

全国地质资料目录服务中心总体设计研究

连 健¹ 王黔驹¹ 颜世强^{1,2}

(1. 中国地质调查局发展研究中心, 北京 100037; 全国地质资料馆, 北京 100037; 2. 北京大学地球与空间科学学院, 北京 100871)

摘要: 本研究以实现全国地质资料目录数据共享服务为目标, 在现有地质资料管理与服务体系架构基础上, 全面了解省级馆藏机构、地调系统资料馆、委托保管单位资料馆及行业资料馆的现有地质资料目录管理与服务现状, 兼容成果、原始、实物三大类地质资料目录, 进行全国地质资料目录服务中心系统的总体设计研究, 涵盖总体技术路线、功能框架以及数据库架构等, 并指出了需要解决的关键问题, 以期能够动态汇聚各部门地质资料目录数据资源, 并以多元的服务方式对公众提供服务。

关键词: 地质资料; 目录; 服务系统; 总体设计

中图分类号: P963 **文献标志码:** A **文章编号:** 1000-3657(2013)05-1675-08

地质资料是地质工作形成的重要基础信息资源, 具有可被重复开发利用、能够长期提供服务的重要功能。新中国成立 60 多年来, 中国形成了海量的地质资料信息, 这些资料数据大部分保存在全国地质资料馆、各省级地质资料馆、委托保管单位资料馆以及各行业地勘单位资料馆等^[1]。中国从 20 世纪 80 年代就开始研建地质资料目录数据库^[2], 迄今大部分地质资料馆藏机构都建立起了目录检索系统并对外提供服务^[2-5]; 此外, 在参考国际国内相关元数据和共享服务标准^[6-8], 以及借鉴多个行业的良好示范应用^[9-12]基础上, 地质资料目录数据的共享与服务在技术层面也有了初步的进展^[13-15]。分析全国地质资料管理与服务的现状可知^[16], 目前全国地质资料保管分散, 提供地质资料目录信息是最普遍、最基础的服务方式。面对国家与国土资源部提出的找矿突破战略行动以及地质资料信息服务集群化产业化的重大需求^[17], 研究并解决应用计算机信息技术、海量数据库存储技术、分布式传输访问技术、地理信息系统技术等实现地质资料目录集群及检索服务, 构建可

互联互通的、分布式共享的全国地质资料目录服务中心, 向政府部门、企事业单位、社会大众提供目录信息检索成为当务之急。

本文详细分析了全国地质资料管理与服务现状, 提出了全国地质资料目录服务中心建设的思路、原则、总体框架、技术路线和功能设计, 探讨了需要解决的关键技术, 对于推进地质资料信息服务集群化产业化具有重要意义。

1 中国地质资料管理及服务现状

1.1 地质资料服务机构及馆藏情况

当前中国地质资料主要分布于 4 大类系统: 一是部省两级地质资料馆; 二是委托保管单位资料馆; 三是地调系统资料馆; 四是行业地勘单位资料馆(室)。其中, 部、省两级地质资料馆共 33 个, 包含全国地质资料馆和国土资源实物地质资料中心两个国家馆以及 31 个省级馆; 委托保管单位资料馆是指国土资源部授权进行原始和实物地质资料委托保管的 35 家油气、海洋地质资料管理部门; 地调系统资料

收稿日期: 2012-12-07; 改回日期: 2013-06-25

基金项目: 中国地质调查局地质矿产调查评价专项“地质资料信息服务集群化产业化综合研究项目”(1212011220335)资助。

作者简介: 连健, 男, 1983 年生, 博士, 从事地质资料管理、服务及信息化研究; E-mail: lianjian2001@163.com。

馆主要为 6 大地调中心资料管理部门;行业地勘单位资料馆包括地矿、有色、冶金、煤炭等行业系统地勘单位的地质资料馆(室)。这 4 类馆藏机构收整、保存了中国大约 90% 以上的成果地质资料以及相当数量的实物和原始地质资料,截止到 2011 年底,全国部省两级地质资料馆藏机构成果地质资料的馆藏总量达 402704 种,以矿产地质、环境地质和水文地质三类资料为最多,占全部汇交成果地质资料的 87.2%。

1.2 地质资料管理与服务现状

从地质资料的保管现状看,由于历史原因,中国地质资料数据分散,从属于不同的馆藏机构和行业部门,缺乏互联互通,集成化程度低;从地质资料的服务利用看,呈“各自为战”的局面,难以提供有效的、集群式的检索服务,况且服务及产品体系尚不健全,缺少应用网络技术所建立的现代化管理与检索平台,尚不能满足快速增长的资源保障、环境保护、工程建设和防灾减灾等工作需求,难以满足公众对地质资料信息的了解、获取等需求。

当前,地质资料的服务渠道主要有网络服务和窗口服务两种,服务模式以网络检索查询、数据下载、窗口复制加工、重点工程专题服务和社会民生应急支撑等为主。2011 年,全国各级地质资料馆藏机构通过传统服务窗口为 4 万多人次提供了地质资料服务,提供地质资料利用达 19 万份次(332.4 万件次),同时,网络服务能力日趋增强,全国地质资料服务网站点击总量超过 400 万次,网络服务方式逐渐占据地质资料服务的主导地位。在地质资料服务中,目录数据服务处于主体地位,大多数的馆藏机构都通过电子阅览室及网络提供目录查询检索服务。

1.3 地质资料目录数据资源现状

目前,全国部省两级地质资料馆、委托保管单位资料馆、地调系统资料馆均建立了目录数据库,研建部署了目录服务系统,开展了窗口和网络目录服务。调研可知,已有的全国性的目录数据资源包括:①成果地质资料案卷级目录数据库;②成果地质资料涉密清理信息数据库(相当于成果地质资料文件级目录库);③实物地质资料目录数据库;④石油天然气原始、实物地质资料目录数据库(文件级);⑤图文数字化形成的地质资料文件级题名资源等。已有的地质资料目录著录、服务的软件或系统有全国成果地质资料目录著录软件、成果地质资料涉密清理登记管理信息系统、全国地质资料业务管理信息系统、地

质资料联合编目与目录服务系统、石油天然气地质资料委托管理系统。

全国地质资料目录服务中心的建设要以这些数据资源为基础,循序渐进,做到案卷级目录与文件级目录数据资源的联动查询,并以此带动地质资料数字资源的公开共享。这将有助于解决各部门地质资料目录数据标准不统一、集群程度低、地质资料目录数据共享服务机制不健全、地质资料目录管理与服务体系滞后等问题,更好地开展服务。

2 地质资料目录服务中心总体设计

2.1 总体设计原则

全国地质资料目录服务中心要以面向公众服务(尤其是检索服务)为宗旨,建立联动式共享服务机制,搭建高效、快捷、灵活、实用的地质资料目录集群服务框架,实现地质资料由传统的信息汇集型向现代的集群服务型转变,遵循以下原则:

(1)数据分布、服务统一原则:结合地质资料目录数据量大、位置分散等特点,结合数据“谁拥有、谁管理、谁发布、谁授权”的原则,通过分布式服务技术手段,搭建逻辑上统一的数据服务平台,使用户可以通过一站式的服务门户,实现对主中心和若干分中心的统一查询浏览和下载定制服务,达到“各中心自由检索,依资源所在地不同开展服务”的目标。

(2)易部署、易管理、易维护原则:分布式应用系统的部署、管理和维护工作不仅包括全国地质资料馆,还涉及各省级地质资料馆、地调系统资料馆、委托保管单位资料馆、行业系统资料馆等,位置分散,工作量大。因此,系统研建应综合应用各种手段,提高对不同软硬件运行环境的适应能力,降低对部署人员的技术水平要求;要提供简洁、易用的系统管理工具,便于对系统用户、数据、运行环境的实时监控和管理。

(3)快速服务、多层次服务原则:对于不可以在网络上服务的数据资料,以目录形式提供服务;对于可以公开进行服务的数据,开展目录服务的同时,提供数据在线浏览和数据下载服务;对于需要申请才能服务的数据,提高数据申请定制需求,从而实现多层次服务。在服务渠道上,结合网上服务和资料馆服务,通过网上服务可以定位资料位置,通过资料馆服务,可以获取网络无法直接公开的数据资料,从而实现多渠道服务,能够使用户快速地查询浏览各类地

质资料。

(4)数据可扩展、功能可拓展原则:目录服务中心建设不但应考虑现势数据的管理,更应该考虑下一步地质资料相关业务的发展趋势。为保证新增加的数据能够方便地进入系统并对外服务,系统研建应保持可扩展性和可配置性,不但能够将现有的各类数据进行管理和查询,而且可以方便地将新增的数据类型纳入管理范围,并可进行定制化的服务。

(5)标准为先、安全为要原则:没有标准规范作为指导,就难以开展资料数据的整理和入库工作,也难以实现地质资料数据统一化、规范化服务。因此目录服务中心建设必须标准先行,以标准来规范数据整合、系统开发和系统上线后的服务工作。同时,系统在数据的发布审核、申请流程及后台管理上面要充分考虑,保障安全。

2.2 目录服务中心系统总体架构

全国地质资料目录服务中心的总体架构可以概括为“两级体系,五层架构”。按照全国地质资料管理与服务职能体系的不同,针对地质资料信息服务互联互通、集群共享的总体要求,全国地质资料目录服

务中心系统由主中心和分中心两级体系构成。主中心(全国地质资料馆)与分中心(各级馆藏机构)互联互通,共同对外提供不同层次的数据。

全国地质资料目录服务中心从总体上可分为五层架构(图 1):

(1)数据层:包括馆藏机构数据、委托机构数据、地调系统数据等多源异构的地质资料目录数据库,以及相关地质资料共享数据;

(2)支撑层:位于数据层与应用层之间,包括两方面,一是分布式数据集中与同步技术等技术支持,二是实现多源异构地质资料目录数据集成的工具集;

(3)应用层:构建数据集中与发布、多元检索服务、专题产品生成等多元多维度的目录数据管理与服务功能;

(4)门户层:全国地质资料目录服务中心系统是以主中心为总出口,以 HTML 页面和 Web Service 接口实现客户端与服务端系统之间交互操作;

(5)外围运管层:建立运行机制,研究标准规范,构筑安全体系,确保系统的顺畅运行。

全国地质资料目录服务中心系统需在全国地质

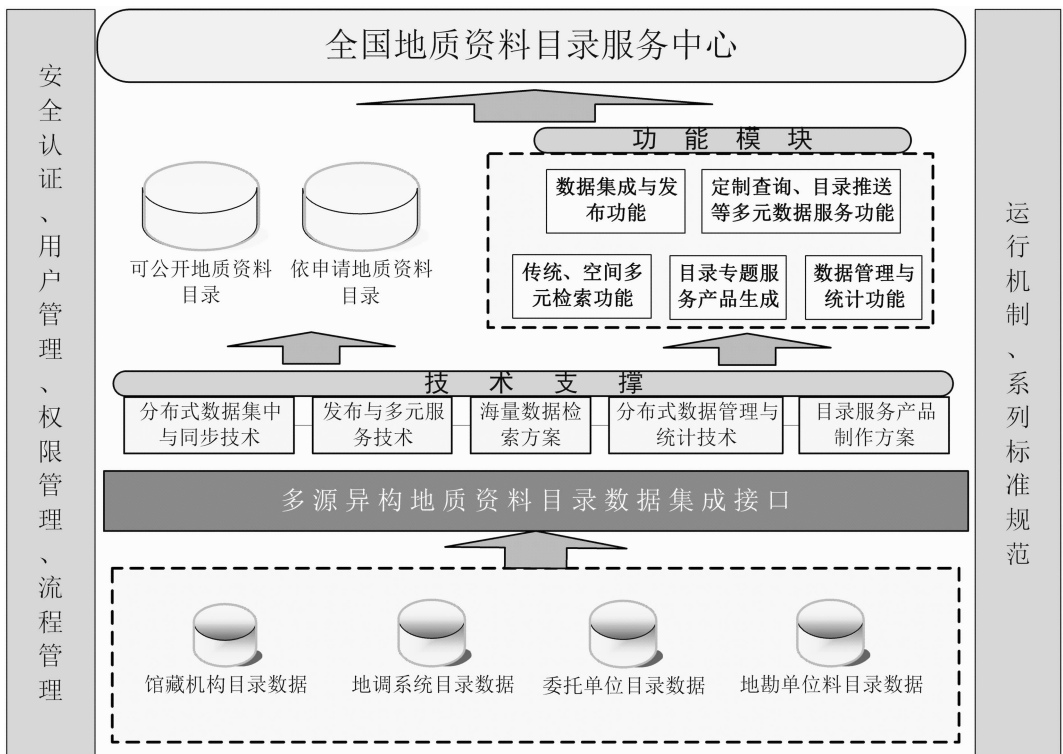


图 1 系统体系结构图

Fig.1 The architecture design of the system

资料馆、各省级地质资料馆、委托保管单位资料馆、地调系统资料馆及其他行业系统资料馆部署和应用,业务逻辑模式如图 2 所示。在目录数据管理上,馆藏单位分布式管理、更新和授权,通过服务的方式对外发布并统一集成到主中心的系统上,并与各分中心保持目录数据同步;在服务管理上,以主中心的目录服务系统为总出口,统一对外提供多种层次的地质资料目录服务。

2.3 目录服务中心建设总体技术路线

在既有研究基础上,对现有地质资料目录数据资源开展全面梳理,收集并了解现有馆藏、委托保管、地调、行业等地质资料目录数据资源的类型、特点、保存、管理及服务现状;明确用户分类,调研用户需求,开展地质资料目录服务中心的总体设计;研究建立地质资料数据共享与服务的运行保障机制;建立并完善目录中心建设需要的相关标准规范,进行关键技术研究;开展目录集成数据库建设,针对不同目录数据来源,开发相应的数据接口,在目录数据集集成整合基础上,扩充、完善系统功能,形成多元的目录数据服务,提供多形式的目录服务产品,实现地质资料目录数据的在线及离线服务。总体技术路线流

程如图 3 所示。

2.4 目录服务中心系统功能设计

全国地质资料目录服务中心提供给六大区地调中心、全国各省馆、委托馆、行业地质资料馆依据自身配备的系统与全国地质资料馆进行联合编目,并通过网络进行信息交换与动态更新;全国各类地质资料馆藏机构通过地质资料目录服务中心系统开展目录检索服务。

目录服务中心整体的功能架构如图 4 所示:

(1) 目录数据的联合编目

通过著录、套录和数据导入方式,实现案卷级目录与文件级目录的联合编目。

(2) 目录数据的更新与发布

目录数据的更新与发布主要包括目录对等发现、目录更新上传和目录审核发布三个功能模块。其中,对等策略主要针对节点群采取间歇式异步嗅探技术,使分布式集群架构与数据存储中心保持一致并提高检索效率;目录更新上传功能用来完成单节点下新增目录数据通过网络向中心节点的更新上传;目录审核发布功能用于各节点,著录、上传或接收到目录的审核与发布。

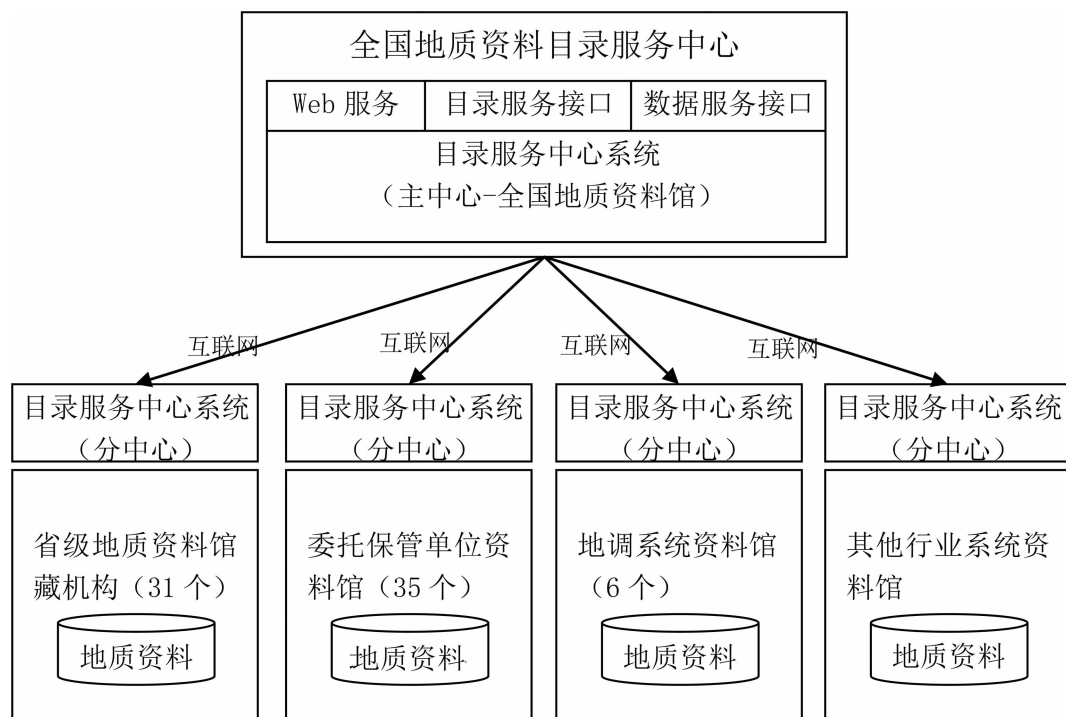


图 2 目录服务中心系统业务逻辑图

Fig.2 The logic diagram of the service center system

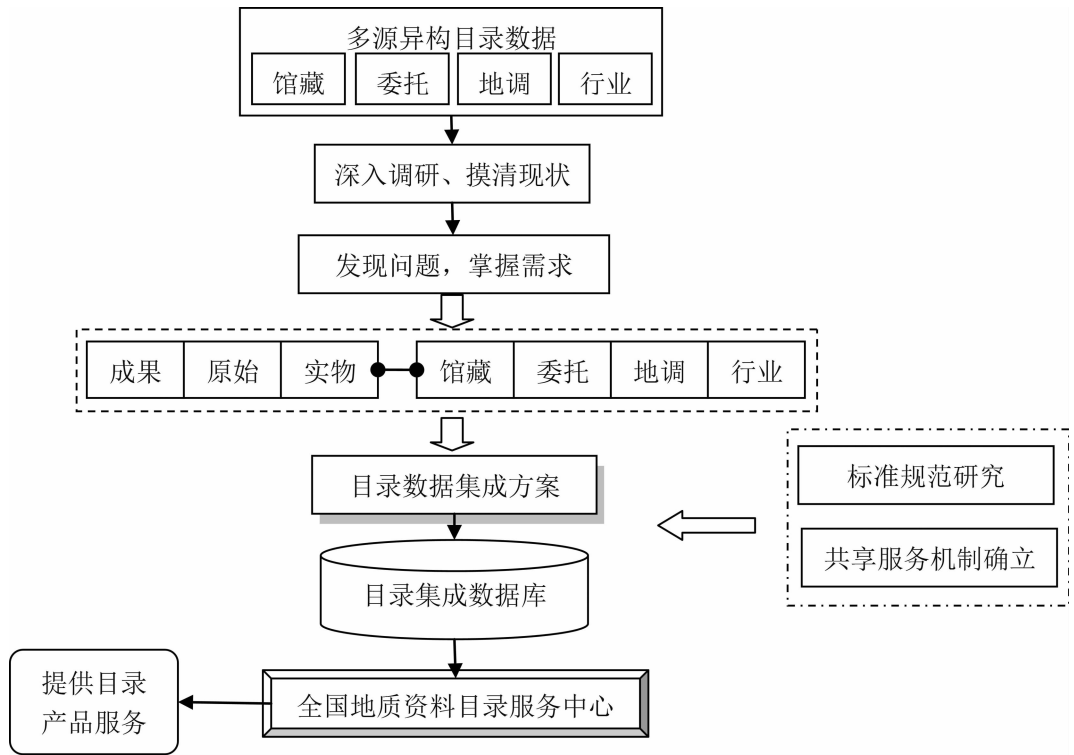


图 3 总体技术路线流程图
Fig.3 The overall technical route chart

(3) 目录数据的查询检索

针对目录数据的查询检索分为传统查询、空间查询和其他数据检索三种方式。其中,传统查询可分为简单查询、模糊查询和高级查询几种情况;而空间查询可进行数图联动,反映地质资料目录关联工作区域所在的位置;另外,可通过分布式的检索方式针对其他地质资料数据进行检索。

(4) 目录数据的统计分析

从所属馆藏机构、工作程度、时间序列等方面,针对目录数据开展专题统计分析,汇总统计数据,并发布最新上传的目录数据。

(5) 目录数据的服务功能

针对汇聚的目录数据资源,根据用户的需求,开展多方式的服务提供,包括接受用户定制查询服务,以及根据用户需求,定制目录推送服务等。

3 目录服务中心系统数据库架构

按照全国地质资料目录服务中心两级体系架构设计,系统数据库架构也划分为两级,即以全国地质资料馆为主中心,以各省馆、委托保管单位资料部门、地调系统资料管理部门、行业部门馆藏机构等为分中

心,开展数据组织管理。其详细架构如图 5 所示。

主中心数据主要包括目录汇集区,通过目录共享服务网关连接、管理各个分中心,并按照一定的标准,对各个分中心的目录数据进行抽取、审核和集成,支撑对外的目录服务,由公开服务目录汇集库和依申请公开目录汇集库组成。同时,主中心还将存储用户身份数据、依申请服务和数据产品服务请求数据等。

4 目录服务中心建设需要解决的关键问题

在全国地质资料目录服务中心建设过程中,需要面临和解决的关键问题有以下几个:

4.1 标准规范的研究与制定

根据中国标准化法的规定和有关标准化文件要求,全国地质资料目录服务中心在系统设计、建设、运行和应用过程中,除了执行既有的国家标准和行业标准之外,仍需要研究与制定一系列标准规范,主要包括:地质资料目录数据资源唯一标识符规范、地质资料目录数据访问服务接口规范、全国地质资料目录服务中心核心元数据规范、地质资料目录数据

跨域互操作技术规范、地质资料目录数据资源加工指导规范等。

4.2 相关关键技术的研究

(1) 分布式环境下高可靠目录同步技术

在分布式环境下,要实现主中心和分中心的目录同步,涉及部省两级地质资料馆、委托保管单位资料馆、地调系统资料馆以及行业地勘单位资料馆(室)。由于各个分节点的运行环境、网络情况各不相同,同步过程中常常会出现各类异常情况,导致主分中心目录同步出现失败。针对这些问题,考虑引入“基于时间点目录同步”技术、“自适应分批同步”技术、“重复数据识别”技术、“XML 数据压缩”技术等来提高目录同步稳定性和效率。

(2) 基于并行、缓存、超时失败的高效分布式检索技术

在覆盖主中心和分中心的分布式系统下,保障系统的稳定性和服务效能是一个巨大的挑战。这里提到的系统的稳定性是指如何保证整体系统不会因为一个分节点的问题而停止服务。这里提到的服务效能是指部分节点由于网络或其他因素导致服务效率或性能下降,如何还能够保障系统的响应速度不受严重影响。针对这一问题,系统将通过“并行检索”

技术、“历史查询缓存”技术、“TopN 反馈”技术、“匹配阈值过滤”技术等来提高系统的性能。

(3) 多源异构数据资源的兼容整合技术

地质资料目录数据的管理软件和保管格式随着采用软件的不同而不同,这就造成了目录数据格式有较大的差异,此外由于目录数据著录软件的不同,也导致相应数据库结构的不同,更为目录数据融合造成障碍。针对这一问题,采用中间件技术,研究不同目录数据的库结构及其存储格式,定制通用的目录数据转换中间件,从而为多源目录数据的转换、存储与发布提供便利。

(4) 数据资源的空间检索查询机制

传统的目录数据查询检索是通过关键字或主题词,通过简单检索或高级检索的方式实现目录数据的列表查询。新形势下对于地质资料空间位置分布的需求越来越多,通过空间上的检索查询来获取目录数据及其分布位置成为一个不可或缺的检索方式。采用嵌入式底层接口服务模式来达到并完成空间检索,通过接口服务来嵌入业已发布的空间信息检索系统,通过参数的交换与传递,来实现目录资料发现、明码或压缩流传输、数据聚类、关系表存储、多协议跨域松散对等、分类树重构、多元检索、分页定

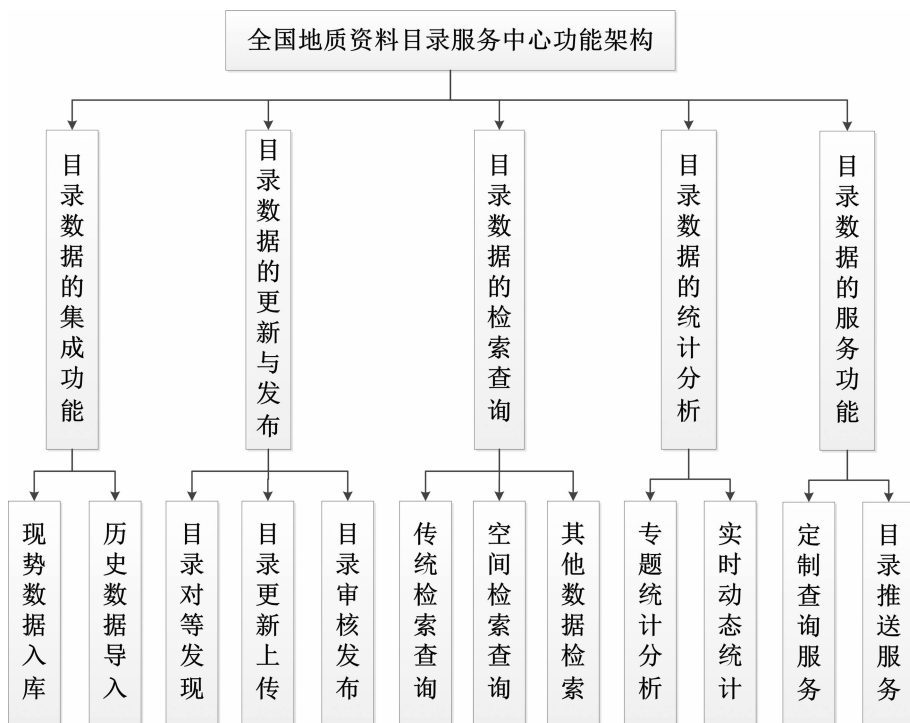


图 4 目录服务中心功能架构图

Fig.4 The function structure diagram of the service center system

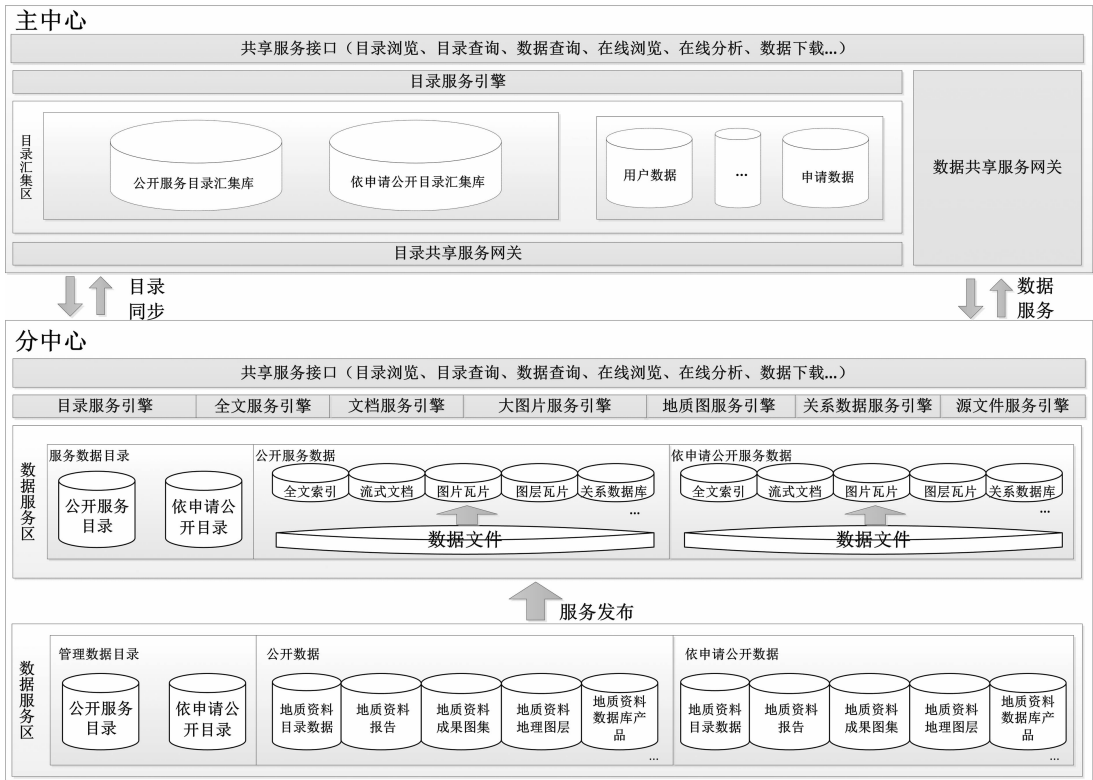


图 5 系统数据库架构设计图

Fig.5 The architecture design diagram of the system database

制推送、依比例尺权重地质体属性编码解析、大型地质图栅格影像切片服务等功能。

5 总结

本文通过对目前地质资料管理与服务现状的调研,分析目录资料管理与服务的现状,在此基础上,研究相关关键技术,探讨架构设计,制定功能方案,完成全国地质资料目录服务中心的总体设计研究,充分发挥资料目录的功用,以期达到更好地为社会大众服务的目的。

参考文献(References):

[1] 贾文珏, 申世亮, 周舟. 地质资料信息集群化共享服务平台构架设计[J]. 国土资源信息化, 2011, 6: 3-7.
 Jia Wenjue, Shen Shiliang, Zhou Zhou. Architecture design of shared services platform and clustering information of geological data [J]. Land and Resources Informatization, 2011, 6: 3-7 (in Chinese with English abstract).

[2] 王春宁, 肖世伟, 尚武, 等. 全国地质资料目录数据库系统建设[J]. 国土资源信息化, 2004, 1: 12-15.
 Wang Chunning, Xiao Shiwei, Shang Wu, et al. The construction of geological data catalog database system of China [J]. Land and

Resources Informatization, 2004, 1:12-15(in Chinese with English abstract).

[3] 温雪茹, 翟国平, 李银罗. 地质资料目录检索服务系统的建设[J]. 国土资源信息化, 2007, 5:10-12.
 Wen Xueru, Zhai Guoping, Li Yinluo. Construction of catalog retrieval system of geological data [J]. Land and Resources Informatization, 2007, 5:10-12(in Chinese with English abstract).

[4] 冯钰. 浅谈成果地质资料目录数据库建立及利用[J]. 内蒙古地质, 2002, 3:35-37.
 Feng Meng. Talking about development and utilization of data base system on contents of geological report [J]. Geology of Inner Mongolia, 2002, 3:35-37(in Chinese with English abstract).

[5] 卜小平, 张翠光, 赵亚利. 全国地质资料数据中心建设方案初探[J]. 国土资源科技管理, 2007, 24(6):127-131.
 Bo Xiaoping, Zhang Cuiguang, Zhao Yali. A discussion on construction scheme of state center for geological documents and data [J]. Scientific and Technological Management of Land and Resources, 2007, 24(6): 127-131(in Chinese with English abstract).

[6] TC 211. Revised Changes to the ISO/DIS 19115: Geographic Information Metadata, 2005.

[7] 何建邦. 地理信息国家标准手册 [M]. 北京: 中国标准出版社, 2004.
 He Jianbang. National Geographic Information Standards Manual [M]. Beijing:China Standard Press, 2004(in Chinese).

- [8] FGDC. Content standard for digital geospatial metadata, 1998.
- [9] 逯永光, 丁孝忠, 李廷栋, 等. “OneGeology 计划”及其在中国研究新进展[J]. 中国地质, 2011, 38(3):799-808.
- Lu Yongguang, Ding Xiaozhong, Li Tingdong, et al. The 'One Geology Project' and its research progress in China [J]. Geology in China, 2011, 38(3):799-808(in Chinese with English abstract).
- [10] 罗雁, 邹梦秋, 王刚, 等. 城市空间信息服务共享平台研究与应用[J]. 城市勘测, 2012, 5:20-24.
- Luo Yan, Zou Mengqiu, Wang Gang, et al. The reasearch and application of urban spatial Information service platform [J]. Urban Geotechnical Investigation & Surveying, 2012, 5: 20 -24 (in Chinese with English abstract).
- [11] 赵静, 史天运, 李平. 铁路信息共享平台中共享数据库结构和关键技术研究[J]. 铁路计算机应用, 2010, 9: 1-3.
- Zhao Jing, Shi Tianyun, Li Ping. Study on structure and key technologies for railway information sharing platform of sharing database [J]. Railway Computer Application, 2012, 9:1 -3 (in Chinese with English abstract).
- [12] 池天河, 王雷, 王钦敏, 等. 数字省信息共享平台的设计与实现[J]. 地理研究, 2003, 3:281-288.
- Chi Tianhe, Wang Lei, Wang Qinmin, et al. Design and realization of Information sharing platform for digital province [J]. Geographical Research, 2003, 3:281-288(in Chinese with English abstract).
- [13] 吴晓红, 王杨刚, 高志新. 基于元数据的地质调查网站群系统研究[J]. 中国地质, 2011, 38(5):1405-1411.
- Wu Xiaohong, Wang Yanggang, Gao Zhixin. Research on the geological survey website group based on metadata [J]. Geology in China, 2011, 38(5):1405-1411(in Chinese with English abstract).
- [14] 周春磊. 基于 ASP.NET 技术的全国地质资料目录服务系统[J]. 国土资源信息化, 2005, 1:24-26.
- Zhou Chunlei. ASP.NET technology based service system of the national geological data catalogue [J]. Land and Resources Informatization, 2005, 1:24-26(in Chinese with English abstract).
- [15] 汪宙峰, 何政伟, 李敏, 等. 面向服务的地质灾害空间数据共享技术研究[J]. 长江科学院院报, 2011, 3: 72-74.
- Wang Zhoufeng, He Zhengwei, Li Min, et al. Service-oriented spatial data sharing technology of geo-hazard assessment[J]. Journal of Yangtze River Scientific Research Institute, 2011, 3: 72-74(in Chinese with English abstract).
- [16] 王黔驹, 颜世强, 王永生, 等. 全国地质资料馆藏机构现状、问题与对策建议[J]. 中国国土资源经济, 2011, 1:18-21.
- Wang Qianju, Yan Shiqiang, Wang Yongsheng, et al. The current situation, problems and countermeasures of national geological data collection agencies[J]. Natural Resource Economics of China, 2011, 1: 18-21(in Chinese with English abstract).
- [17] 王黔驹, 颜世强, 丁克永, 等. 推进地质资料信息服务集群化产业化的初步思路与构想[J]. 中国矿业, 2011, 12:20-23.
- Wang Qianju, Yan Shiqiang, Ding Keyong, et al. Preliminary ideas on the promotion of the clustering and industrialization of geodata information service [J]. China Mining Magazine, 2011, 12: 20-23 (in Chinese with English abstract).

Research on the architectural design of the National Geological Data Catalog Service Center

LIAN Jian¹, WANG Qian-ju¹, YAN Shi-qiang^{1,2}

(1. Development and Research Center, China Geological Survey, Beijing 100037, China; 2. School of Earth and Space Sciences, Peking University, Beijing 100871, China)

Abstract: In order to achieve the target of the national geological data sharing services, this paper based on the existing geological data management and service architecture, first analyzed the current situation of the geological data management and service in the geological data collection agencies, and then designed and dealt with the architecture of national geological data catalog service center system. The authors provide the overall technology roadmap, functional framework, database schema, and point out the key issues that need to be addressed, in the hope that, with the research work, the geological data resources from various agencies will be integrated, and diverse service styles will be offered to the public.

Key words: geological data; catalog; service center; architectural design

About the first author: LIAN Jian, male, born in 1983, doctor, engages in the study of geological data management and geological information analysis; E-mail: lianjian2001@163.com.