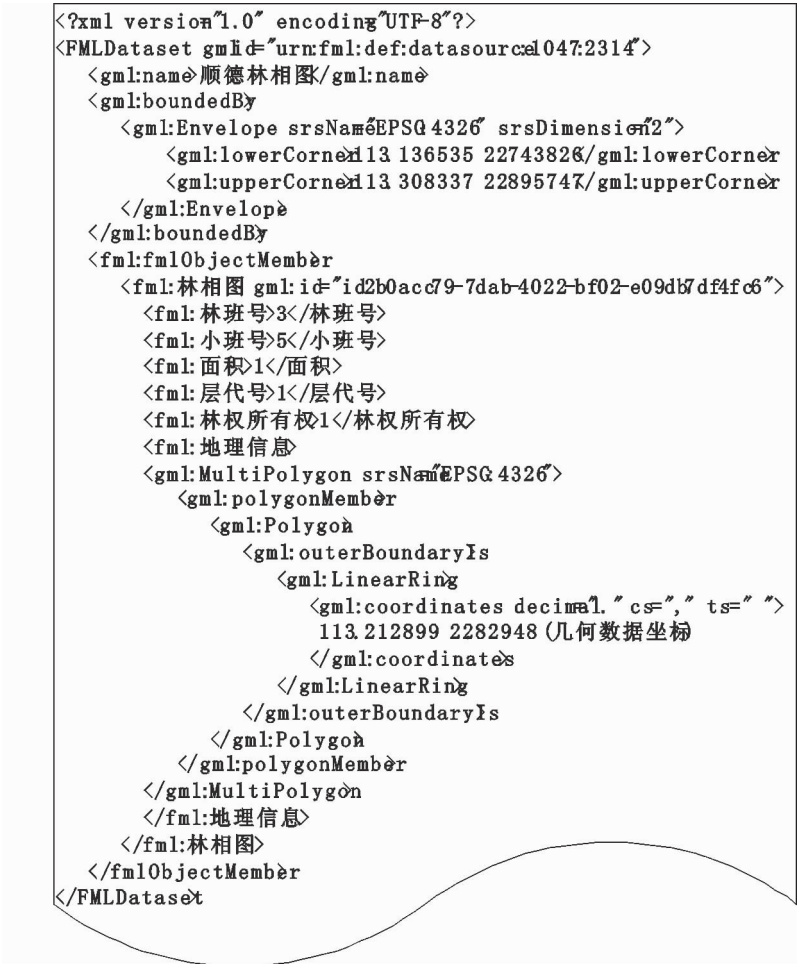


图 11 FML 数据描述

Fig. 11 FML data describe

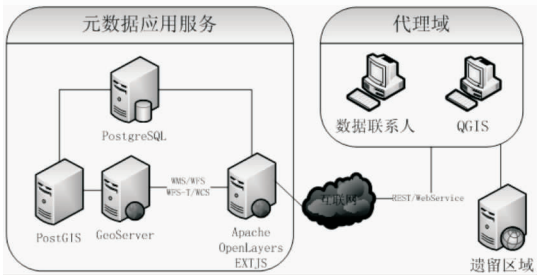


图 12 系统原型架构图

Fig. 12 System prototype chart

2.2.2 元数据应用服务主要业务流程

2.2.2.1 数据注册 在 MetaAS 中,数据管理采用分层管理,分为数据源和数据集,两者之间采用 1:n 的映射关系。从上面 FML 中可知,元数据实际由两部分组成,数据集系列级和数据集级 2 层组成,分别对应现实中的数据源和数据集。用户进行元数据登记时,对于数据集范围的描述,地理标识名称和数据集最小范围都过于抽象,对于查询空间分布范围比较大的数据比较适用。但是,对于微小数据集往往被淹没在这些大的数据集中,不能够被体现出来。

如果数据信息比较多,数据使用者在不能明确知道地理编码或地理坐标的情况下,想找到数据集将比较困难。因此,在数据联系人登记属性资料时,并在 OpenLayer 展示的地图上以可视化的方式加以标注,使用 WKT(well-known text)格式以 POINT 或 POLYGON 类型通过对数据集的位置进行地理编码插入到 Postgresql 数据库中。这样便于数据使用者在查询数据集的同时,在地图上展示出数据集所在的位置。

2.2.2.2 数据获取 获取方式采用 REST Web 服务方式,数据源基于 ROA 方法将自身数据资源发布变成一个个 REST 服务。MetaAS 统一管理数据源数据资源 REST 服务,并统一对外发布。数据使用者从 MetaAS 中查询到所需要的数据,从 MetaAS 中发送数据请求信息,MetaAS 对数据使用者的请求信息进行验证,验证通过的,提供给数据使用者数据获取方式。数据使用者通过得到的获取方式,获取符合 FML 数据。这样既利于降低数据获取差异性和难度,也利于消除数据交换障碍(图 13)。

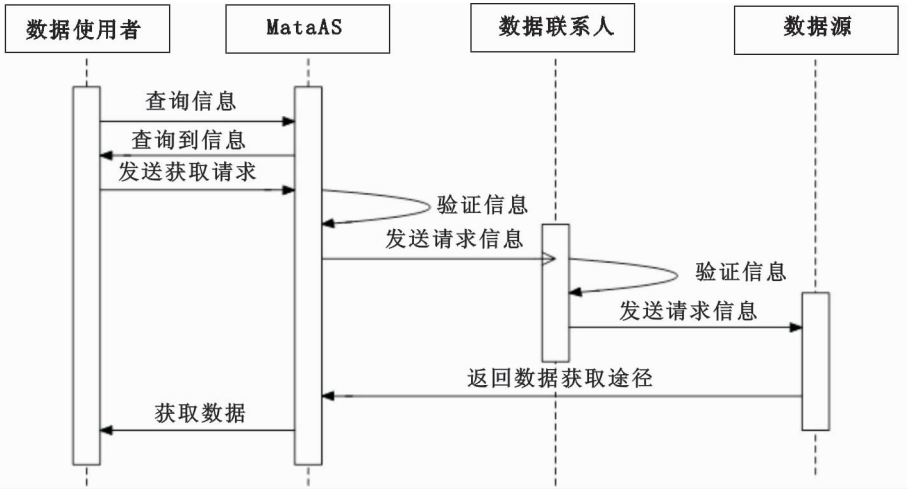


图 13 数据获取示意图

Fig. 13 Schematic diagram of data acquisition

2.3 遗留区域和代理域

遗留区域就如第 1.2 中所述,数据的原始文件区域。代理域利用从 MateAS 中获取到的 FML 模式规范化遗留区域数据,并对外发布 FML 规范化数据。规范化遗留区域数据的方式有很多,如:重新构建新的规范化数据库或采用 ETL^[10] 技术对外提供规范化数据。

在遗留区域通过 ETL 方式,抽取转换成 FML 模式数据。QGIS 是由 OSGeo (Open Source Geo-Spatial Foundation) 采用 python 和 C++ 语言开发,以 GUN(General Public License)授权发布的一

个跨平台开源地理信息系统平台,支持用户自行扩展。本系统基于 QGIS,使用 python 进行功能扩展,为符合 FML 规范数据的操作提供了一个统一的前台数据操作平台。支持对多种数据格式与标准的处理,包括数据库数据,如 PostgreSQL/PostGIS、Oracle/OracleSpatial,以及 OGR 函数库;多种矢量与栅格数据,如 ESRI Shapefiles、MapInfo、SDTS、GML、GeoTiff、Erdas Img、ArcInfo Ascii Grid、JPEG、PNG; GRASS; OGC 标准服务,如 WMS、WFS 等。数据使用者可以通过 QGIS 将抽取到的本地数据转换成符合 FML 规范的数据,进行发布

和获取符合 FML 规范的 XML 文件,通过 QGIS 验证后,直接以图层方式加载,操作各类要素信息,并可以转换成不同的格式。

3 原型测试

3.1 测试步骤

3.1.1 数据资源准备与发布 根据《林业科学数据仓库和数据共享技术标准语规范》(第一辑)中的内容,生成了符合 FML 规范的 3 种模式:森林基本统计属性模式、小班模式和林相模式,登记到 MateAS 中。代理域分别对湖南省 2007 年林业统计 XLS 文件格式数据、湖南省宜章 2005 年北京 1954 shp 文

件格式小班图数据和广东省佛山市顺德区 2004 年西安 1980 mapinfo 文件格式林相图数据这 3 种数据,依据从 MateAS 获到相应模式,采取 ETL 方式,通过 QGIS 转换成符合 FML 规范的数据。在代理域采用 GlassFish 应用程序服务器和 Java EE 5 支持的 JAX-WS2.0 实现 REST 风格的 WEB 服务,并将资源 URI 及其元数据信息和其他信息发布到 MataAS 中。

3.1.2 获取数据资源 数据使用者在 MateAS 中采用全文搜索方式检索所需数据资源元数据,选取数据资源,发送获取请求(图 14)。



图 14 元数据数据查询页面

Fig. 14 Meta data query page

MateAS 验证通过后,数据使用者调用对应资源 URI 获取所需数据。客户端提供展示功能,数据使用者可以根据需要定制更小的数据集结果。代理

域 web 服务根据为数据使用者的要求,提供一个 .gz 格式的 XML 数据压缩包(图 15)。

查询项目 [森林] 查询内容 [宜章县森林资源]		搜索结果													
县名	乡名	村名	小班号	地类	林种	立地类	蓄积量	蓄积量	平均年	平均年	平均年	树种组	每公顷	每公顷	每公顷
宜章县	浆水乡	新华村	431022150700005	111	33	17	12.000000	10.000000	10	1850	180	杉木组	145	0	0
宜章县	浆水乡	浆水村	431022150100106	131	33	0	0.000000	1.500000	0	0	0	杉木组	40	0	0
宜章县	浆水乡	浆水村	431022150100103	111	33	18	12.000000	8.000000	10	1680	180	杉木组	40	0	0
宜章县	浆水乡	天塘村	431022150900060	131	51	0	0.000000	1.600000	0	501	0	杉木组	0	0	0
宜章县	浆水乡	新石村	431022151200021	111	11	8	10.000000	5.000000	0	140	0	马尾松	42	0	0
宜章县	浆水乡	编溪村	431022150300068	132	11	0	0.000000	1.700000	0	0	0	杉木组	0	0	0
宜章县	浆水乡	编溪村	431022150300134	113	33	5	15.000000	6.000000	0	2000	0	杉木组	0	0	0
宜章县	浆水乡	洛洞村	431022150400124	111	33	7	10.000000	6.000000	10	1401	150	阔叶松	9	0	0
宜章县	浆水乡	编溪村	431022150300104	111	16	17	12.000000	10.000000	0	180	0	杉木组	22	0	0
宜章县	浆水乡	编溪村	431022150300118	132	16	0	0.000000	2.000000	0	0	0	杉木组	0	0	0
宜章县	浆水乡	编溪村	431022150300118	131	33	0	0.000000	1.500000	0	0	0	杉木组	0	0	0
宜章县	浆水乡	编溪村	431022150300133	210	0	0	0.000000	0.000000	0	0	0	杉木组	0	0	0
宜章县	浆水乡	编溪村	431022150300132	132	16	0	0.000000	1.400000	0	0	0	杉木组	0	0	0
宜章县	浆水乡	编溪村	431022150300150	161	0	0	0.000000	0.000000	0	0	0	杉木组	0	0	0
宜章县	洛洞村	洛洞村	431022150300112	210	0	0	0.000000	0.000000	0	0	0	杉木组	0	0	0
宜章县	洛洞村	洛洞村	431022150400056	111	33	23	16.000000	14.000000	10	1443	180	杉木组	198	0	0
宜章县	浆水乡	洛洞村	431022150400057	111	33	15	14.000000	10.000000	10	975	140	马尾松	80	0	0
宜章县	浆水乡	洛洞村	431022150400079	111	33	14	12.000000	10.000000	10	1950	150	阔叶松	23	0	0
宜章县	浆水乡	洛洞村	431022150400081	111	33	8	10.000000	8.000000	10	1410	150	阔叶松	13	0	0
宜章县	浆水乡	洛洞村	431022150400160	131	51	0	0.000000	1.700000	0	501	0	杉木组	0	0	0

图 15 数据获取界面

Fig. 15 Data collecting surface

3.2 测试结果

数据使用者获取到压缩包,解压后在 QGIS 中进行操作(图 16)。

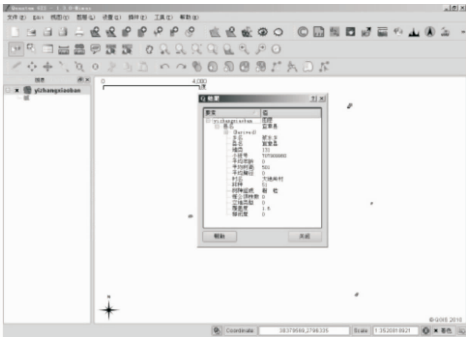


图 16 浆水乡 FML 小班模式数据

Fig. 16 Jiangshui FML subcompartment of mode data of iangsu Township

4 结论

本研究提出了一个多源分布式森林资源数据集成系统设计方案,方案探讨了多源分布式森林资源数据集成共享的设计思路、结构与功能。并通过对系统原型的实际测试,验证了此设计方案的可行性和实用性。该方案中主要是对多源分布式森林资源数据共享方式和内容进行研究,兼扩展到整个林业

领域。此设计中对空间地理数据共享研究还不足,如:三维数据、遥感数据和可视化统一标准。对此,将在以后的工作中进一步完善。本研究提出的系统设计和实践经验,可供类似数据集成系统作参考。

参考文献:

[1] 中华人民共和国中央人民政府门户网站. 林业局发全国林业信息化建设纲要(2008—2020 年)[EB/OL]. http://www.gov.cn/gzdt/2009-02/17/content_1233970.htm,2009.

[2] 张会儒,李春明. 森林资源信息共享中信息的分类与编码研究[J]. 西北林学院学报,2006,21(4):189-192.

ZHANG H R,LI C M. Classifying and coding of information in forest resource information sharing[J]. Journal of Noahwest Forestry University, 2006,21(4):189-192. (in Chinese)

[3] GeoServer[EB/OL]. <http://geoserver.org/display/GEOS/Welcome>,2009.

[4] QGIS[EB/OL]. <http://www.qgis.org>,2009.

[5] WILLIAM A R,FRANCIS X M,WILLIAM J B. 企业应用集成[M]. 北京:机械工业出版社,2003:20-23.

[6] 用 Restlet 创建面向资源的服务[EB/OL]. <http://www.infoq.com/cn/news/2008/05/restlet-for-restful-services>, 2009.

[7] OGC[EB/OL]. <http://www.opengeospatial.org>,2009.

[8] ISO19115:2003[EB/OL]. <http://www.iso.org>,2009.

[9] EPSG [EB/OL]. <http://www.epsg.org>,2009.

[10] Extract,transform,load[EB/OL]. http://en.wikipedia.org/wiki/Extract,_transform,_load.htm,2009.