

由图 1 可以看出,此平台主要包括 4 个层次:各部门分布式的异构数据层,元数据库数据层,Web services 层,Web 站点。其中,异构数据层分布于 Internet 网络的不同节点上,其数据管理和维护由各部门独立进行,这也是本文平台需要实现共享的数据源。

元数据库记录了各部门供公众共享的数据的元数据信息,每条元数据记录对应数据提供部门的一个相应数据集。各部门通过元数据编辑器把相应数据的元数据信息注册到平台的中心元数据库,实现元数据信息的发布。元数据库的建设是实现数据共享的前提。

Web services 是上海市资源与环境数据共享平台实现数据共享的关键技术,实现了对各部门异构数据的标准化处理。它可以标准化处理以文件形式和数据库形式存储的数据。各部门的服务器上部署相应 Web services,Web services 通过请求参数对相应数据进行处理,实现异构数据的标准化。

Web 站点是和终端公众交互的接口,是实现元数据库查询和 Web services 调用的一个桥梁。公众进入门户网站,查询元数据库,获取元数据信息,并发出请求调用相应的 Web services 标准化相应元数据描述的数据,最终下载 Web services 处理后的数据集,实现数据共享。

3 平台实现

上海市资源与环境数据共享平台的实现步骤如下:(1)构建一个集中的元数据库,各部门的元数据通过元数据编辑器被编辑并存入数据库;(2)在每个部门的服务器上部署 Web services 组件;(3)公众在门户网站中通过基于 Web 的元数据查询系统查询元数据;(4)具有权限级别的用户根据不同元数据的描述信息调用不同网络节点的 Web services 处理数据,最后下载数据资源。具体实现流程如图 2 所示。

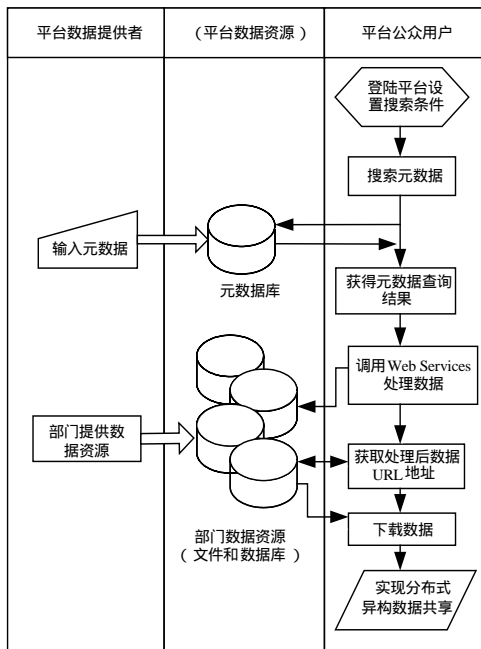


图 2 上海市资源与环境数据共享平台流程

3.1 元数据的注册

各数据提供部门通过元数据编辑器注册元数据,本文平台的元数据编辑器是基于 C/S 架构、利用 VB 开发的,元数据编辑人员根据需要选择元数据标准中的内容项,输入相应元数据信息后,通过 Internet 存储到数据库中。

平台元数据库软件采用 SQL Server 2000 数据库软件,最

主要的数据库是元数据信息表(T_METADATA_INFO),其中包括分布式异构数据的相关元数据信息。数据提供部门可以根据数据特点将相关信息通过元数据编辑器存入元数据库,其中分发信息描述了关于数据的物理地址等相关信息,是公众能否通过 Web 站点请求调用 Web services 进行数据处理并实现下载共享的关键信息。元数据表的分发信息字段具体描述如表 1 所示。

表 1 T_METADATA_INFO 表分发信息字段说明

字段名	数据类型	字段长度/bit	字段描述
fb_zyzxurl	varchar	80	网上资源定位符说明
fb_fxgs	varchar	100	分发数据的格式说明
fb_sjl	varchar	60	数据量
fb_fxbb	varchar	40	发行版本
fb_ccjz	varchar	100	资源存储介质说明
fb_sjkglrj	varchar	100	管理资源的软件信息
fb_sjkczt	text	16	管理资源的服务器信息

元数据分发信息填写是否正确,关系到此元数据描述的数据能否实现共享。其中,fb_zyzxurl(资源在线定位符)字段是说明分布式异构数据位置的信息,必须按标准格式要求填写。在本文系统中如果数据是文件型的,则格式如下:

文件(夹)名@服务器 IP

如果是数据库类型,则格式如下:

用户名:密码@主机 IP:端口:数据库名或实例名:表名

以上格式可根据需求更改和扩展,只要 Web 应用可以解析其信息并作为参数传输给相应的 Web services 实现相应操作。由于元数据的分发信息字段中出现了 IP、用户名及密码等相关敏感信息,因此为保证安全性,本文系统在数据通过元数据编辑器存入元数据库时进行了加密,且公众用户在 Web 站点无法查看相关信息的字段值。

3.2 Web services 的开发

Web services 是面向服务的技术,可以通过 Internet 进行应用程序到应用程序的通信,并提供不同环境下应用程序和设备的可访问性,Web services 实现了跨平台、跨操作系统、跨语言的解决方案^[6]。目前 Web services 技术有 2 大体系:Microsoft 公司的 .NET 和 Sun 公司的 Sun One。本文采用了 Sun 的 J2EE 技术体系。

本文平台的 Web services 实现了对各部门动态异构数据进行标准化处理,主要针对文件存储形式的数据和数据库存储形式的数据,对于不同形式的数据分别开发相应的 Web services,具体实现如下:

(1)对于文件存储形式数据,如果是压缩文件(.zip 文件或.rar 文件),则 Web services 直接将此文件复制到 service 容器的指定目录下(服务部署目录/download/);如果是非压缩文件,对需要共享的数据文件(单个文件或一个文件夹)压缩生成一个 zip 文件放置在 service 容器的指定目录下,Web services 返回此压缩文件网络可以访问的 URL 地址字符串,以下为程序主要代码框架:

```
//处理文件格式数据
Public class FileSharingservice
{
    //返回处理完的压缩文件名
    Public String getZipFileName()
    {
        ...
        String zipFileName = handleFile(file);
    }
}
```

```

if (zipFileName == null)
{
    return "";
} else
{
    return zipFileName;
}
}
//根据指定的请求参数压缩文件(文件夹), 返回一个压缩文件
Private File handleFile(String fileURL){
... } }

```

(2)对数据库存储形式的数据集, Web services 根据资源在线定位符参数连接数据库, 读取相应数据表(视图)的数据记录集字段值并写入一个动态生成的文本文件, 然后将此文本文件压缩生成一个.zip 文件并放在 service 部署容器的 Web 可访问目录下, Web services 返回该.zip 文件的 URL 地址字符串, 以下为程序主要代码框架:

```

//处理数据库存储的数据
Public class DBSharingService{
//返回数据库资源数据压缩后的文件名
public String getFileName(String resourceLocate, String dbType)
{
...
File zipFile;
zipFile = handleDB(resourceLocate, dbType);
return zipFile.getName();
...
}
//获取数据库数据读取到文件中, 并压缩此文件, 返回一个压缩文件
private File handleDB(String resourceURL, String dbType){
...
} }

```

3.3 Web services 的部署

Web 服务部署就是向服务请求者或服务注册中心发布服务接口和服务实现的定义, 并把 Web 服务的可执行文件部署到运行环境(典型情况下是 Web 应用程序服务器)中^[7]。

本文平台以 Apache Axis 为 Web services 引擎, Apache Axis 是一种开放源代码的实现方案, 为 Web 服务提供了一套完整的解决方案。Axis 提供了 2 种 Web 服务部署方式: 即时发布和定制发布。本文采用定制发布方式, 利用 Axis 提供的客户端管理工具——AdminClient 来完成服务的定制发布。

3.4 Web 站点的建设

Web 站点是基于网络的应用程序, 是整个平台与公众用户的交互端口, 它利用浏览器网页作为用户界面, 通过与用户的交互获取用户的请求参数, 根据用户输入条件, 检索元数据库并返回用户浏览器相应元数据结果集; Web 站点也是 Web services 的客户端, 根据请求参数的不同, 调用部署于不同网络节点终端上的 Web services 来实现服务的调用并进行数据处理, 最后实现数据下载。

平台的 Web 站点以 Tomcat 为 Servlet 容器, 采用 Jsp/Servlet+ JavaBean 方式实现。用户通过 jsp 页面输入元数据检索条件发送请求, Servlet 处理请求并转发给相应的 JavaBean, JavaBean 查询元数据库并返回结果记录集。用户根据元数据信息的详细描述, 向 Web 应用服务器发送下载请

求, Web 应用服务器根据元数据的参数信息, 建立 RMI 并向相应 Web services 发送调用请求, 部署在相应部门节点上的 Web services 根据请求参数对数据进行处理, 最终 Web services 返回数据文件的 URL 地址。Web 应用服务器将用户请求转发至数据文件的 URL 地址, 实现数据下载。整个流程如图 3 所示。

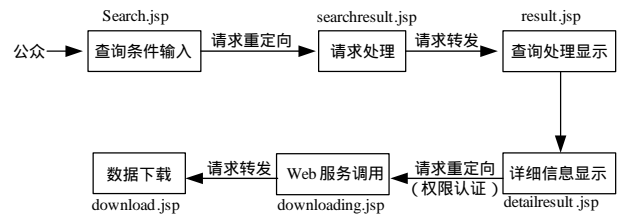


图 3 元数据检索和 Web services 调用流程

4 应用实例

上海市资源与环境数据共享平台基于元数据和 Web services 技术实现了包括上海市气象局、上海市地震局、上海市环境保护局以及中国极地研究中心 4 个部门数据的共享。各部门的数据都分布在各自的服务器节点上, 公众用户可以通过浏览器实现元数据的查询、浏览和数据下载。具体应用流程如下:

(1)用户打开浏览器输入本文平台 Web 站点网址(<http://202.96.202.181:8080/redata/index.jsp>)进入平台首页, 用户进行登录。

(2)进入元数据查询页面, 定制查询条件, 搜索已经在数据库中存储的元数据信息, 例如关键词为温度。

(3)符合搜索条件的元数据信息将以列表的形式显示在网页上, 通过点击查看详细资料可以查看元数据的详细信息。

(4)根据元数据的描述, 具有权限级别的用户可以对需要下载的数据, 点击下载数据按钮, 实现数据下载。

5 结束语

本文实现的平台在不影响各部门现有应用系统业务流程的前提下, 通过元数据的注册和 Web services 的开发与部署实现了数据共享及分布式异构数据的分布存储和集中获取, 效果良好。但目前此平台对各部门共享数据的处理相对简单, 地理空间数据共享处理问题、数据的集成和进一步挖掘等仍有待研究。

参考文献

- [1] 上海市公共研发服务平台[Z]. (2006-12-30). <http://db.sgst.cn/index1.jsp>.
- [2] 辜寄蓉, 苗放, 王成善. 基于元数据的信息共享机制研究[J]. 物探化探计算技术, 2006, 28(1): 76-79.
- [3] 石伟鹏, 杨小虎. 基于 SOAP 协议的 Web services 安全基础规范[J]. 计算机应用研究, 2003, 20(2): 100-105.
- [4] McIlraith S A, Son T C, Zeng Honglei. Semantic Web services[J]. IEEE Intelligent Systems, 2001, 16(2): 46-53.
- [5] 胡学骏, 曾凡智. 一种基于 XML Web services 的分布式解决方案[J]. 计算机工程, 2005, 31(13): 204-205.
- [6] 吴迪, 陈刚. 新一代的 Web services 技术[J]. 计算机应用研究, 2003, 20(3): 4-9.
- [7] 段江涛, 王保保. 基于 Web services 的应用系统构架研究[J]. 计算机仿真, 2005, 22(5): 259-271.