

# 基于元数据的江苏省水利数据资源目录服务系统

司存友<sup>1</sup>, 高祥涛<sup>1</sup>, 李政道<sup>2</sup>, 胡金龙<sup>1</sup>, 胡文斌<sup>1</sup>

(1. 江苏省水文水资源勘测局, 江苏 南京 210029; 2. 河海大学计算机与信息学院, 江苏 南京 210098)

**摘要:** 随着江苏省水利信息化建设进程加快, 各个水利部门积累了海量的数据资源, 如何实现跨部门跨平台的异构数据资源的整合与共享, 成为当前迫切需要解决的问题。在研究元数据模型的基础上, 提出了基于三级元数据的水利数据资源目录服务系统。系统运用 J2EE 技术、基于 web 服务方式设计并实现动态可配置的水利数据资源目录, 包含数据目录定制、元数据录入、数据下载、数据查询和访问统计等功能。

**关键词:** 水利数据资源; web 服务; 元数据; 数据目录; 资源共享

**中图分类号:** TP311.13 **文献标识码:** B **文章编号:** 1007-7839 (2018) 01-0041-05

## Metadata based directory service system for hydraulic data resource in Jiangsu Province

SI Cunyou<sup>1</sup>, GAO Xiangtao<sup>1</sup>, LI Zhengdao<sup>2</sup>, HU Jinlong<sup>1</sup>, HU Wenbin<sup>1</sup>

(1. *Jiangsu Hydrology and Water Resources Survey Bureau, Nanjing 210029, Jiangsu;*  
2. *School of computer and information, Hohai University, Nanjing 210098, Jiangsu*)

**Abstract:** With the acceleration of water information construction process in Jiangsu, various water resources departments have accumulated huge amounts of data resources. How to integrate and share heterogeneous data resources across departments and platforms has become an urgent problem to solve now. In this paper, a hydraulic data resource directory service system based on three level metadata was proposed based on the research of metadata model. The system used J2EE technology and web service mode to design and implement dynamically configurable water resources data directory, including data catalog customization, metadata entry, data download, data query and access statistics and other functions.

**Key words:** hydraulic data resource; web services; metadata; data directory; resource sharing

## 0 引言

江苏省水利信息化建设起步较早, 各业务部门先后建设了很多业务系统, 有力支撑了部门的各项管理工作。但各个系统之间相互独立, 拥有不同的运行环境和数据库系统, 不能很好地支撑各个部门之间业务上的关联, 数据共享和业务协同困难。为提高数据资源的利用效率、数据资源利

用价值, 同时提升业务系统的应用水平, 为水利业务管理提供高效决策支持, 需要在研究水利信息核心元数据基础上, 进行数据分类, 编制数据目录, 建立元数据库, 设计并实现水利信息资源目录系统, 整合分散的数据资源, 为数据共享、数据交换提供一个集成平台<sup>[1-2]</sup>。

## 1 总体设计

收稿日期: 2017-11-15

基金项目: 江苏省水利科技项目 (2015001)

作者简介: 司存友 (1978-), 男, 硕士, 高级工程师, 研究方向为水文水资源与水利信息化。

## 1.1 系统设计总体思路

依据元数据、Web Service、数据目录等技术<sup>[3-5]</sup>, 系统设计的总体思路如下:

(1) 基于水利核心元数据标准, 采用三级元数据模型, 形成“数据库—数据表—表字段和记录”的三级元数据体系, 并对水利核心元数据语义加以扩充, 加入表字段描述信息和数据来源信息, 支持数据的溯源能力, 保证数据的权威性和可信性;

(2) 采用数据目录与元数据结合的方式, 根据统一的标准规则对数据资源进行分类, 将数据资源整理形成有序组织, 从而形成水利信息资源目录, 方便用户对水利数据的定位、查询、下载与共享, 支持数据资源目录的动态可配置;

(3) 基于 Web Service 技术体系进行设计开发, 降低系统耦合性, 提高内部系统和外部系统调用的便捷度, 提高系统扩展性;

(4) 采用 MVC 设计模式设计系统, 将业务逻辑、数据、界面显示分离组织, 更好的支撑个性化定制及用户交互。

## 1.2 系统总体框架

系统基于 Web Service 技术体系、采用基于 J2EE 的 MVC 模式开发, 整体上由展示层、控制层和数据层等三层结构组成。位于系统最顶层的展示层为用户提供一个屏蔽系统差异的功能入口。该层通过应用生成的接口访问位于系统中间层的控制层。控制层通过提供基于元数据方面的操控,

对数据进行分析、抽取和使用。数据层则为整个系统提供数据支持, 位于整个系统的底层。系统的下载服务、元数据管理和访问统计服务都是基于 Web Service 开发的, 外部的 Web Service 服务可以注册到系统中, 并且被系统调用, 提高了扩展性。系统总架构如图 1。

## 2 元数据设计

### 2.1 元数据建模

根据《水利信息核心元数据》(SL 473-2010) 标准, 结合江苏省水利数据资源目录的建设需求, 元数据建模采用元数据三级应用模式。三级元数据具体包括:

(1) 一级元数据, 是顶层元数据, 即在数据库和文件集层面上进行描述的元数据, 表征基础数据库、监测数据库、业务数据库、空间数据库、多媒体数据库等特征;

(2) 二级元数据, 即在某类数据库中围绕某些相关对象及其业务活动的一组数据表或空间数据文件中某要素类的结构、特性的描述;

(3) 三级元数据, 即一般在管理、交换、发布等活动中, 按照一个批次进行处理的某要素类下的一组数据表的内容子集或空间数据文件中某要素类的一组要素的特征。

元数据三级应用模式具体如图 2 所示。

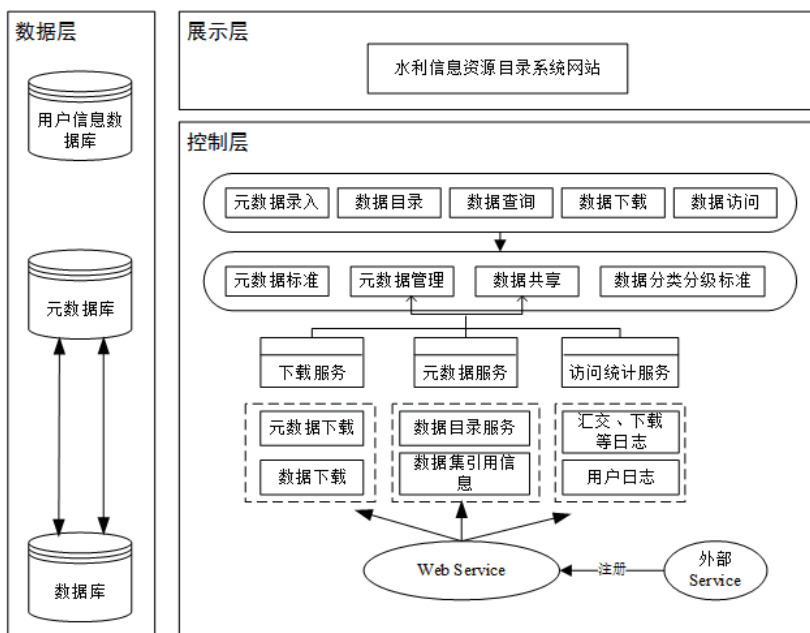


图 1 基于元数据的数据资源目录系统总体框架图

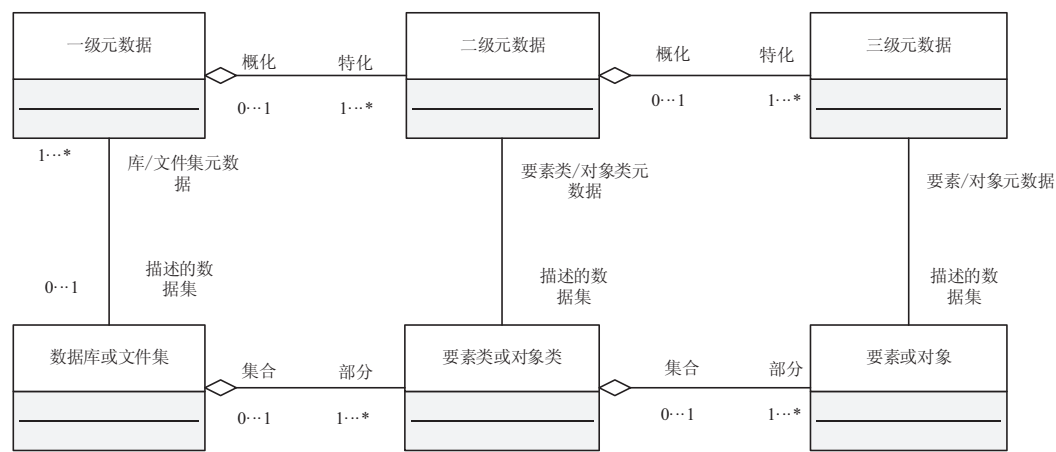


图 2 元数据三级应用模式图

2.2 元数据库设计

在上述三级元数据模式下，结合水利核心元数据，设计出三级元数据表。一级元数据表储存数据库的描述信息，其包含标识信息、内容信息、交换分发信息、维护信息、限制信息、覆盖范围信息等；二级元数据表储存数据表描述信息，其包含标识信息、维护信息、交换分发信息、限制信息表

等；三级元数据表储存数据表字段和记录的描述信息，其包含标识信息、维护信息、交换分发信息、限制信息等。具体元数据库对应表如表 1 所示。

3 系统功能设计与实现

3.1 基于的系统开发架构

系统基于 Web Service 开发，屏蔽了逻辑实现

表 1 元数据库对应表

表级别	表名	表标识符	表包含信息
一级元数据表	元数据信息表	MD_Metadata	元数据信息
	数据库级标识信息表	MD_PrimIdentifier	标识信息、内容信息、覆盖范围信息
	数据库级维护信息表	MD_PrimMaintain	维护信息
	数据库级交换分发信息表	MD_PrimExchDistr	交换分发信息
	数据库级限制信息表	MD_PrimConstraints	限制信息
二级元数据表	表级标识信息表	MD_SecIdentifier	标识信息、内容信息
	表级维护信息表	MD_SecMaintain	维护信息
	表级交换分发信息表	MD_SecExchDistr	交换分发信息
	表级限制信息表	MD_SecConstraints	限制信息
三级元数据表	记录级标识信息表	MD_TertIdentifier	标识信息、内容信息
	记录级维护信息表	MD_TertMaintain	维护信息
	记录级交换分发信息表	MD_TertExchDistr	交换分发信息
	记录级限制信息	MD_TertConstraints	限制信息
	字段级标识信息表	MD_ColIdentifier	标识信息、内容信息
	字段级维护信息表	MD_ColMaintain	维护信息
	字段级交换分发信息表	MD_ColExchDistr	交换分发信息
	字段级限制信息	MD_ColConstraints	限制信息

部分, 对服务使用者透明, 系统只关注输入参数和输出结果。基于 Web Service 的系统开发架构如图 3 所示。

系统由元数据录入、数据目录、数据查询、数

方式一适用于没有数据库或者数据库保密程度很高的单位, 直接进行人工填写录入, 不必进行数据源配置这一流程, 提交过后系统接收到用户提交的元数据信息, 会自动对元数据必填项进

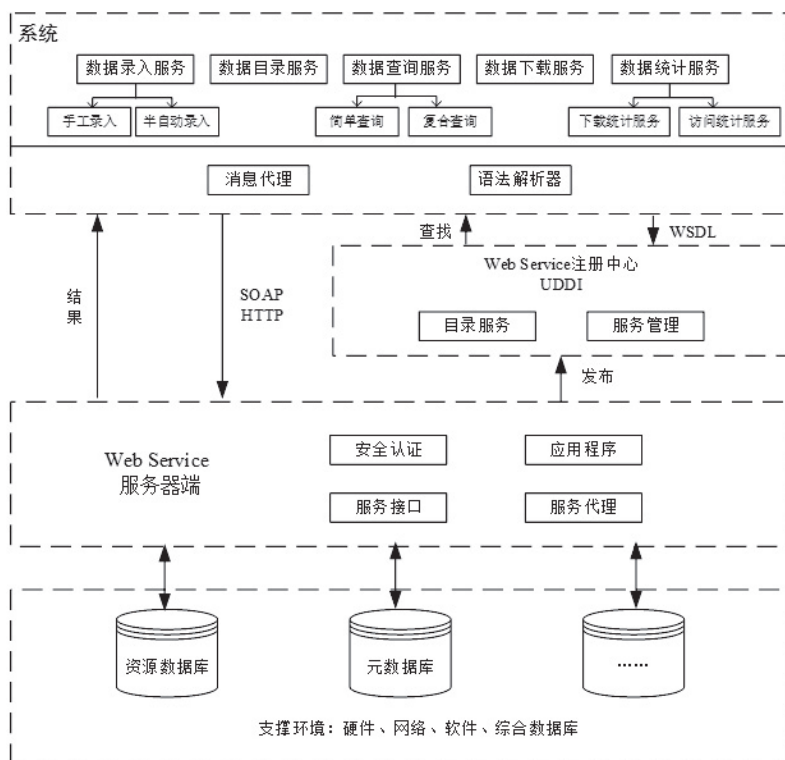


图3 基于 Web Service 的系统开发架构图

据下载和访问统计 5 个功能模块组成, 建立在底层硬件、软件、网络和综合数据库的支撑之上。在底层和系统应用层之间, 由 Web Service 起到桥梁作用, 将底层的数据和上层的请求加以逻辑处理, 简化了系统结构也降低了耦合性, 同时向外部系统提供了接口, 提高可扩展性。元数据录入划分为手工录入和半自动录入; 数据查询划分为简单查询和符合查询; 访问统计划分为下载统计和访问统计。

### 3.2 系统总体设计与实现

系统共有元数据录入、数据目录、数据查询、数据下载和访问统计 5 个功能模块组成, 系统总体功能框架如图 4 所示。系统实现界面如图 5 所示。

### 3.3 系统主要功能

#### (1) 数据录入模块

数据录入模块提供 2 种方式: 一是完全的手工录入; 二是在配置好数据源后通过内置程序萃取元数据信息, 在用户对必要的信息进行完善后提交进行元数据注册。

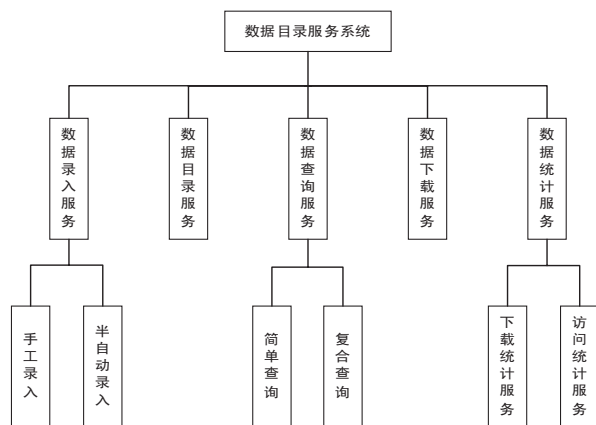


图4 系统功能框架图

行验证。若验证通过, 则存入系统中等待管理员审核; 若验证未通过, 则系统定位到第一个出错的元数据项, 并给出错误提示, 提醒用户修改。成功录入后会在页面显示刚刚录入的信息以便用户核对查错。

方式二便捷度较高, 只需部分人工参与。系统要求首先配置好需要注册的数据库, 配置成功后,



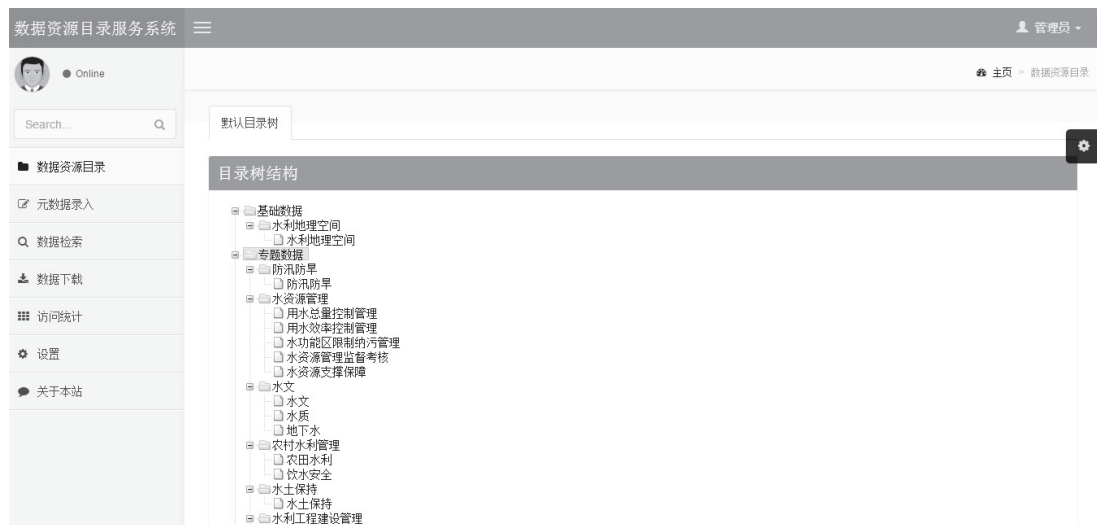


图 5 系统实现界面图

后台自动萃取该数据库的元数据信息并保存到临时元数据库中, 同时将萃取到的数据信息显示给用户, 供用户补充并核对。对于抽取不到的数据项, 需要用户手工填写。为便于区分, 系统对必填项和选填项进行标注, 必填项数据不完整时, 用户必须将信息补全后, 才可进入后续流程。

无论是方式一还是方式二, 都将繁琐的工作交给系统处理, 最大化地给用户提供了便利。各用户单位只需要提供一个数据库只读用户, 即可萃取出元数据信息。系统支持 Oracle、Mysql、SQLServer 等主流数据库。

#### (2) 数据目录模块

数据目录模块提供 2 种操作方式。一是使用系统默认生成的目录树结构进行数据资源浏览等操作; 二是用户自己定制目录树分类标准, 生成个性化的目录树结构再加以操作。

方式一中, 目录树结构模块会自动根据元数据库内信息按照默认分类层级生成一棵资源目录树。用户可以对目录树进行点击展开直到最后一个层级。在对最后一个层级对象点击时, 目录检索结果模块会显示出此对象对应的元数据信息。用户可以对元数据信息进行详细查看、数据预览、数据下载等操作。

方式二中, 用户首先对分类层级定制, 最多可以定制 4 个层级。定制层级后点击生成目录树则在目录树结构模块生成一棵个性定制的目录树。

#### (3) 数据查询模块

数据查询模块提供简单查询、复合等查询方式。

简单查询是最基本的查询方式, 只要求用户输入一个检索词, 系统根据这个检索词对摘要、关键词等进行全文匹配查询。无论摘要, 关键词, 还是标题中出现该检索词均认为该条记录符合检索要求, 将其作为检索结果返回。

复合查询缩小了数据的查询范围, 实现对查询数据结果的快速准确定位, 提高查询效率, 同时, 也为了满足不同用户的查询需要。复合查询为用户提供更多的检索条件, 用以更加精确地检索出元数据。选择数据库字段名称进行检索包括元数据名称、关键词和摘要等; 每种分类可以输入不同的检索词进行多重查询。

#### (4) 数据下载模块

对于公开和在用户下载权限内的数据资源, 在数据下载模块可以使用数据下载功能在线浏览或下载该数据资源; 而对于不公开和受用户权限限制不能访问或下载该数据, 用户可以通过填写数据申请表向管理员或数据所在单位申请该数据。通过数据下载申请授权后, 用户可通过管理员或数据所在单位提供的下载方式在共享服务系统直接下载。

#### (5) 数据统计模块

访问统计服务功能实现信息交换过程中对系统操作信息和各个节点交换的业务数据信息的记录和展现。如在线数据下载统计、共享资源服务访问统计等。用户通过限定时间范围, 选择展现方式(柱状图、饼图等)即可搜索出结果, 并将结

(下转第 50 页)

（上接第 45 页）

果按照选定方式展现出来。

## 4 结语

元数据模型基于水利核心元数据标准、扩充元数据语义进行设计,实现了采用三级元数据结构全面地描述数据。系统采用 MVC 架构、Java EE 技术和 B/S 模式设计开发,满足跨平台要求。在技术体系中采用了 AJAX、Mybatis、及 Spring MVC 等核心技术,在保证技术先进性的同时兼顾了技术的实用性。采用数据访问对象(Data Access Object)来实现对数据库的存取,采用异步任务来处理长时间请求,采用 O-R MAPPING 技术保证公共数据库的可扩展性;采用基于 Web Service 的方式开发接口,使得在构建各种各样的系统中的服务时可以采用一种统一和通用的交互方式,从而实现数据共享。通过实践运行表明,水利数据资

源目录服务系统提高了水利数据资源的组织、共享、管理和统计能力,提高了水利部门的工作效率。

## 参考文献:

- [1] 程益联,付静.水利数据整合共享研究[J].水利信息化,2014(6):13-17.
- [2] 朱跃龙,许峰,冯钧,等.水利信息资源目录体系构建研究[J].水利信息化,2010(02):4-8.
- [3] 成建国,冯钧,杨鹏,等.水利数据资源目录服务关键技术研究[J].水利信息化,2014(06):18-21+35.
- [4] 李萍.基于WebService的数据共享交换平台的设计与实现[J].计算机时代,2016(07):34-37+41.
- [5] 陆佳民,冯钧,唐志贤,等.水利大数据目录服务与资源共享关键技术研究[J].水利信息化,2017(4):17-20,27.