

# 国家地球系统科学数据共享平台 数据分类编目与特征分析

王卷乐<sup>1,5</sup> 宋佳<sup>1,5</sup> 卜坤<sup>2</sup> 王末<sup>1,3</sup> 赵红伟<sup>1,3</sup> 田奋民<sup>1,4</sup> 柏永青<sup>1,3</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. 中国科学院东北地理与农业生态研究所, 吉林长春 130102; 3. 中国科学院大学, 北京 100049;  
4. 中国矿业大学(北京), 北京 100083; 5. 江苏省地理信息资源开发与利用协同创新中心,  
江苏南京 210023)

**摘要:** 数据分类编目在数据共享平台管理中发挥着核心作用。以已建立的“面向共享的地球系统科学数据分类”为基础, 实现了国家地球系统科学数据共享平台1832条数据集的实体分类编目, 定量分析数据共享平台数据资源在分类体系内和各分平台内的分布特征, 提出潜在的数据资源建设策略建议。研究产生的数据分类编目体系已在国家地球系统科学数据共享平台新版平台实现在线管理和应用。

**关键词:** 国家科技平台; 地球系统; 数据共享; 分类体系; 数据资源建设

中图分类号: P9

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2015.06.012

## Data Classification Cataloguing and Feature Analysis of National Data Sharing Platform of Earth System Science

Wang Juanle<sup>1,5</sup>, Song Jia<sup>1,5</sup>, Bu Kun<sup>2</sup>, Wang Mo<sup>1,3</sup>, Zhao Hongwei<sup>1,3</sup>, Tian Fenmin<sup>1,4</sup>, Bai Yongqing<sup>1,3</sup>

- (1.State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101; 2. Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, 130102; 3. University of Chinese Academy of Sciences, Beijing 100049; 4. China University of Mining & Technology (Beijing), Beijing 100083; 5. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023)

**Abstract:** Data classification plays core role in the data management of data sharing platform. Based on the built “Data Sharing Oriented Earth System Science Data Classification System”, totally, 1832 data entities are classified

**作者简介:** 王卷乐\*(1976-), 男, 博士, 中国科学院地理科学与资源研究所研究员, 研究方向: 资源环境科学数据集成与共享标准研究; 宋佳(1980-), 男, 博士, 中国科学院地理科学与资源研究所助理研究员, 研究方向: 地理本体和语义搜索; 卜坤(1980-), 男, 博士, 中国科学院东北地理与农业生态研究所高级工程师, 研究方向: 遥感与GIS应用研究; 王末(1987-), 男, 中国科学院地理科学与资源研究所博士研究生, 研究方向: 空间数据挖掘; 赵红伟(1987-), 女, 中国科学院地理科学与资源研究所博士研究生, 研究方向: 地理空间大数据挖掘; 田奋民(1993-), 男, 中国矿业大学(北京)硕士研究生, 研究方向: 科学数据分类与可视化平台开发; 柏永青(1992-), 男, 中国科学院地理科学与资源研究所硕士研究生, 研究方向: 格网化资源环境数据管理。

**基金项目:** 国家科技基础条件平台: 地球系统科学数据共享平台; 中国科学院信息化专项项目“资源学科领域基础科学数据整合与集成应用”(XXH12504-1-01)。

**收稿时间:** 2015年10月25日。

and coded in National Data Sharing Platform of Earth System Science (DSPESS). Data sets distribution features are quantitatively analyzed respectively from the view of classification system and data center system. Some potential data resources development suggestions are proposed based on the analysis. This classification and management system has been applied in the updated online website of DSPESS (<http://www.geodata.cn/>).

**Keywords:** Science & Technology Platform of China, Earth System, data sharing, classification system, data resource development

## 1 引言

地球系统科学数据共享平台于2002年作为国家科学数据共享工程首批试点之一启动<sup>[1]</sup>,其主要目标是整合集成科研院所、高等院校和科学家个人通过科研活动所产生的研究型科学数据,服务于地球系统科学与全球变化等基础和前沿科学研究<sup>[2]</sup>。2005年该平台纳入国家科技基础条件平台建设,2011年通过国家评议正式成为首批进入运行服务阶段的国家平台<sup>[3]</sup>。该平台采用分布式的共享架构,包括一个总中心和14个分平台,即长江三角洲、青藏高原、寒区旱区、黄河中下游、黄土高原、东北黑土、南海及其毗邻海区(以下简称南海)、新疆与中亚、极地等区域分平台和湖泊-流域、地质与地球物理(以下简称地球物理)、空间科学、天文科学、全球变化模拟等学科和综合分平台<sup>[3]</sup>。

地球系统科学数据共享平台中的数据资源涉及地球系统的各个圈层、领域、学科和特色区域,如何科学、有序、长期、可持续地积累和管理这些数据资源是一个关键问题,也是一个难点问题。数据分类编目是数据共享平台管理数据的重要抓手。本文在该平台初步建立的“面向共享的地球系统科学数据分类”<sup>[4]</sup>基础上,开展了现有全部数据实体的分类编目,实现了全部数据的在线分类管理,进而定量分析共享平台数据资源在分类体系内和各分平台内的分布特征,提出潜在的数据资源建设方向的建议。

## 2 平台数据分类编目

### 2.1 平台数据分类体系

王卷乐等在《面向数据共享的地球系统科

学数据分类探讨》<sup>[5]</sup>中提出了地球系统科学数据的分类体系。基于该分类,适当应用优化后,形成10个一级类和114个二级类分类体系(表1)。其中,一级类包括:大气圈、陆地表层、陆地水圈、冰冻圈、自然资源、海洋、极地、固体地球与古环境、日地空间环境与天文、遥感数据。

### 2.2 平台数据编目

为了便于数据实体的分类管理,以Web开发的方式实现了国家地球系统科学数据共享平台的所有数据在线分类管理,且已部署应用于国家地球系统科学数据共享平台新版平台<sup>[6]</sup>([www.geodata.cn](http://www.geodata.cn))中,具体包括数据分类定制和数据实体分类管理两部分功能。

数据分类定制采用树形结构,支持在线的数据分类添加、修改和删除操作,如图1所示。数据实体分类管理可对每个数据指定其所属类别,同时可以指定该数据的属性信息。以“2002年福建省1:10万地形地貌数据”为例,可将其赋予“陆地表层”一级类和“地形”二级类。与此同时,可赋予该数据的数据产生方式为“考察调查”,数据类型为“栅格”,比例尺为“1:10万”,空间尺度为“省级尺度”,空间分辨率为“公里级”,时间分辨率为“年”。通过数据分类和属性项绑定,可为用户在检索数据时提供更多筛选条件,更快速、准确地获得所需数据。

按以上分类体系,对共享平台全部1832条数据(截至2014年4月)进行了分类编目。分类的技术流程包括两个步骤,一是组织专业人员对所有数据进行初步分类;二是将分类结果发送给该数据所对应归属的总中心或相关分平台,由其审核并修改后,再统一汇总完善。

表1 地球系统科学数据共享平台分类编目体系

一级分类	二级分类
大气圈	综合观测、气候、温度、降水、湿度、日照、风速、蒸发、气压、辐射、温室气体、气溶胶、大气质量、大气成分、其他
陆地表层	基础地理、土地利用/覆盖、人口、社会经济、区划、地形、地貌、土壤、沙漠、荒漠、湖泊、水库、湿地、植被、生态、环境、灾害、其他
陆地水圈	水文、地表水、地下水、水循环、水利工程、水环境、水化学、其他
冰冻圈	冻土、冰川、冰、雪、其他
自然资源	气候资源、生物资源、水资源、土地资源、农业资源、矿产资源、药物资源、旅游资源、能源资源、可再生资源、海洋资源、其他
海洋	基础地理、海洋水文、海洋气象、海洋化学、海洋生物、海洋生态、海洋水产、海洋地质、海洋物理、海岸过程、海洋工程、海平面、海岸带、其他
极地	极地海洋、极地地球物理、极地大气、极地生物、极地环境、极地地理、极地地质、极地工程、极地冰川、极地天文、其他
固体地球与古环境	地震、地磁、重力、岩石/矿物、地质、地热、火山、地球化学、冰芯记录、海洋/湖泊记录、陆地记录、古气候重建、其他
日地空间环境与天文	太阳、太阳系、银河系、河外与宇宙、行星际空间环境、地球卫星轨道空间环境、电离层、中高层大气、地球磁场、地面宇宙线、空间天气灾害事件、空间环境效应、其他
遥感数据	航片、卫星影像、雷达影像、地物波谱、反演数据产品、遥感解译产品、其他



图1 数据分类编目定制界面

### 3 数据分类应用分析

#### 3.1 数据一级分类总体特征

共享平台数据按一级类的分类汇总情况如表2所示。由表2可见,赋予不同一级类的数据集总数达2878个,多于1832条数据集的总数,这说明存在着某一个数据集对应着多个数据分类的情况。所属一级类的数据集平均数为288,这总体反映了国家地球系统科学数据共享平台的编目体量。在数据分类结果中,占比重最大的是“陆地表层”数据,占31.03%,这反映出该平台当前以陆地表层(地理学)数据为主的特征。其次多的是“极地”数据,占17.58%,这反映出我国南极考察(30次)、北极考察(6次)数据集成内容较为丰富、数据集条目较多,同时也反映出极地数据编目较细(例如,一条航次的某一类要素观测数据被列为一条独立的数据集)。再其次的是“自然资源”和“海洋”分类数据,均占10.98%,这反映出海洋数据的体量也较大,其原因可能是海洋领域的二级分类范围较大,几乎陆地上的所有领域类别在海洋中均有体现。自然资源在该共享平台中没有专属的数据分平台,但也占到了较高的分类比重,这说明一级类别中的许多数据资源兼具自然资源的属性,在一个数据允许有多个分类的情况下,有许多专题数据都被同时分到自身的专业类别和自然资源类别中。“大

气圈”的数据类目所占比重为8.51%,这一比重与地球系统科学研究热点的地位是不相称的,这也反映出当前共享平台尚未有独立的“大气”数据分平台,因此缺少相应的基础和动态增长数据资源。“固体地球与古环境”与“日地空间环境与天文”数据所占比重分别为6.22%和4.90%,这反映出其在该共享平台内的资源量较少,这与这两类数据的“大数据”特征不相符。其原因:一方面可能是承担该数据资源建设的分平台尚未将大量数据整合进来,另一方面恰恰是这些数据量大而难以整合进来。例如:天文和空间探测数往往以日均TB的数据量增长,如何整合这些以自动巡天观测为主的数据资源则是一个难题。“陆地水圈”和“冰冻圈”的数据所占比重为3.13%和1.29%,这反映出有关湖泊、湿地、冰雪等研究的数据较少,有关寒区旱区和湖泊一流域的相关平台应进一步增加冰、雪数据产品的体量。“遥感数据”已占到总体量的5.39%,这与遥感数据类别和条目不多、但数据量大的特点相一致。

#### 3.2 数据二级分类特征

将各一级类数据展现在其二级类层次上,可更清楚地反映出共享平台数据资源的类别分布特点。具体分布情况如图2所示。

在大气圈数据中,占比最多的是气象、大气成分、温度、降水、气溶胶和温室气体数据,相应的日照、蒸发、气压、辐射等数据较为缺

表2 地球系统科学数据共享平台数据集按一级分类编目统计表

一级类	所属该类的数据集个数	比例/%
大气圈	245	8.51
陆地表层	893	31.03
陆地水圈	90	3.13
冰冻圈	37	1.29
自然资源	316	10.98
海洋	316	10.98
极地	506	17.58
固体地球与古环境	179	6.22
日地空间环境与天文	141	4.90
遥感数据	155	5.39

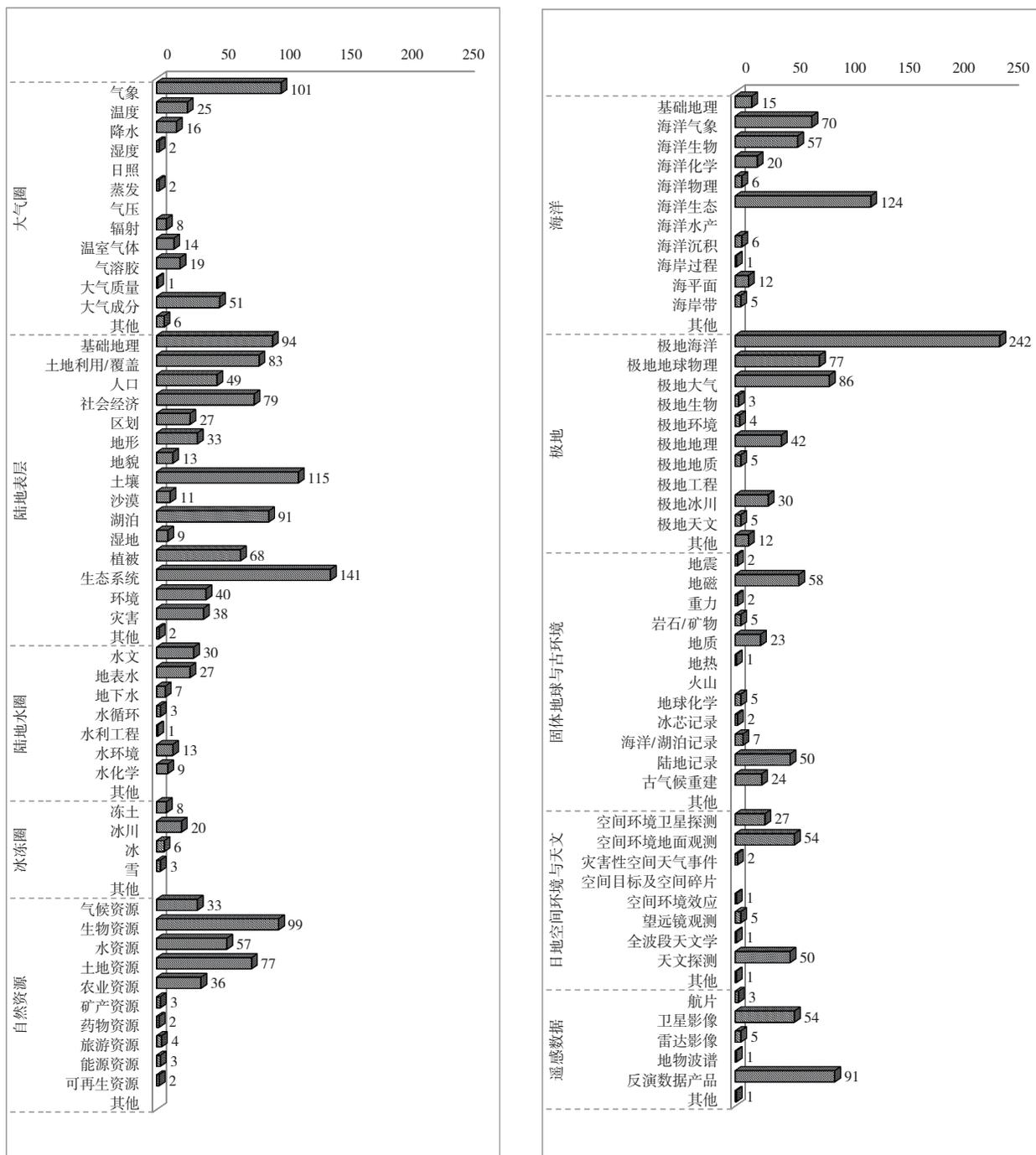


图2 地球系统科学数据共享平台数据在二级类目上的分布

乏，甚至有些数据类别没有任何数据。陆地表层数据的二级类目数据分布较为均衡，各门类均有数据，占优的数据包括生态系统、土壤、基础地理、土地利用/覆盖、人口与社会经济等，这与其当前的主要用户需求特点是相符的。而较少的数据类目为沙漠、湿地、地貌等专题数据。陆地水圈数据中相对较多的是水文、地表水和水环境

数据，而水循环、水利工程等较少。冰冻圈数据中以冰川数据为主，其他数据较少。在自然资源数据中，以生物资源、土地资源、水资源、农业资源和气候资源为主体，而涉及矿产、药物、旅游、能源、可再生资源等数据较少。海洋数据的二级类目分布较不均衡，海洋生态、气象、生物等数据较多，而海洋物理、水产、沉积、过程、

海岸带等数据非常少。极地数据的分布也存在不均衡性。极地海洋、地球物理、大气等数据较多,而极地生物、环境、地质、工程、天文等数据明显较少。在固体地球与古环境数据中,地磁数和陆地记录数据居多,地质领域有关地震、重力、岩石/矿物等数据较少,古环境领域的海洋/湖泊、冰芯等记录也较少。日地空间环境与天文数据中天文探测和空间环境观测数据较多,而其他专题数据较少,如涉及灾害性空间天气事件、空间目标及空间碎片、空间环境效应、全波段天文学等数据。在遥感数据中主要以反演数据产品和航天卫星影像数据为主,其他地基遥感、航空遥感和主动雷达遥感等数据较少。

### 3.3 二级类数据的学科和区域分布特征

不同类别数据分别归属在总中心或不同的分平台中(图3)。大气圈数据主要集中在全球变化分平台和极地分平台。陆地表层数据突出集中于总中心。陆地水圈数据主要集中在总中心和湖泊分平台。冰冻圈数据主要集中在寒区旱区分平台和总中心,而极地分平台所占的比重并不高。自然资源数据主要集中在总中心。海洋数据主要集中在极地分平台和南海及其毗邻海区分平台,相比之下南海及其毗邻海区分平台的体量较小。极地数据主要集中在极地分平台。固体地球与古环境数据则依存在总中心和极地、空间和地质与地球物理分平台。这是因为古环境没有分平台,其数据直接存储在总中心。日地空间环境与天文数据主要集中在空间、天文、极地和地质与地球物理分平台。遥感数据主要集中在全球变化、极地、寒区旱区等大尺度研究领域的分平台。

### 3.4 学科和区域数据分平台的数据分类特征

在图4中,以总中心及各学科和区域数据分平台为对象,显示了数据资源在各科目中的分布特点。在总中心数据中,以陆地表层和自然资源数据为主,兼顾部分古环境数据,这与总中心依托中科院地理资源所、以陆地表层和自然资源数据为主的特色相吻合。长江三角洲分平台以区域的陆地表层和自然资源数据为主。青藏高原分

平台以区域陆地表层、大气、固体地球等数据为主。寒区旱区分平台以区域陆地表层和遥感数据为主,反映了其在本区域最敏感的水、冰等陆地水圈和冰冻圈数据尚有提升空间。黄河中下游分平台以区域陆地表层数据为主,也同样反映了其在黄河流域的水文水资源数据方面的欠缺。黄土高原分平台以区域陆地表层数据为主。东北分平台以区域陆地表层和自然资源数据为主。南海及其毗邻海区分平台以海洋数据为主。新疆与中亚分平台以区域陆地表层、遥感影像和自然资源数据为主。极地分平台以极地和海洋数据为主。湖泊-流域分平台以区域陆地表层和自然资源数据为主,但其在陆地水圈的数据相对较少。地质与地球物理分平台以固体地球和日地空间环境数据为主。空间科学分平台以日地空间环境与天文数据为主。天文科学分平台以天文数据为主。全球变化分中心则以大气层和遥感数据为主。从以上数据资源分布情况可以看到,各分平台的定位与其数据资源的布局基本合理,但部分平台在本学科或本区域特色数据资源内容上有缺失,存在提升的空间。

## 4 讨论与结论

本文在“面向数据共享的地球系统科学数据分类”的基础上,对国家地球系统科学数据共享平台的1832条数据集资源(截至2014年4月)进行了分类编目,分析了数据资源在分类体系结构和分平台中的布局,初步形成以下认知。

(1)从数据资源的一级类型分布特征来看,陆地表层数据占3成,为第一方阵;极地数据、自然资源和海洋数据为第二方阵,几乎占数据资源4成;大气数据、固体地球与古环境、日地空间环境与天文、陆地水圈和冰冻圈等数据占了其他数据的3成。以上布局揭示出“大气圈”的数据类目占比较低,这与其地球系统科学研究热点的地位是不相称的,未来可以考虑建立独立的“大气”数据分平台。固体地球数据占比不高,这与其具有的地学“大数据”特征不相符合,这说明需要进一步加强固体地球(地球物理)数据

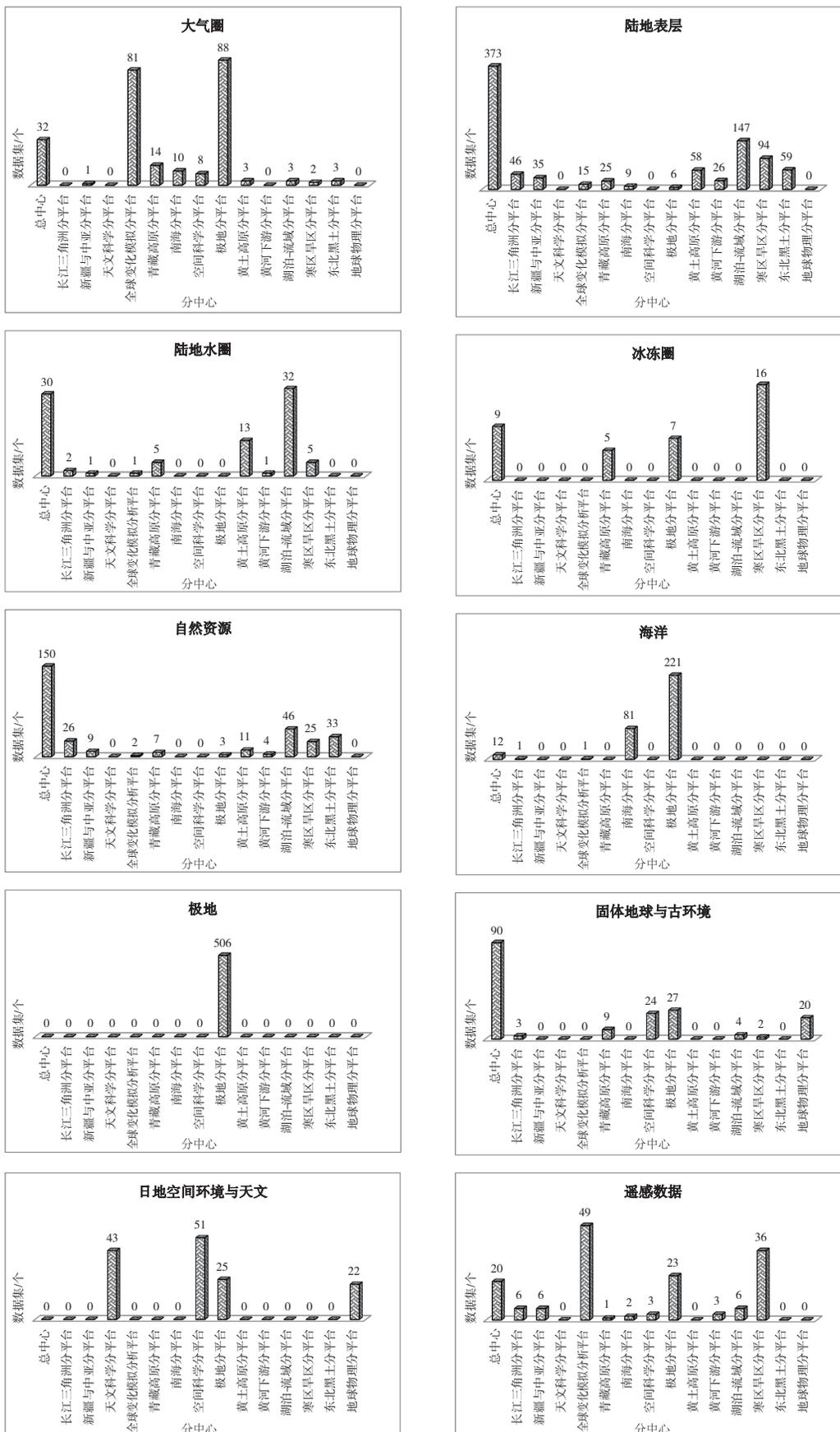
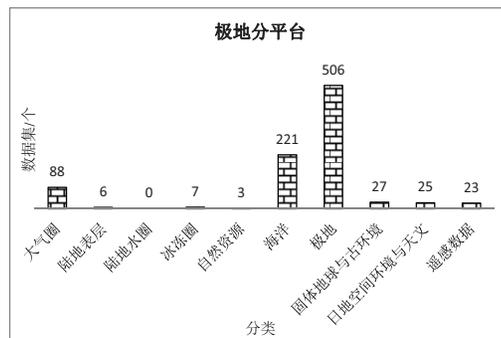
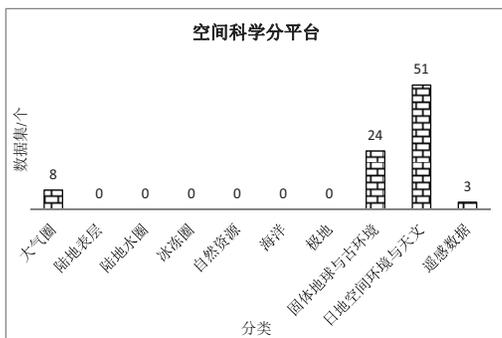
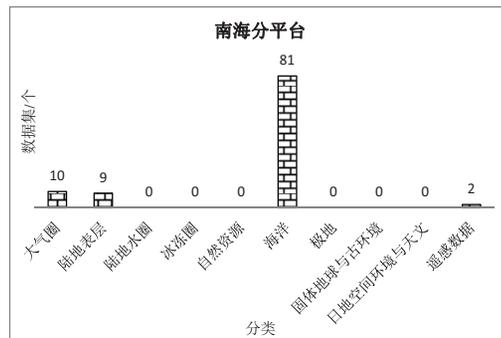
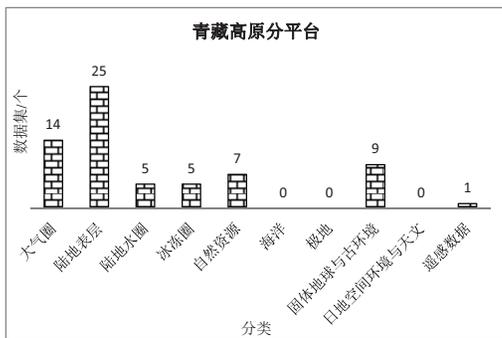
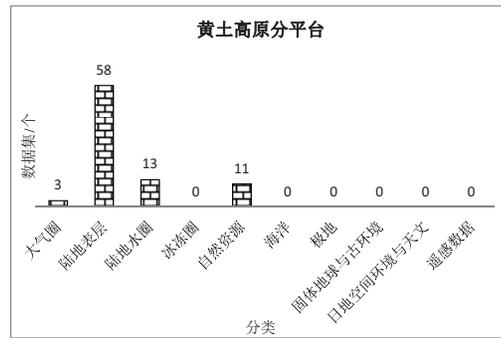
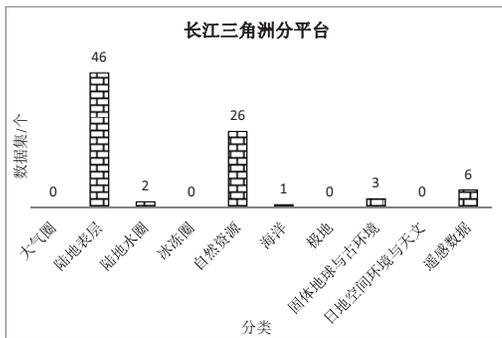
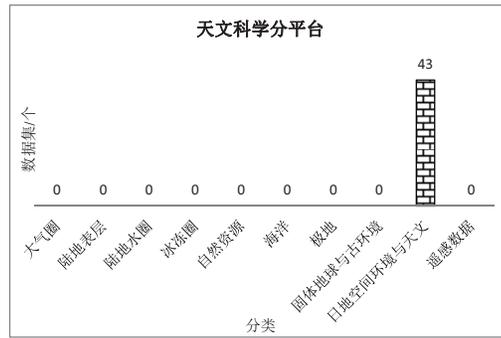
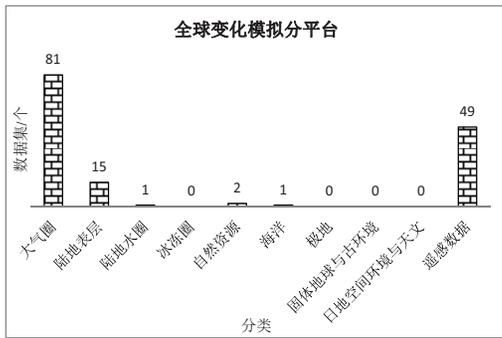
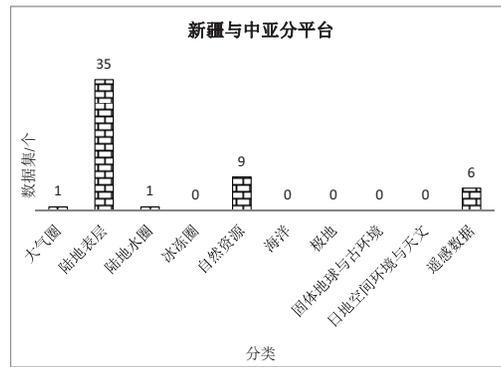
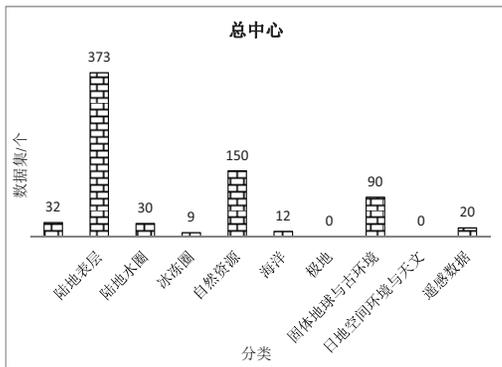


图3 不同类目数据在共享平台体系中的分布



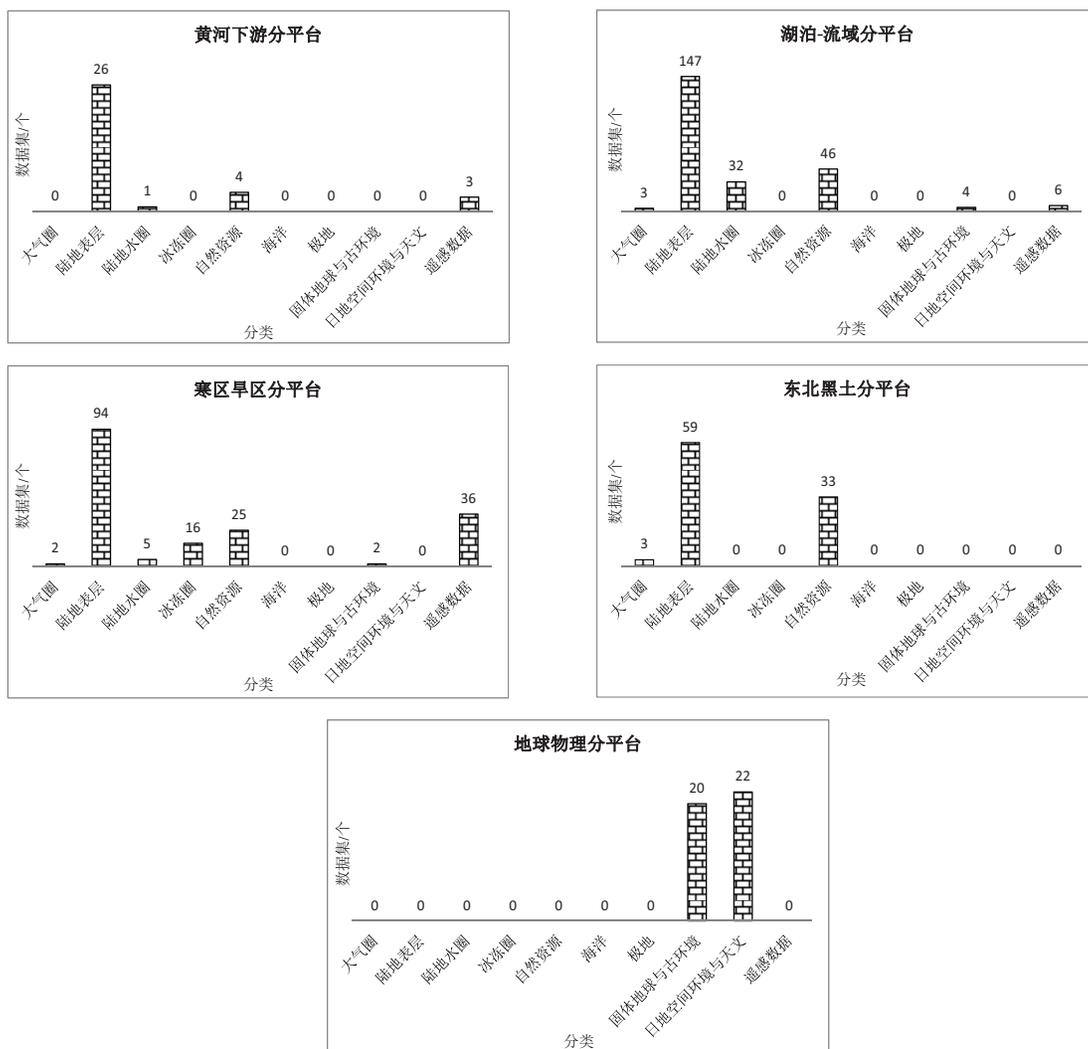


图4 学科和区域数据分平台的数据分类分布图

分平台建设。“陆地水圈”和“冰冻圈”的数据占比最低，这说明需加强寒区旱区、湖泊一流域等相关分平台中的水文、水资源、冰、雪数据产品等建设。

(2)从数据资源的二级类型分布特征来看，陆地表层数据资源的布局相对均衡，这体现了共享平台依托单位的整体优势。在大气圈、海洋、陆地水圈与冰冻圈、固体地球与古环境、日地空间环境与天文、自然资源等都存在着数据资源类型布局不均衡，甚至存在部分二级分类体系空白的问题。这表明共享平台中的数据资源体系在地球系统科学的分类框架上仍显不足，需要根据各分类特点，加强数据共享联盟建设，促进数据资

源的综合和协调发展。

(3)从各数据资源分布于总中心或不同的分平台特点来看，每一数据分类体系应该主要突出集中在某一平台中，如果分散在多个平台中，则暗示其重点不够突出，或应考虑新建一些分平台。例如，大气圈数据分布于总中心、全球变化与极地分平台，显示出大气数据的需求强烈，可以考虑未来新增大气分平台的建设；陆地水圈数据分散于总中心与湖泊一流域分平台、黄土分平台，说明本应以此数据为重点的寒区旱区数据分中心需要加强；海洋数据集中在南海和极地分平台，但南海的占比较低，反映出需要加强海洋分

(下转第101页)

机构参与的积极性, 实现政府、中小企业与科研机构之间相互联动与共赢, 发挥创新券最大的效用<sup>[9]</sup>。

### 参考文献

- [1] 张贵红, 朱悦, 朱金鑫. 发达国家或地区创新券制度分析[J]. 科技进步与对策, 2014, 31(12): 119-123.
- [2] 姜桂兴, 庄楠, 程如烟, 等. 财政资金支持企业技术创新的机制研究[J]. 全球科技经济瞭望, 2013(10): 43-50.
- [3] 陈法玉. 地方政府施行科技创新券政策探析[J]. 行政管理改革, 2014(5): 49-53.
- [4] 朱绍斌. 中国为中小企业发展绘宏图: 从“铺天盖地”到“顶天立地”[EB/OL]. (2014-10-30)[2015-08-27]. [http://news.xinhuanet.com/mrdx/2014-10/31/c\\_13375](http://news.xinhuanet.com/mrdx/2014-10/31/c_13375)

6539.htm.

- [5] 郭丽峰, 郭铁成. 用户导向的政府创新投入政策——创新券[J]. 科技创新与生产力, 2012(8): 10-13.
- [6] 郭效丁, 仝爱华. 农业科技发展的金融支持研究——以宿迁市为例[J]. 安徽农业科学, 2014, 42(14): 4507-4509.
- [7] 陆海霞. 科技型中小企业技术创新问题研究[D]. 福州: 福建师范大学, 2006.
- [8] 郭铁成, 郭丽峰. 以创新券政策促进中小企业创新[EB/OL]. (2012-04-02) [2015-08-27]. [http://www.wokeji.com/pl/kjzl/201204/t20120411\\_168000.shtml](http://www.wokeji.com/pl/kjzl/201204/t20120411_168000.shtml).
- [9] 吕宏志. 国有中小企业发展问题探讨[J]. 经济经纬, 2002(1): 50-53.
- [10] 范旭, 谷飞. 科研机构在创新型城市建设中的作用分析[J]. 科技管理研究, 2010(13): 9-14.

(上接第73页)

平台建设; 固体地球与古环境数据主要分布于总中心, 这反映出在我国地学界占有较高学科地位的古环境数据缺乏独立的位置, 应该考虑建立独立的“古环境”数据分平台。

以上讨论表明, 数据分类编目是数据共享平台的核心管理任务, 只有具有完整的数据分类信息, 才便于数据的组织、管理和服务。本文结合国家地球系统科学数据共享平台分类编目, 分析了当前该平台的数据资源体系架构及其在不同类别上的分布特征, 提出了相应的数据资源建设策略建议。本文的分析也反映出在当前全球数字对象标识(DOI)<sup>[7]</sup>和我国科技资源标识快速发展及应用的背景下, 及时加强科技平台的分类编目和资源标识, 对于促进科学数据资源的全生命周期<sup>[8]</sup>管理与开放共享具有重要意义。

致谢: 感谢国家地球系统科学数据共享平台总中心和各分平台及相关专家对数据资源分类的反馈和建议。感谢孙九林、林海、黄鼎成等专家对本文的指导。

### 参考文献

- [1] 徐冠华. 实施科学数据共享, 增强科技竞争力[J]. 中国基础科学, 2003(1): 5-9.
- [2] 孙九林, 施慧中. 中国地球系统科学数据共享服务网的构建[J]. 中国基础科学, 2003(1): 76-82.
- [3] 科技部. 关于国家生态系统观测研究网络等23个国家科技基础条件平台通过认定的通知[EB/OL]. [2011-12-31]. [http://www.most.gov.cn/tztg/201111/t20111115\\_90870.htm](http://www.most.gov.cn/tztg/201111/t20111115_90870.htm).
- [4] 孙九林, 王卷乐. 探索分散科学数据资源共享之路[M]//整合共享创新——国家科技基础条件平台建设回顾与展望. 北京: 中国科学技术出版社, 2009.
- [5] 王卷乐, 林海, 冉盈盈, 等. 面向数据共享的地球系统科学数据分类探讨[J]. 地球科学进展, 2014, 29(2): 86-95.
- [6] 苗茹, 宋佳, 王卷乐, 等. 地球科学数据共享平台建设思路与实践[J]. 中国科技资源导刊, 2014, 46(4): 64-72.
- [7] 吴立宗, 王亮绪, 南卓铜, 等. DOI在数据引用中的应用: 问题与建议[J]. 遥感技术与应用, 2013(3): 377-382.
- [8] 叶玉江. 加强科技平台工作推进科技资源管理[J]. 中国科技资源导刊, 2015, 47(2): 1-6.