

# 东北亚资源环境综合科学考察数据集成体系的构建

王卷乐<sup>1</sup>, 宋佳<sup>1</sup>, 朱立君<sup>1,2</sup>

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101;

2. 中国科学院研究生院, 北京 100049)

**摘要:** 东北亚区域是一个资源相对集中, 生态环境格局复杂, 气候地带性多样, 人地关系对比显著的区域。集成该区域的资源环境综合科学考察数据, 对于支撑全球变化等前沿科学研究和区域可持续发展等具有重要意义。针对该区域数据资源缺乏的现状, 研究构建了包括三大体系的东北亚资源环境综合科学考察数据集成系统, 即数据资源采集与管理标准规范体系、数据分类体系、数据应用平台软件体系。其中, 标准规范体系包括数据采集规范、数据管理规范 and 综合考察规范 3 类共 23 项; 数据分类体系包括 4 个大类、25 个小类、128 个要素; 数据平台软件系统包括数据目录检索管理、元数据管理、数据发布与可视化、数据浏览与获取等 5 个功能模块; 鉴此, 用地理信息技术和网络技术, 实现了东北亚资源环境综合科学考察数据集成原型系统。目前, 已集成 144 个数据集(库), 数据量超过 2TB, 为东北亚综合科学考察数据的深度集成和应用奠定基础。

**关键词:** 资源环境; 综合科学考察; 数据体系; 东北亚区域

**DOI:** 10. 3724/SP. J. 1047. 2012. 00074

## 1 引言

东北亚区域(本文系指我国黄河以北的东北、华北和西北地区, 蒙古全境, 俄罗斯西伯利亚和远东地区)是一个资源相对集中, 生态环境格局复杂, 气候地带性多样, 人地关系显著的区域<sup>[1]</sup>。该区域内的自然资源、生态环境与人类活动等具有典型的梯度变化特点。例如, 年平均温度从高于 20° 到低于 -20°; 年平均降水量从 2000mm 到 100mm; 每平方公里人数从大于 1000 人到小于 10 人; 土地利用从集约化程度非常高到人类活动干预非常少等。这些区域特征对于研究全球变化在该区域的响应、自身的可持续发展研究等具有重要意义。国际上包括美国、德国, 以及亚洲的日本、韩国等国家长期在该区域开展了科学研究合作和综合科学调查活动<sup>[2-3]</sup>。本世纪初, 中国科学院加强了与俄罗斯科学院、蒙古科学院等机构的科研合作, 联合在该区域开展了综合科学考察。科技部于 2008 年专门资助并启动了科技基础性工作科学考察项目“中国北

方及其毗邻地区综合科学考察”, 希望能通过联合科学考察和调查, 加强对该区域本底资料的获取和分析, 为在这一区域深入开展地球系统、全球变化和区域可持续发展研究提供数据支撑<sup>[4]</sup>。

综合科学考察数据具有多来源、多类型、多学科等特点, 这不仅影响到大量已有考察活动积累数据的集成, 而且给当前正在开展的许多科学考察研究活动的系统集成管理带来困难<sup>[5]</sup>。国际上在综合科学考察与调查领域的发展趋势表现在以下几个方面: (1) 新技术、新方法在科学考察领域广泛应用, “遥感”+“数据收集整理”+“野外调查”+“定位观测”+“模型模拟”+“综合分析”成为主流技术路线。(2) 注重资源环境的长期联网观测和数据积累, 有关地球环境、资源变化的长期观测、监测与信息网络正在快速形成。(3) 注重数据的积累与长期管理, 例如, 英国洛桑农业实验站的样品数据积累达 160 余年。我国在科学考察与调查的数据积累与管理方面, 仍然非常薄弱, 大量历史的和现状的科学考察资料缺乏集成和长期管理<sup>[6]</sup>。这些问题

收稿日期: 2011-05-25; 修回日期: 2011-12-23.

基金项目: 科技基础性工作重点项目——中国北方及其毗邻地区综合科学考察(2007FY110300); 格网化资源环境综合科学调查规范(2011FY110400); 环保公益性行业科研专项项目(201109075)资助。

作者简介: 王卷乐(1976-), 男, 博士, 副教授, 研究方向: 资源环境空间数据管理、地理信息共享、环境遥感。

E-mail: wangjl@igsrr.ac.cn

可以归结为 3 个方面:①多学科综合科学考察数据采集、管理的标准化和规范化问题,②如何建立科学、合理的数据分类体系,③如何实现这些多源、异构数据的空间化展示与可视化访问。针对这些问题,本文提出了东北亚资源环境综合科学考察数据集成体系的构架并开展了相关实践研究。

## 2 东北亚资源环境综合科学考察数据集成体系框架设计

面向东北亚资源环境综合科学考察对数据资源集成管理的需求,设计了数据集成体系框架,如图 1 所示。整个框架包括 3 部分,即考察数据的采集与规范化整理,数据目录体系构建,不同类型数据的数据库建设及多维数据可视化浏览与访问。

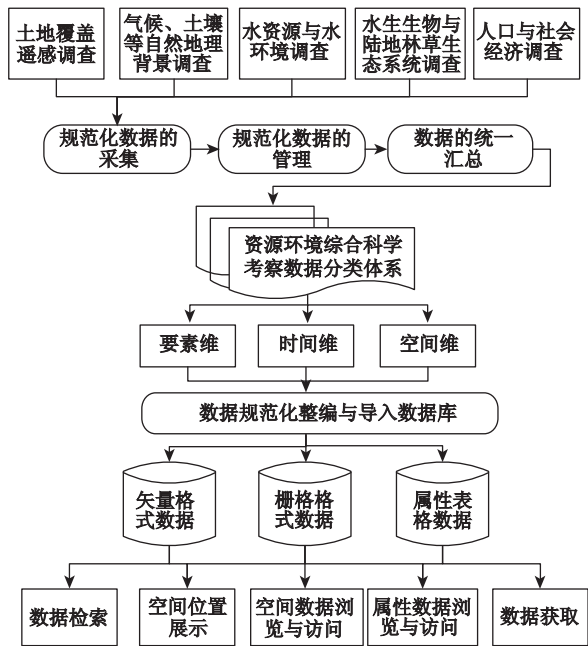


图 1 东北亚资源环境综合科学考察数据集成体系框架

Fig. 1 Architecture of resource and environment science expedition data resources integration system in Northeast Asia

(1)考察数据的采集与规范化整理。东北亚资源环境综合科学考察内容包括遥感面上土地覆盖调查、自然地理环境(土壤、气候等)调查、水资源与水环境调查、水生生物与水鸟调查、陆地生态系统(森林、草地等)调查、人居环境调查、人口与社会经济调查等。确保各专题调查活动采集到的数据资源,符合相对一致的标准规范,是考察数据集成的基础。为此,需要制定这些考察活动的数据采集与

整理规范。

(2)考察数据目录体系构建。由于采集到的综合科学考察数据来源各异,其在时间尺度、空间尺度,以及数据的要素粒度等方面存在很大的差异,因此,有必要建立整体数据目录体系,确定公共的时间尺度、空间尺度及要素粒度分类维<sup>[7-8]</sup>。

(3)不同类型数据的数据库建设及多维数据可视化访问。将所有综合科学考察获得的数据资源,按照矢量、栅格、属性等 3 种数据类型,规范化整编所有数据的时间、空间、要素等信息及元数据信息,建立数据库和元数据库。需要实现包括数据检索、空间位置展示、空间数据浏览与访问、属性数据浏览与访问,以及数据获取等功能的数据展示平台。

针对以上数据集成体系框架,本文结合科学考察活动的实践,进行了系统的梳理和分析,提出相应的解决方案,并借助于地理信息技术和计算机技术,实现整个数据体系的原型系统。

## 3 东北亚资源环境综合科学考察数据的采集与整理的规范化

### 3.1 标准规范体系架构

东北亚资源环境综合科学考察数据的获取包括 3 个环节:(1)采集野外数据和收集历史数据;(2)分析和整理数据;(3)系统管理数据并为数据共享作准备。鉴此,其在数据标准规范方面的需求可以归为 3 个方面,即数据采集和处理类标准规范、数据分析与整编类标准规范、数据管理与共享类标准规范。据此,建立的东北亚资源环境综合科学考察数据标准规范体系结构如图 2 所示。其中,数据采集与处理类包括 10 项规范,数据分析与整编类包括 7 项规范,数据管理与共享类包括 6 项规范。总计 23 项标准规范。

#### (1)数据采集与处理类规范

主要包括遥感面上调查与考察数据采集与处理规范、土壤生态样方调查技术规范、森林生态样方调查技术规范、草地生态样方调查技术规范、水资源科学考察数据采集与处理规范、水生生物及生态系统考察数据采集与处理规范、典型湖泊环境科学考察数据采集与处理规范、社会经济调查与考察数据采集与处理规范、人居环境调查与考察数据采集与处理规范、大气气溶胶数据采集与处理规范等。其内容主要包括各专题调查的前期准备、外业调查、考察资料内业

初步整理、考察数据整理、考察成果提交、质量管理、资料更新与归档等的标准规范要求。

(2)数据分析与整编类规范

主要包括综合科学考察核心元数据标准和综合科学考察数据文档编写规范。其中,综合科学考察核心元数据标准内容包括数据集标识模块、数据集内容模块、分发信息模块、质量信息模块及相应的代码表等;综合科学考察数据文档编写规范包括数据集名称、数据集内容说明、数据源描述、数据加

工方法、数据应用成果、知识产权等。

(3)数据管理与共享类规范

主要包括中国北方及其毗邻地区综合科学考察管理办法、综合科学考察数据汇交细则、考察报告撰写格式规范、野外考察日志撰写格式规范等。其中,数据汇交规范明确了考察数据汇总的技术要求,包括数据汇交组织管理、考察数据汇交内容、数据汇交流程、数据管理、权益保护、奖惩及数据汇交计划格式等内容。

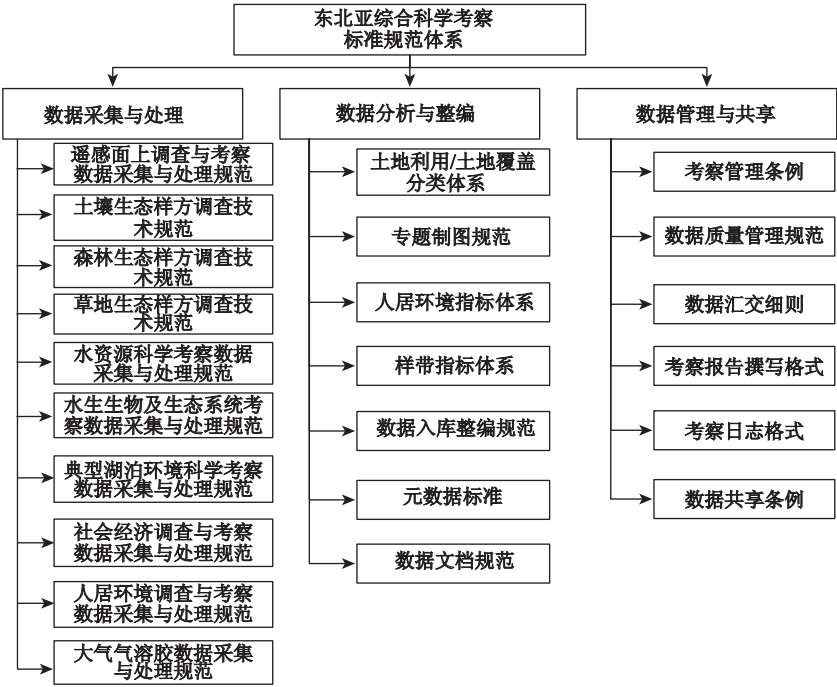


图 2 东北亚资源环境综合科学考察数据标准规范体系

Fig. 2 Standard and specification system of resource and environment science expedition data integration in Northeast Asia

3.2 标准规范研制

东北亚资源环境综合科学考察数据标准规范采用 3 个步骤开展编制:(1)明确标准化对象;(2)确定标准的规范性技术要素;(3)编写标准。在编写阶段,首先,应从标准的核心内容——规范性技术要素开始编写(包括规范性技术要素、规范性附录或资料性附录等);然后,编写标准的规范性一般要素(如规范性引用文件、标准的范围等);最后,编写资料性要素(包括引言、参考资料、索引和目次等)<sup>[9]</sup>。在各专题考察队的积极配合下,目前所有的 23 项标准规范全部形成征求意见稿,且已在科学考察数据的采集、分析、管理和共享活动中得到了应用。

4 东北亚资源环境综合科学考察数据目录体系的构建

东北亚资源环境综合科学考察的内容主要包括自然地理环境、林草生态系统、水资源、水环境、水生生物、湖泊、社会经济与人居环境、全球变化样带监测等。据此,设计其数据体系及相关要素如下:数据资源体系总体包括 4 个大类、25 个小类、128 个要素。统一为各类别和要素制定了编码体系,其中,小类按 6 位数字编码,便于分类系统的扩展、更新和维护。图 3 显示了数据资源体系的大类和小类。

该数据目录是所有专题科学考察数据集成的

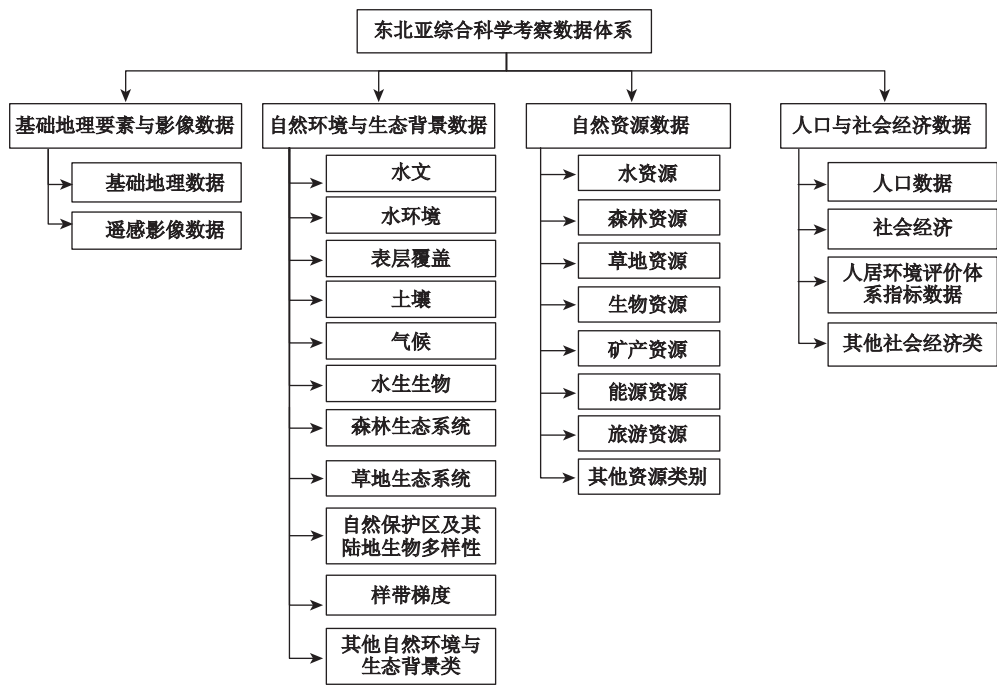


图 3 东北亚资源环境综合科学考察数据分类结构图

Fig. 3 Data classification system of resource and environment science expedition data integration in Northeast Asia

主线,最终各类考察数据将汇总在不同的数据类别中。为此,需要确立各专题数据的统一时间、空间和数据要素基准。

(1)时间基准。本次东北亚资源环境综合科学考察的时间周期是 2008 - 2012 年,结合现势和历史资料的获取情况,设立的时期基准为 3 个基准年份,即 2000、2005 和 2010 年。部分实测考察资料,以考察年当年为准。其他相关考察资料,则以这 3 个基准年份为基础进行整合。

(2)空间基准。确定考察资料收集的行政区划边界,以 WGS84 坐标系统为基准,按经纬度信息集成所有考察资料的空间信息。包括森林和草地样方数据库、水资源与水环境采样数据库、土壤剖面采样数据库、遥感解译标志数据库等。

(3)要素基准。依据各专题考察的具体资源内容,确定各类数据资源的指标要素定义,统一数据指标的语义说明,建立数据要素字典,最大限度地避免同名异义或同义异名问题。

5 东北亚资源环境综合科学考察数据平台设计与构建

为了便于东北亚资源环境综合科学考察各考

察队汇总和内部共享数据,并为后续对外共享做好准备,借助于地理信息技术与网络数据库技术,设计并研发了东北亚资源环境综合科学考察数据集成的平台软件原型系统。其直接服务对象是综合科学考察的各专题科考队员,间接的服务对象是对这一区域和相关研究领域感兴趣的科学家。根据用户的特点,其主要功能需求可以概括为以下 3 点:(1)所有入库数据遵从前文制定的标准规范;(2)数据按分类体系和编码统一管理;(3)空间数据能够可视化展示;(4)具有网络平台界面。

5.1 数据平台逻辑结构

数据平台逻辑结构分为 4 层,即原始数据层、关系型数据库层、时空数据管理功能层、可视化展示与用户交互层。平台逻辑结构如图 4 所示。

5.2 数据平台物理结构

数据平台的物理结构基于 PostGIS 数据库,Java 语言,利用 Openlayer 容器和 Flex 技术实现空间数据发布和属性数据浏览和显示<sup>[10-12]</sup>(如图 5 所示)。

数据平台中建立的数据库表主要包括 geo\_spatial, geo\_meta\_data, geo\_meta\_field, geo\_field

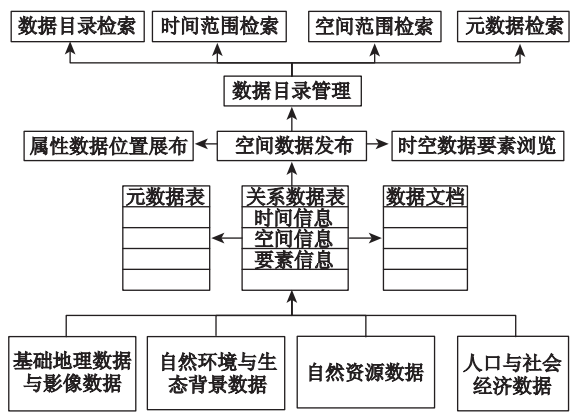


图 4 数据管理平台逻辑结构图

Fig. 4 Logical model of data management software platform

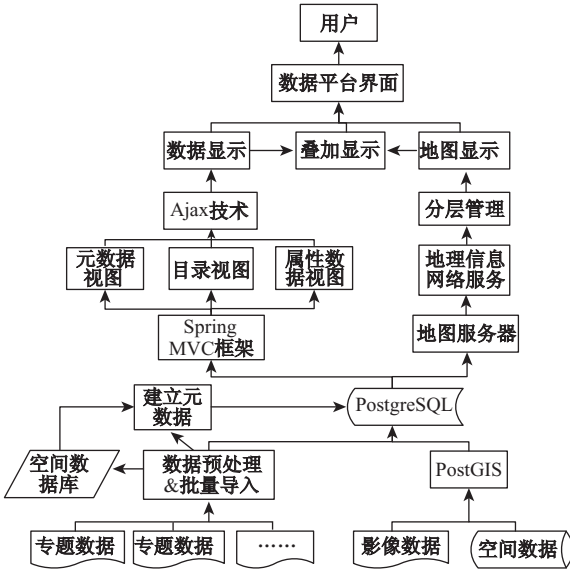


图 5 数据平台物理结构

Fig. 5 Physical architecture of data management software platform

和 geo\_category 5 个表,用以建立数据表格的元数据和目录结构。

(1)Geo\_spatial 表:记录空间地点名称及其编码,空间地点名称字段取值于数据文件中相应字段,并替换为空间位置编码进行存储。

(2)Geo\_meta\_data 表:记录数据文件的元数据内容,包括数据文件编码,数据文件名称,数据文件所属专题数据集编码,数据文件格式类型,数据采集时间,数据描述等信息。

(3)Geo\_field 表:用以记录数据文件出现过的所有字段名称,并替换成编码存储。

(4)Geo\_meta\_field 表:将 Geo\_meta\_data 表中

文件编码和 Geo\_field 表中字段编码进行关联。

(5)Geo\_category 表:用以存储专题分类类别,及分类编码,其中分类编码被引用至 Geo\_meta\_data 表。

5.3 数据集成平台原型系统的应用

初步完成的数据集成平台原型系统能够满足以上设计功能。以“东北亚土壤调查及土壤样品分析数据”为例,对数据的集成及可视化访问功能进行概要说明。

(1)土壤剖面采样数据的规范化整编:换算采样点的经纬度信息,增加野外采样点的背景信息说明,将野外采样数据与室内分层化验数据进行统一汇总。

(2)数据库存储:将以 WORD 和 EXCEL 格式存储的野外考察数据,按照建立好的数据结构,存储到关系型数据库或空间数据库中(本平台采用 PostGIS 数据库存储)。

(3)数据检索:通过数据目录树、时间范围、空间范围、属性要素信息等均可以查询到所需的数据集。如图 6 显示的是以数据目录树的方式查询到的“东北亚土壤调查及土壤样品分析数据集”。



图 6 数据目录方式检索数据

Fig. 6 Data resources search through data catalogue

(4)空间位置显示及数据浏览:点击该数据集名称后,能够将该数据集的采样点位置展布在地理底图上,点击各相应的采样点,可以浏览到采样点数据的元数据说明、东北亚区域的采样点位置分布及其数据描述信息(见图 7)。

(5)数据内容访问:点击详细数据内容,则可以打开采样点数据的属性信息列表,查看详细内容。包括土壤剖面的位置(经纬度和海拔)、所在位置地名、土壤分层、有机质含量、全氮含量、全磷含量、土



图 7 东北亚区域土壤剖面样点的分布  
Fig. 7 Distribution of soil profile sampling location in Northeast Asia

壤质地、生境描述等(如图 8 示)。

东北亚土壤剖面及土壤样品分析数据									
开始时间	结束时间	空间单元名称	实验编号	剖面编号	层次	有机质(g/kg)	全氮(g/kg)	全磷(g/kg)	全钾(g/kg)
2008-07-01	2008-08-20	俄罗斯	1	BP01001	1	44.18	1.851		
2008-07-01	2008-08-20	俄罗斯	2	BP01002	2	5.79	0.266		
2008-07-01	2008-08-20	俄罗斯	3	BP01003	3	2.37	0.286		
2008-07-01	2008-08-20	俄罗斯	4	BP01301	1	81.01	2.573		
2008-07-01	2008-08-20	俄罗斯	5	BP01302	2	181.99	4.715		
2008-07-01	2008-08-20	俄罗斯	6	BP01303	3	63.02	1.491		
2008-07-01	2008-08-20	俄罗斯	7	BP01304	4	32.71	0.144		
2008-07-01	2008-08-20	俄罗斯	8	BP01305	5	8.96	0.329		
2008-07-01	2008-08-20	俄罗斯	9	BP00701	1	21.73	1.587		
2008-07-01	2008-08-20	俄罗斯	10	BP00702	2	36.41	1.525		
2008-07-01	2008-08-20	俄罗斯	11	BP00703	3	10.11	0.587		
2008-07-01	2008-08-20	俄罗斯	12	BP00704	4	2.29	0.26		
2008-07-01	2008-08-20	俄罗斯	13	BP00201	1	53.16	3.309		
2008-07-01	2008-08-20	俄罗斯	14	BP00202	2	28.09	1.332		
2008-07-01	2008-08-20	俄罗斯	15	BP00203	3	18.97	1.116		
2008-07-01	2008-08-20	俄罗斯	16	BP00204	4	10.24	0.519		
2008-07-01	2008-08-20	俄罗斯	17	BP00205	5	5.66	0.432		
2008-07-01	2008-08-20	俄罗斯	18	BP01201	1	82.08	3.933		

图 8 土壤剖面样点调查与分析属性数据列表  
Fig. 8 Attribute information list of soil profile sampling investigation and analysis in Northeast Asia

(6) 获取数据: 点击获取数据可以打包下载 Shapefile 格式的空间数据及 EXCEL 格式的属性能数据。

6 结语

东北亚区域资源环境综合科学考察数据集成体系,已在科学考察的数据集成中得到初步应用。制定的标准规范体系已在 8 个专题考察队中得到应用,所有采集到的科考数据,均遵从采集规范,且具有元数据、数据文档等相关说明。在统一的数据分类体系下,现已汇集了近千个数据集,数据库存量超过 2TB,经过规范化整理的数据集实体有 144 个。建立的东北亚资源环境科学考察数据平台,允许各类符合标准规范的数据在统一的数据分类体系下进行集成和管理,实现了以空间数据展示为主线的可视化访问,能够满足当前各科学考察队员和科学家用户的需求。然而,由于资

源环境综合科学考察获取的数据历史跨度大,许多数据极不规范,非常难以建库集成,后续还将针对历史数据和长期监测数据等类型的差异,分别完善相应的整合集成标准规范和技术方法,促进这些宝贵的科学考察数据能进入我国科学数据集成与共享体系,服务于我国的科技创新和国家经济社会的发展。

致谢:感谢中国北方及其毗邻地区综合科学考察各专题考察队的合作支持,感谢俄罗斯科学院西伯利亚分院贝加尔湖自然资源管理研究所、蒙古科学院地理所等机构的支持。

参考文献:

[1] 孙鸿烈主编. 中国自然资源综合科学考察与研究[M]. 北京:商务印书馆,2007.

[2] Northern Eurasia Earth Science Partnership Initiative [DB/OL]. <http://neespi.org/>

[3] Yuri MUN, Ick Hwan KO, Luntun JANCHIVDORJ, et al. Integrated water management model on the Selenge River Basin Status Survey and Investigation[DB/OL]. 2008, <http://www.kei.re.kr>

[4] 孙九林,董锁成,等. 中国北方及其毗邻地区综合科学考察实施方案[R]. 2008.

[5] 王卷乐,朱立君,孙崇亮. 资源环境综合科学考察中的多维数据集成管理模式研究与实践——以中国北方及其毗邻地区综合科学考察为例[J]. 自然资源学报, 2011,26(7):1129-1137.

[6] Wang J L. Study on geosciences multi-dimensions data model Integration and sharing oriented[C]. The 18th International Conference on Geoinformatics, Beijing, China,2010.

[7] 王卷乐. 科学数据整合集成与共享中的关键技术问题研究——以研究型、参考型数据为例[D]. 中国科学院地理科学与资源研究所博士后出站报告,2007.

[8] 孙九林,林海主编,地球系统研究与科学数据[M]. 北京:科学出版社,2009.

[9] 白殿一,等. 标准的编写[M]. 北京:中国标准出版社, 2009.

[10] Pinde Fu, Jiulin Sun (ed.). Web GIS principles and applications[M]. ESRI Press, 2010.

[11] 刘光,唐大仕(编). Web GIS 开发——ArcGIS Server 与 .Net[M]. 北京:清华大学出版社,2009.

[12] 方元,赵冠伟,何观生. 基于 Ajax 和 GeoServer 的 WebGIS 设计[J]. 微计算机信息, 2009,25(1).

## Construction of Resource and Environment Science Expedition Data Resources Integration System in Northeast Asia

WANG Juanle<sup>1</sup>, SONG Jia<sup>1</sup> and ZHU Lijun<sup>1,2</sup>

(1. *State Key Lab of Resources and Environment Information System, Institute of Geographic Sciences and Nature Resources Research, CAS, Beijing 100101, China;*

2. *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*)

**Abstract:** Northeast Asia is a key area for Earth system study, global change frontier science research and regional sustainable development research, where has complex ecological environment pattern, variety of climate zones, and typical human-earth relationship comparing characteristics. Large amount of scientific expedition data are accumulated in the past several years in Northeast Asia (here refers to Northern China, Mongolia, and Siberia and Far East area of Russia) under the cooperation of China, Mongolia and Russia scientists. These data have multi-disciplinary features. How to integrate these scientific expedition data public services is a problem. According to these scientific expedition data accumulation and management requirements, we designed a data resources integration system, which has 3 sub systems, i. e. , data resources collect and management standards and specifications system, data classification system, data management and publication software platform. Data resources collection and management standard and specification system has 23 specifications, which can be divided into 3 types, i. e. , data collection and processing specification type, data analysis and archiving specification, and data management and sharing specification. Data resources classification system has 4 classes, 25 sub classes and 128 data elements. Data management and publication software platform has 5 function models, i. e. , data catalogue search model, metadata management model, data publication and virtualization model, data view model and data download model. Based on the designed data integration system, a prototype system has been developed supported by computer technology and Web GIS technology. As far 144 datasets have been integrated in this data system. With more and more data are accumulated and integrated in this system, it will play an important role for the scientific expedition data application and analysis in the near future.

**Key words:** resource and environment; integrated science expedition; data system; Northeast Asia