引用格式:Wang Liangxu,Nan Zhuotong,Ge Jinsong,*et al*. Design and Application of Data Center for Eco-environment Protection in the Qinghai Lake Basin[J]. Remote Sensing Technology and Application,2013,28(1):166-172.[王亮绪,南卓铜,葛劲松,等.青海湖流域生态环境综合数据平台的设计与应用[J].遥感技术与应用,2013,28(1):166-172.]

青海湖流域生态环境综合数据平台的设计与应用

王亮绪[1],南卓铜[1],葛劲松[2],吴立宗[1],聂学敏[2],鲁子豫[2]
(1.中国科学院寒区旱区环境与工程研究所,甘肃 兰州 730000;
2.青海省生态环境遥感监测中心,青海 西宁 810001)

摘要:国家发改委《青海湖流域生态环境保护与综合治理规划》自2008年正式启动以来,有力地推动了青海湖流域的生态环境保护,积累了丰富的生态环境数据。青海湖流域生态环境综合数据平台是生态监测体系中的数据集成中心,为发掘生态环境监测数据的科学价值,辅助生态环境变化过程中的科学研究,充分发挥青海湖流域生态环境监测体系建设项目的投资效益,并面向青海湖流域的生态环境研究提供科学数据服务。平台依托地理元数据技术,集成了开源GeoNetwork元数据系统,构建了一个集在线元数据编辑、数据可视化、数据汇交、管理、发布和共享等功能于一体的WEB系统,支持缩略图导航、分类导航、时空导航、关键词导航等多种数据导航方式,定义了一套快速数据服务体系,实现了从数据汇交、数据发布到数据共享的业务流程。该系统已初步实现了9个青海湖生态环境监测单位的数据整理与汇交,数据量约为200 GB。

关 键 词:青海湖;生态环境;数据共享;元数据;GeoNetwork
中图分类号:TP 79 **文献标志码**:A **文章编号**:1004-0323(2013)01-0166-07

1 引 言

青海湖是我国最大的内陆高原咸水湖,是维系青藏高原东北部生态安全的重要水体,也是控制西部荒漠化向东蔓延的天然屏障。根据国家发改委《青海湖流域生态环境保护与综合治理规划》,青海省于2008年启动了青海湖流域生态环境监测体系建设,有力地推动了青海湖流域的生态环境保护。青海湖流域生态环境监测体系建设的目标是及时掌握流域生态环境现状和变化规律,为流域资源、环境、人口的管理和经济发展提供重要决策依据。目前该监测体系已积累了丰富的生态环境数据,迫切需要建立一个数据集成平台,实现数据的长期安全存储、高效管理、资源共享和及时发布,并为政府决策提供数据支持。

数据共享已在科学界取得共识,如科技部组织的科学数据共享工程[1]、中国科学院启动的科学数据库[2]、国家自然科学基金委在中国西部环境与生态科学研究计划中设立的中国西部环境和生态科学数据中心[3,4]等一批重大工程和项目,明确体现了政府和科学界在数据共享方面的重视。在流域尺度上也开展了很多数据共享研究,如黄河流域[5]、黑河流域[6]都开展了数据共享工作,青海湖流域也曾在鸟类迁徙方面开展过数据挖掘工作[7],但还没有建立一个完整的以数据共享为目标的青海湖流域科学数据平台。

青海湖流域生态环境综合数据平台(以下简称数据平台,Data Center for Eco-Environment Protection in the Qinghai Lake Basin,DEEP)于2011年启动,由青海省生态环境遥感监测中心组织实施,

收稿日期:2012-05-24;**修订日期**:2012-11-22
基金项目:国家自然科学基金项目"面向黑河流域生态—水文过程集成研究的数据整理与服务"(91025001),青海湖流域生态环境数据库项目资助。
作者简介:王亮绪(1976-),男,江苏连云港人,高级工程师,主要从事科学数据共享研究。E-mail:wangliangxu@lzb.ac.cn。

中国科学寒区旱区环境工程研究所负责技术协作,青海省环境监测中心站、青海省草原总站、青海省水文水资源勘测局、青海省林业调查规划院、青海湖国家级自然保护区管理局、青海省气象科学研究所、青海湖裸鲤救护中心和青海省水土保持局 8 个单位合作建立的。目前已经完成了数据平台的汇交、管理、发布和共享功能,实现了青海湖流域已有生态监测数据的入库管理,于 2012 年 7 月开始对外发布数据(http://deep.qherc.org)。本文将系统地介绍青海湖流域生态环境综合数据平台的总体框架、系统设计、数据资源和共享服务,并对该平台的未来建设进行讨论。期望更多对青海湖流域感兴趣的科学家和研究者了解和使用数据平台成果,促进青海湖流域的发展和保护。

2　总体框架

2.1　总体目标

青海湖流域生态环境综合数据平台的目标是依托数据库、互联网和地理信息系统技术,发掘生态环境监测数据的科学价值,辅助生态环境变化过程中的科学问题研究,充分发挥青海湖流域生态环境监测体系建设项目的投资效益。

数据平台的主要任务包括整理现有青海湖流域各类生态环境数据,撰写对应元数据和数据文档,实现数据和元数据的汇交、管理、发布、共享、可视化系统,为今后青海湖流域生态环境数据入库提供支持。

2.2　系统框架

数据平台的总体框架如图 1 所示,以文件形式和数据库形式保存的数据为核心,以对数据描述的元数据为数据平台交互的桥梁,通过元数据实现数据汇交、管理、发布、共享与服务的所有功能,并以这些功能服务于不同的目标用户群。目标用户群分为 3 类:普通用户、数据生产用户以及数据平台管理者。普通用户可以查看所有的元数据信息并进行数据服务,数据生产用户(包括青海省环境监测中心站、青海省林业调查规划院、青海省草原总站、青海省水保局、青海省水文局、青海湖国家自然保护区管理局、青海省气象科学研究所、青海湖裸鲤救护中心和青海省生态环境遥感监测中心等 9 家数据生产单位)可以进行数据汇交与管理,管理者(即青海省生态环境遥感监测中心)负责数据审核与发布。

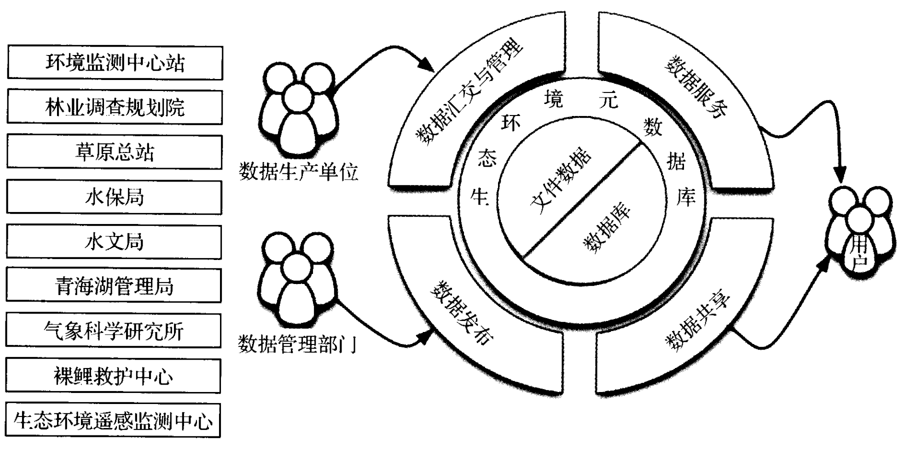


图 1　系统框架

Fig. 1　System architecture of Data Center of Eco-environment Protection in the Qinghai Lake basin (DEEP)

开源技术在地学数据中心得到了广泛应用[8],本数据平台也采用了大量的开源软件作为系统的基础框架(图 2),在数据存贮底层同时支持文件和数据库,文件服务器采用 Pureftp,数据库服务器采用 PostgreSQL,Web 服务器采用 Apache2,服务器端脚本采用 PHP5,应用层采用 PHP5 进行自主开发以构建数据平台的核心功能,同时数据平台也集成并扩展了开源的 GeoNetwork 作为元数据编辑器。

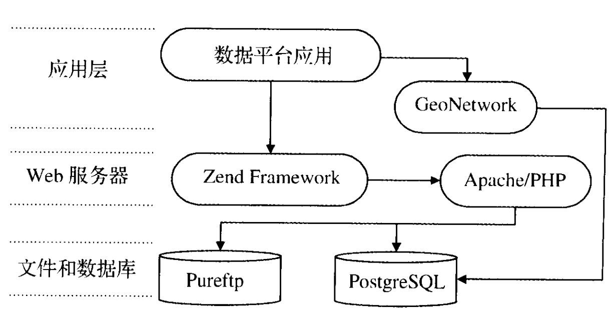


图 2　软件架构

Fig. 2　Implementation details of DEEP

2.3 元数据标准

生态环境监测数据来源丰富、格式多变、类型复杂，为支持所有的生态环境数据支持，在数据平台中就必须使用元数据技术。元数据在数据平台发挥着重要作用，是联接数据用户与数据作者之间的重要中介，也是数据平台进行系统实现时的一个重要桥梁。数据平台选择国际元数据标准ISO 19115作为数据描述标准。针对青海湖流域生态环境监测系统的数据特点，本系统对ISO 19115标准进行了简单定制，如对数据标识符采用UUID的形式来定义，对数据的服务方式以数据分发要素的在线链接或离线获取作为判断标准等，集成了在线元数据撰写系统，提高了元数据的撰写效率。

2.4 数据服务方式

数据平台提供2种数据服务方式：在线数据下载和离线数据申请。其中，在线数据是公开的数据，用户可以快速获取；离线数据则由于政策原因不能直接提供服务，申请的周期也比较长，离线数据的提供者需要对用户的申请进行审核。数据平台所有的数据服务都是通过FTP来提供服务的，在线下载或离线申请通过后，数据平台都会给一个FTP的帐号来获取其申请的数据，用户需要使用FTP工具来获取对应的数据资源。数据服务流程如图3所示，用户在数据平台上通过导航或搜索查找到数据后，查看其感兴趣数据的详细信息，以确定其是否是自己所需要的数据，确认后即可根据不同的数据类型进行对应的数据服务方式，即在线数据或离线数据，在线数据则在填写完对应的用途信息即可获取对应的FTP账号信息，离线数据则还需要打印邮寄对应的申请表给数据平台的服务组，数据平台收到用户的申请后，邀请数据资源的提供者开展对用户申请的审核，审核通过后即通过邮件通知用户进行FTP下载。

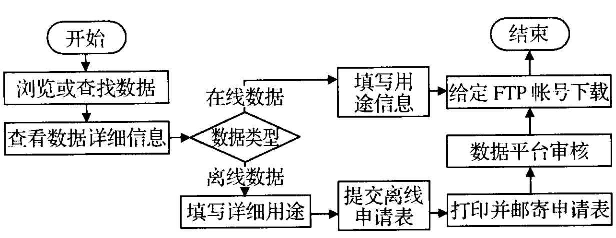


图3 数据服务流程

Fig. 3 Workflow of data services in DEEP

3 系统实现

数据平台中实现的功能主要包括青海湖流域生态环境数据的汇交、管理、发布、审核、发现、共享和反馈等功能。数据生产部门负责数据汇交和管理，数据汇交到数据管理部门后由管理部门进行质量审核，在质量满足要求的情况下即可对外发布，发布后用户即可通过数据功能实现搜索、查看、下载、申请和反馈，数据管理部门收到用户的离线申请后，就可以进行离线的申请审核，各功能之间的相互关系如图4所示。

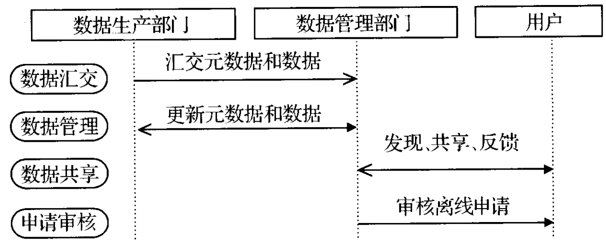


图4 系统功能流程图

Fig. 4 Functionalities offered by DEEP and their interactions

3.1 数据汇交

数据平台的核心是数据，而元数据是数据平台开展共享服务不可或缺的一个部分。因此要实现青海湖流域生态环境数据汇交系统，就应该同时支持新元数据的撰写以及数据的上传汇交。数据平台在实现数据汇交功能时，采用集成GeoNetwork系统的方法来实现元数据的创建和编辑。进行数据汇交时，首先要创建元数据，可以通过元数据模版或已有元数据进行创建，元数据撰写完成后再根据系统提供的FTP账号实现对应数据的上传，最后提交评审发布。管理者收到数据生产用户的数据评审申请后就可以开展数据审核工作，审核完成即可对外发布。

3.2 数据管理

数据平台必须保护数据提供者的权益，因此在数据管理系统的设计与实现过程中，数据提供者的权益保护是首先考虑的要素。用户在申请离线数据时，数据服务组收到用户的纸质申请后即发送申请审核邮件给数据提供者，数据服务组收到数据提供者的反馈后才能开始离线数据服务。同时，数据平台还提供了管理元数据、管理数据反馈、管理数据文档、管理数据新闻、管理数据版本的功能，以方便数据提供者更好地管理自己的数据。

3.3 数据发现和共享

随着应用功能的完善，数据平台将会发布越来越多的数据资源。如何保证让用户在众多的数据资源中尽快发现自己所需要的数据，是数据平台必须提供的一个核心功能。因此，良好的数据发现功能，

能够让用户快速地找到数据并很好地使用数据,是提升数据平台服务质量的一个重要途径。数据平台利用ISO 19115元数据的核心要素,实现了特色的数据发现功能,包括根据数据的不同关键词(包括时间、地点、主题和学科关键词)进行导航、根据数据集类别进行导航、根据数据的缩略图进行导航等方式,同时还提供了不同的浏览方式,如时间轴浏览方式、时空浏览方式、列表浏览方式,还可以根据数据提供者的生产单位浏览不同部门提供的青海湖流域生态环境数据。数据平台在每个页面都提供了搜索功能,可以搜索元数据中的标题和摘要,方便用户更快地找到自己所需的数据。

用户找到数据后,即可查看数据的详细信息(图5),展示了元数据中的核心要素,同时数据信息页面还提供了一定的数据关联功能。如相关的数据、网络资源等。用户才根据数据的服务方式来进行服务,在线数据可直接下载,离线数据则需要进行申请并通过审核才能服务。

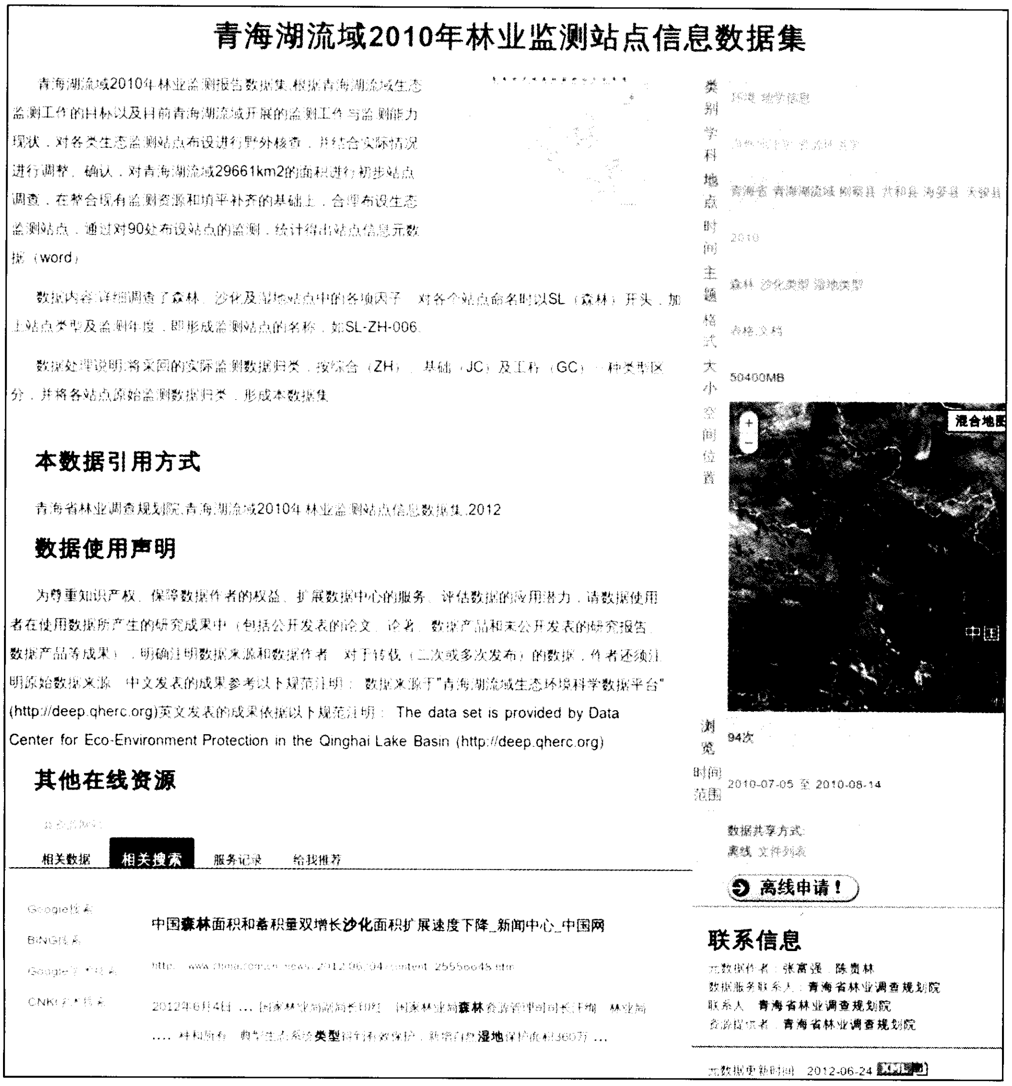


图5　数据平台元数据查看页面示例

Fig. 5　Data detail page in DEEP

3.4　数据审核与发布

数据管理部门通过数据平台管理所有的数据,对汇交数据进行质量审核和发布,对用户的离线申请进行服务。数据只有发布后才能被用户查找、发现和服务。用户提交的离线申请首先要由数据管理部门进行初次审核,然后再发送审核邮件给数据提供者,最后汇总数据提供者的意见反馈给用户。离线数据的审核流程如图6所示,数据平台收到申请后将从元数据的联系人信息中提取合适的邮件地址作为数据提供者,发送邀请数据审核的邮件,然后根据数据提供者的反馈来决定是否通过此次申请。

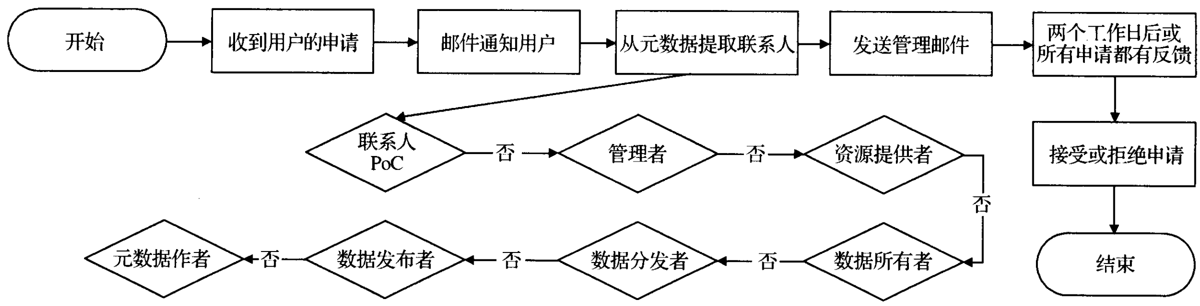
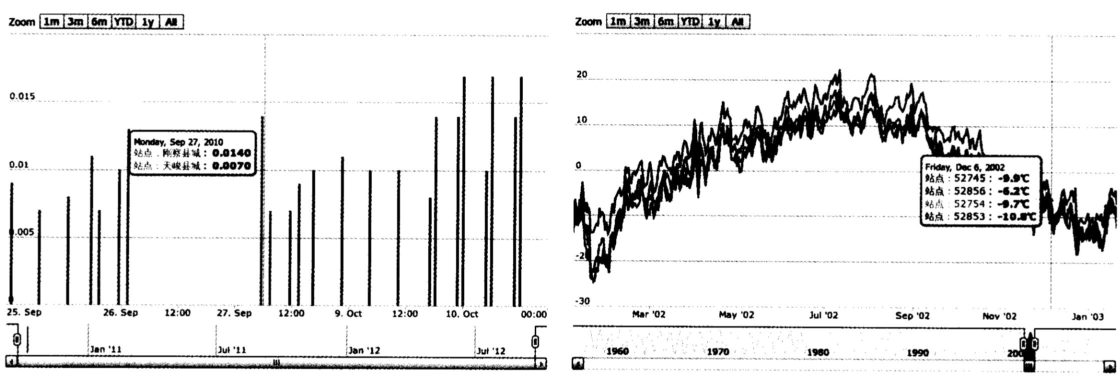


图 6 数据离线申请的审核流程

Fig. 6 Workflow of dealing with offline application in DEEP

3.5 数据可视化

数据可视化将数据从表格形式转变为支持时间序列的动态交互图像，方便用户更好地了解数据，发现其中的规律。数据平台目前提供了气象数据、水文数据和环境监测数据的可视化功能。数据平台的数据可视化系统利用 Highcharts JS 客户端脚本结合数据库功能而实现。气象数据和水文数据通过动态曲线的方式进行可视化，环境监测数据通过柱状图的形式进行可视化，如图 7 所示，所有的可视化都支持时间系列的选择，也可以查询某一个时刻的值。



(a) SO_2 (b)日平均温度

图 7 数据可视化示例

Fig. 7 An example of data visualization in DEEP

4 数据资源

数据是数据平台的核心，而数据的数量和质量决定了数据中心的成效。目前，数据平台已整理数据集 76 条，数据量约 200 GB，共涵盖 9 家青海湖流域相关单位，各单位已整理数据情况见表 1。其中，青海省生态环境遥感监测中心主要整理了基础地理数据和遥感数据；青海省林业调查规划院提供了

表 1 已整理数据集统计表

Table 1 Statistics of published datasets classified by organization in DEEP

数据生产单位	元数据条数	数据量/MB
青海省水土保持局生态环境监测总站	2	43.29
青海湖裸鲤救护中心	5	10.58
青海省林业调查规划院	5	50 577.80
青海湖国家级自然保护区管理局	7	647.60
青海省环境监测中心站	7	5.25
青海省水文水资源勘测局	8	13.65
青海省草原总站	8	4.55
青海省气象科学研究所	12	5 318.06
青海省生态环境遥感监测中心	22	143 670.00

2010年90个监测站点的林业调查资料，包含大量的野外观测照片和实地调查薄；青海省气象科学研究所整理了自1957年以来的青海湖流域逐日、逐月、逐年气象观测资料，覆盖时间长；青海省环境监测中心站整理了2010、2011年的青海湖流域环境空气、地表水水质、生活饮用水水质和土壤环境质量数据；青海省裸鲤救护中心整理了青海湖裸鲤卵苗分布、裸鲤生存环境监测、裸鲤资源蕴藏量、渔业生态环境和水生生物调查等数据；青海湖国家自然保护区管理局整理了鸟类调查、水禽及普氏原羚栖息地植被生物量、生物多样性、普氏原羚监测、科学考察等数据；青海省水土保持局整理了土壤侵蚀数据及相关历史资料；青海省草原总站整理了2005～2010年的草地生态监测数据集；青海省水文局整理了1957～2009年日降水量、蒸发、流量数据。

5 结 语

本文介绍了青海湖流域生态环境综合数据平台的总体框架、系统实现以及目前数据平台在可用数据集方面的初步进展。在数据平台系统方面，实现了一个有自身特色的科学数据平台：①一体化的数据共享业务流程，从数据汇交、发布、共享到管理各个方面；②支持多种特色的数据导航方式，得益于ISO 19115元数据标准，数据平台在元数据的基础上可以实现不同的数据导航，如根据数据的缩略图来查看所有数据、根据ISO 19115元数据标准中定义的分类信息进行数据导航、根据时间信息和空间信息查看处于不同时间范围、空间范围的可视化数据导航方式、根据ISO 19115元数据标准中定义的各种关键词信息进行数据导航(如基于主题关键词、地点关键词、学科关键词等)；③特色的离线数据申请审核系统，在离线数据申请流程中尊重数据生产者的权益和意愿，体现了数据平台作为数据用户和数据所有者的桥梁作用。在数据平台的核心数据上，已整理收集了青海湖流域1957～2011年的包括气象、水文、水保、林业、草原、裸鲤资源、环境监测、自然保护区等方面的数据资源，也将逐步更新、引入更多的青海湖流域生态环境监测方面的科学数据。

数据平台的下一步工作包括：①与国内其他生态环境科学数据中心，如中国西部环境与生态科学数据中心，开展互连互通工作，提高数据平台知名度，吸引更多科学研究人员参与到青海湖的生态环境保护；②数据平台不仅要提供数据，还应开展数据之上的扩展服务，如在线的模型服务、在线的数据处理功能、更丰富的数据可视化功能等。

青海湖流域生态环境环境科学数据平台是对青海湖流域长期生态环境监测的一次信息化总结和示范，是对青海湖流域生态环境保护的一次数据总结，为今后青海湖流域生态环境保护和治理工作奠定了良好的数据基础。

致谢：青海湖流域生态环境综合数据平台在执行过程中得到了项目各参与单位的大力支持，盖春梅、李建轩等参加了不同阶段的工作。在此致谢！

参考文献(References)：

[1] Zhang Xian'en. China National Scientific Data Sharing Program[J]. Scientific Chinese, 2004, (9): 11-13. [张先恩. 国家科学数据共享工程[J]. 科学中国人, 2004, (9): 11-13.]

[2] Gui Wenzhuang. The 20-year Development of CAS Scientific Database[J]. Bulletion of the Chinese Academy of Sciences, 2007, (1): 87-89, 91. [桂文庄. 迎接科学数据库发展的新阶段——中国科学院科学数据库发展20年的回顾与思考[J]. 中国科学院院刊, 2007, (1): 87-89, 91.]

[3] Li X, Nan Z T, Cheng G D, *et al*. Toward an Improved Data Stewardship and Service for Environmental and Ecological Science Data in West China[J]. International Journal of Digital Earth, 2011, 4(4): 347-359.

[4] Li Xin, Nan Zhuotong, Wu Lizong, *et al*. Environmental and Ecological Science Data Center for West China: Integration and Sharing of Environmental and Ecological Data[J]. Advances in Earth Science, 2008, 23(6): 628-637. [李新, 南卓铜, 吴立宗, 等. 中国西部环境与生态科学数据中心：面向西部环境与生态科学的数据集成与共享[J]. 地球科学进展, 2008, 23(6): 628-637.]

[5] Qin Yaochen, Qin Fen, Li Shuang. Research and Practice of the Earth System Science Data Sharing in the Yellow River Basin[J]. Yellow River Civilization and Sustainable Development, 2008, 1(2): 168-180. [秦耀辰, 秦奋, 李爽. 黄河流域地球系统科学数据共享研究与实践[J]. 黄河文明与可持续发展, 2008, 1(2): 168-180.]

[6] Li Xin, Wu Lizong, Ma Mingguo, *et al*. Digital Heihe River Basin. 2: Data Integration[J]. Advances in Earth Science, 2010, 25(3): 306-316. [李新, 吴立宗, 马明国, 等. 数字黑河的思考与实践2: 数据集成[J]. 地球科学进展, 2010, 25(3): 306-316.]

[7] Zhang Haiting, Tang Mingjie, Zhou Yuanchun, *et al*. Design and Implementation of a Data Mining System for Qinghai Lake Bird Migration[J]. Computer Systems & Applications, 2009, (11): 1-6. [张海亭, 唐明洁, 周园春, 等. 青海湖鸟类迁徙的数据挖掘系统设计与实现[J]. 计算机系统应用, 2009, (11): 1-6.]

[8] Wang Liangxu, Wu Lizong, Nan Zhuotong, *et al*. Application of Open Source Technologies in Geoscientific Data Centers[J]. China Science & Technology Resources Review, 2010, 42(3): 17-23, 35. [王亮绪，吴立宗，南卓铜，等. 开源技术在地球科学数据中心中的应用[J]. 中国科技资源导刊，2010，42(3)：17-23，35.]

Design and Application of Data Center for Eco-environment Protection in the Qinghai Lake Basin

Wang Liangxu[1], Nan Zhuotong[1], Ge Jinsong[2], Wu Lizong[1], Nie Xuemin[2], Lu Ziyu[2]

(1. *Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences*, *Lanzhou* 730000, *China*;
2. *Qinghai Ecological and Environmental Remote Sensing Monitor Center*, *Xining* 810001, *China*)

Abstract: Ecological and environmental data in the Qinghai Lake Basin (QLB) have been systematically collected since the official launch of the Plan of Eco-Environmental Protection and Integrated Management of the Qinghai Lake Basin in 2008. The Plan has established a comprehensive monitoring system for ecology and environment studies in the QLB. The Data Center for Eco-Environment Protection in QLB (DEEP) is a data portal of the established monitoring system in QLB, aims which to collect, manage and disseminate ecological and environmental data to facilitate scientific studies in QLB. DEEP is built upon a geographic metadata architecture that is actually supported by the open source GeoNetwork, by which creation and edition of geographic metadata is implemented. DEEP offers full data sharing functionalities, including metadata editing, data collecting, auditing, managing and publishing, as well as data visualization currently only for partial datasets. At present, DEEP holds 76 metadata and a total size of ecological and environmental datasets of around 200 GB from9 participating organizations. The functional design and key technologies are presented in this paper.

Key words: Qinghai Lake Basin; Eco-environment; Data sharing; Metadata; GeoNetwork