

黄河流域生态保护与高质量发展体系化 科学数据建设与实践

康健芳^{1,2} 张耀南^{1,2} 王家耀³ 贾泽祥⁴ 韩立钦⁵ 刘 春³ 敏玉芳^{1,2} 李红星^{1,2} 吴亚敏² 张彩荷²
(1. 中国科学院西北生态环境资源研究院, 甘肃兰州 730000; 2. 国家冰川冻土沙漠科学数据中心, 甘肃兰州 730000; 3. 河南省时空大数据产业技术研究院, 河南郑州 450000; 4. 黄河上中游管理局, 陕西西安 710000; 5. 河南师范大学, 河南新乡 453007)

摘要: 黄河是中华民族的母亲河, 中华文明的发祥地, 黄河流域生态保护、高质量发展研究, 需要覆盖全流域的高质量科学数据支持。针对黄河流域基础科学数据碎片化、零散化、难整合、难集成、难共享、难以系统支持科学研究与应用等问题, 国家冰川冻土沙漠科学数据中心联合沿黄若干机构, 收集来自于统计年报、文献文档、政府数据、调查考察、试验示范、“空天地”观测、数值模拟、交换交流等数据, 通过整编、分析、处理、质量控制与专家审查, 构建了20大类174个小类的黄河流域基础科学数据资源分类体系, 形成了600多个数据集和143个相关研究模型, 建立了黄河流域专题服务云。数据在国家科学数据中心上线发布后, 截至目前数据访问量已经超过40万人次, 为高等院校、科研院所、企业、政府机构、个人以及国外开展黄河流域研究者提供了数据支持, 形成了黄河流域科学数据开放共享示范实践, 取得了良好的科学数据共享成效。本文阐述黄河流域数据资源建设的现状, 分析存在的问题, 分类建设数据资源, 提出黄河流域生态保护与高质量发展体系化科学数据开放获取的服务方式, 并介绍了数据服务模式及其取得的成效。

关键词: 黄河流域; 科学数据; 数据共享; 生态保护; 高质量发展; 水土保持

DOI: 10.3772/j.issn.1674-1544.2022.01.007

CSTR: 15994.14.issn.1674-1544.2022.01.007

中图分类号: K90

文献标识码: A

The Yellow River Basin Ecological Protection and Development of High Quality Systematic Scientific Data Construction and the Practice

KANG Jianfang^{1,2}, ZHANG Yaonan^{1,2}, WANG Jiayao³, JIA Zexiang⁴, HAN Liqin⁵, LIU Chun³, MIN Yufang^{1,2}, LI Hongxing^{1,2}, WU Yamin², ZHANG Caihe²

作者简介: 康健芳 (1981—), 女, 中国科学院西北生态环境资源研究院工程师, 研究方向为科学数据管理与应用; 张耀南 (1966—), 男, 中国科学院西北生态环境资源研究院研究员, 研究方向为数据工程与应用 (通信作者); 王家耀 (1936—), 男, 院士, 河南省时空大数据产业技术研究院院长, 研究方向为地图学理论及时空大数据应用; 贾泽祥 (1966—), 男, 黄河上中游管理局高级工程师, 研究方向为水土保持科研与管理; 韩立钦 (1983—), 男, 河南师范大学副教授, 研究方向为遥感大数据与应用; 刘春 (1982—), 男, 河南省时空大数据产业技术研究院副教授, 研究方向为时空大数据与软件工程研究; 敏玉芳 (1983—), 女, 中国科学院西北生态环境资源研究院工程师, 研究方向为科学数据管理与应用; 李红星 (1981—), 女, 中国科学院西北生态环境资源研究院工程师, 研究方向为科学数据开放共享与服务; 吴亚敏 (1988—), 女, 国家冰川冻土沙漠科学数据中心助理工程师, 研究方向为科学数据制备与挖掘; 张彩荷 (1994—), 女, 国家冰川冻土沙漠科学数据中心助理工程师, 研究方向为数据分析与处理。

基金项目: 中国科学院信息化专项“寒旱区环境演变研究‘科技领域云’的建设与应用”(XXH13506-103); 中国科学院科学数据中心体系建设项目“中国科学院冰川冻土沙漠数据中心能力建设”(XXH-13514-0209); 国家科技平台专题服务项目“国家冰川冻土沙漠科学数据中心运行建设”(E01Z790201)。

收稿时间: 2021 年 12 月 3 日。

(1. Northwest ecological environment resources research institute, Chinese Academy of Sciences, Lanzhou 730000; 2. National Cryosphere Desert Data Center, Lanzhou 730000; 3. Henan Industrial Technology Academy of Spation-Temporal Big Data, Zhengzhou 450000; 4. The Yellow River administration of aggregate, Xi'an 710000; 5. Henan Normal University, Xinxiang 453007)

Abstract: The Yellow River is mother river of the Chinese nation and birthplace of its civilization. The ecological protection and high-quality development of the Yellow River Basin needs high-quality scientific data covering the whole basin. The data from statistical annual reports, literature documents, government data, surveys, experimental demonstrations, “space-air-ground” observations, numerical simulations, exchanges, etc. are collected, sorted, analyzed, quality controlled by National Cryosphere Desert Data Center with institutions along the Yellow River, to solve the problems of fragmentation, difficult integration and poor sharing of basic scientific data of Yellow River Basin to systematically support scientific research and applications. 20 categories and 174 sub-categories, and a special service cloud has been established under the a classification scheme of basic scientific data resources of the Yellow River Basin. More than 400, 000 times so far access the data after it released online at the National Science Data Center and support the research and decision support on the Yellow River Basin of universities, research institutes, enterprises, governments, individuals and foreign researchers, formed a demonstration of the open sharing of scientific data and achieved good sharing results in the Yellow River Basin. This paper expounds the current situation of the construction of data resources in the Yellow River Basin, analyzes the existing problems, classifies the construction of data resources, puts forward the service mode of open access to systematic scientific teaching data for ecological protection and high-quality development in the Yellow River Basin, and introduces the daata service mode and its achievements.

Keywords: Yellow River Basin, scientific data, data sharing, ecological protection, high quality development, water and soil conservation

0 引言

黄河发源青藏高原巴颜喀拉山，流经青藏高原、黄土高原和黄淮海平原，是世界第五大长河我国第二长河，是中华民族的母亲河中华文明的主要发祥地，全长 5 464 公里，流域面积 75 万多平方公里，流经 9 个省区^[1]（图 1）。习近平总书记在黄河流域生态保护和高质量发展座谈会上强调，黄河流域作为我国重要的生态屏障，是全国重要的经济地带和打赢脱贫攻坚战的重要区域，在我国经济社会发展和生态安全方面具有举足轻重的地位^[2-3]。同长江经济带发展、京津冀协同发展、长三角一体化发展、粤港澳大湾区建设一样，黄河流域的生态保护和高质量发展也是重大国家战略。

在黄河流域生态保护与高质量发展中，黄河上游和源区主要是提升水源涵养功能、增强水汽

循环、恢复生态、提升水资源利用效率和高质量发展，中游主要是做好生态保护修复、水土流失治理、水沙调控、污染防治、水资源利用效率提升和高质量发展，下游主要是保障河道安全、洪水防范、河口湿地保护恢复、近海海水入侵、富营养化和有毒物质治理。按照习近平总书记“共同抓好大保护、协同推进大治理，为全面加强黄河流域生态保护，坚持山水林田湖草生态空间一体化保护和环境污染协同治理”的指示，急需构建能够支持黄河流域生态保护、长治久安、水资源利用、高质量发展和文化保护传承弘扬的基础科学数据与信息化环境，形成支持幸福黄河建设的智能化决策平台，来支持黄河大保护、大治理和高质量发展研究。为黄河大保护、大治理和高质量发展提供“空—天—地—河”一体化的感知手段、科学数据支撑，为黄河流域的保护治理、安全保障、高质量发展与水旱灾预测预警与决策

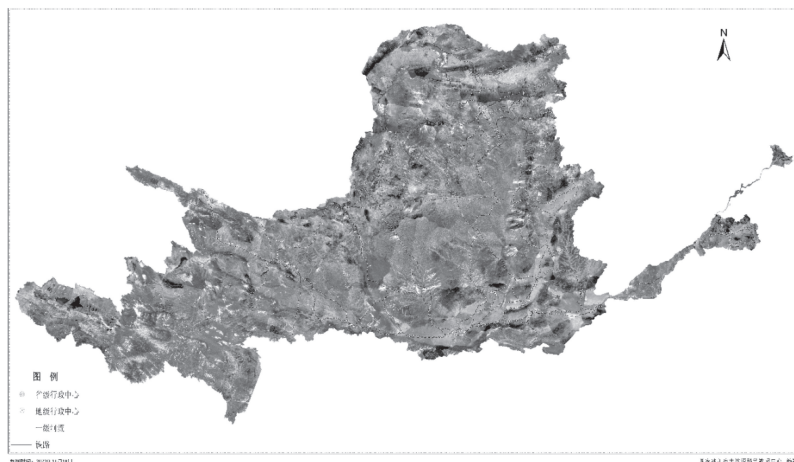


图1 黄河流域高分卫星正射影像图

注：基于高分1、高分2的黄河流域正射影像数据产品。

支持提供坚实的数据基础。从实施的角度来看，黄河流域生态保护、高质量发展、水旱灾害的研究、治理和规划等，都离不开完善的体系化的全流域基础科学数据和相关研究模型的支持。特别是黄河源区水汽循环、水源涵养功能提升、源区产水与蓄水能力的预估、上游生态环境变化预测预警、中游水污染监测与干支河流健康评估预警以及水土流失、滑坡泥石流、洪水溃坝、堤岸崩塌等自然灾害的预测预警与决策支持、下游湿地与河口生态功能预测预警等，都需要覆盖全流域的基础、完整、完善的体系化科学数据和各类研究模型的支持。因此，本文将针对黄河流域基础科学数据体系建设和应用支持服务，围绕黄河流域20大类174子类数据资源，建设数据资源和应用研究模型，构建黄河流域时空数据平台，提供黄河流域科学数据开放共享服务，产生的共享服务成效显著。

1 数据资源分类与建设

1.1 数据资源分类

目前，黄河流域基础科学数据零星散落，在科学数据实时观测获取、科学数据产品研发、科学数据体系化系统集成、基于科学数据的应急响应、决策支撑和服务支持方面，受到缺失、不足与壁垒等制约；流域内涉及的科学研究模型仍存在分散、适应性验证不够、综合集成支持能力低

等不足；针对性应用需要的专题数据制备关键方法技术缺失，不具备实时现场专题数据生产能力，在线流域中数据驱动、人工智能融合模型进行环境现势性数据制备与状态评估、发展预测预警的可用性模型综合支持能力不足。为此，国家冰川冻土沙漠科学数据中心开展了黄河流域科学分类建设、体系化科学数据收集加工和专题数据制备，完成了黄河流域自然科学数据、人文经济社会数据，集成验证了解黄河流域各类研究模型，并针对有效支持水土流失与治理、水污染与水沙调控、支流可持续承载力评估、河道安全与洪水灾害、岸线生态保障修复、河口湿地保护、文化保护传承弘扬、新旧动能转换以及经济社会高质量发展等研究和决策需要，将黄河流域基础科学数据划分为20大类174个小类。20个大类主要包括：黄河流域基础地理数据、黄河流域气候数据、黄河流域水文水资源数据、黄河流域冰川冻土积雪数据、黄河流域湿地数据、黄河流域草地数据、黄河流域水土保持治理数据、黄河流域沙地数据、黄河流域生态环境保护数据、黄河流域自然资源数据、黄河流域水利水电工程数据、黄河流域社会发展数据、黄河流域经济数据、黄河流域文化旅游数据、黄河流域农业数据、黄河流域畜牧业数据、黄河流域工业数据、黄河流域城镇发展数据、黄河流域承载力数据、据黄河流域自然灾害数据。具体数据体系结构见图2。



图2 黄河流域生态保护与高质量发展数据体系结构

1.2 数据资源建设

国家冰川冻土沙漠科学数据中心（以下简称“中心”），联合河南大学、河南省时空大数据产业研究院、黄河上中游管理局、河南师范大学、甘肃省祁连山水源涵养林研究院、甘肃省祁连山研究中心，针对黄河流域划分的20大类基础数据资源体系，基于历年统计年报资料、历史文献文档资料、政府公开数据、“空一天一地”观测

数据、调查考察数据、试验示范数据、数值模拟数据、交换交流数据等，建立数据整理流程、质量控制流程，利用“空一天一地一河”数据融合的方法对这些数据资源进行挖掘、分析、处理、加工、质控，形成时空基准一致、规范化的数据资源。

考虑到数据来源不同、结构不同、格式不同，本文分析其关联性，对其进行分类处理。并

参照数据分类,根据合作单位提交的数据实体,制定数据标准体系规范,对数据实体、元数据标准、数据分类规范等内容的建设进行约束。

1.2.1 数据分类整编

从多尺度、多方位对包括野外观测数据、互联网数据、遥感对地观测数据、视频图片、文献数据、数值模拟数据和测试分析数据等进行如下整编。

(1) 黄河流域基础地理数据。包括流域基础地理,水土保持基础边界、地质图、地表覆盖等数据。

(2) 黄河流域基础卫星遥感数据。包括流域高分正射影像、Landsat8、MODIS、哨兵1号等数据。

(3) 黄河流域地形因子和水文分析数据。包括流域地形起伏度、地表切割深度、坡向、平面曲率、坡向变率、坡度、流量、水流长度、河流网络、流向等数据。

(4) 黄河流域气候气象数据。包括流域月度降水量、平均降水量、月度气温、气温平均空间分布、湿润指数、气象灾害、风沙天数统计等数据。

(5) 黄河流域水系及水文水资源数据。包括黄河流域三级河流水系、河流湿地、湖泊湿地、沼泽湿地、水库池塘湿地、养殖池塘湿地、人工水渠湿地、近海与海岸湿地、盐田湿地、冰雪湿地、上游河段汇入支流数据、中游河段汇入支流数据、下游河段汇入支流数据、水资源流域分区、市域降水量、水资源量统计等数据。

(6) 黄河流域水利水电工程数据。包括流域水电站、水利枢纽、水库等数据。

(7) 黄河流域水土流失与水沙变化数据。包括流域水土流失、极强度水蚀分布、沙漠化、石漠化、水土流失、盐渍化、泥沙观测、水沙变化、淤地坝、降雨侵蚀因子、土壤侵蚀因子、平均植被净初级生产力等数据。

(8) 黄河流域生态环境数据。包括流域河湖、水面、湿地分布、自然保护区分布、植物多样性、动物多样性、生态脆弱区、水源涵养保护

区、生态环境受灾情况、黑臭水体分布及综合整治规划、主要河道水环境容量测算、饮用水水源地保护规划、各县区生态红线、土壤污染状况、重要污染源环境统计、县区垃圾处理、沙漠分布、冰川分布、冰湖分布、阿尼玛卿山冰川监测等数据。

(9) 黄河流域土地及矿产资源数据。包括流域多年土地利用类型分布、土壤、土种志、耕地分布、林地分布、牧草地分布、森林分布、灌丛分布、草地分布、湿地分布、生态公益林、矿产资源储量、有色金属资源分布、煤田分布、稀土资源分布、石油天然气资源分布、矿产资源生产基地分布等数据。

(10) 黄河流域自然灾害数据。包括流域近几年洪水灾害、旱灾、地质灾害、抗震薄弱区、冰凌洪水、地震、地质环境等数据。

(11) 黄河流域社会经济数据。包括流域人口、农业、工业、城乡和旅游等数据。

(12) 黄河流域城市规划数据。包括流域经济开发区总体规划、近期建设规划、城市控制性详细规划、新农村建设规划、城乡统筹规划、乡村振兴规划、美丽乡村规划等数据。

(13) 黄河流域文化旅游数据。包括流域物质文化遗产、非物质文化遗产、古都文化、红色文化、文化保护、各级风景名胜区分布等数据。

(14) 祁连山专题数据。包括祁连山排露沟流域生态监测、祁连山老虎沟冰川监测、七一冰川监测、八一冰川监测、天老池流域监测、大野口流域监测、祁连山植被分布、祁连山土壤有机碳密度、祁连山平均干燥度指数、归一化植被指数等数据。

(15) 黄河流域中上游水土保持数据。主要包括西峰、天水、绥德等3个观测站和检测中心的径流泥沙、降雨、径流场、标准小区长期观测数据及植被和下垫面的调查报告数据。黄河流域重点地区的土地利用、植被覆盖和土壤侵蚀等数据资源。

(16) 黄河流域支流数据。目前中心制备了渭河流域基础地理、遥感专题、土壤、自然状况

统计、水环境情况统计、大气环境统计、固体废物统计、自然生态统计、土地利用情况统计、林业资源状况统计、自然灾害及突发事件统计、环境与投资状况统计、城市环境状况统计、农村环境状况统计、气象站点监测（日）、空气质量监测（日）等数据。

1.2.2 数据实体建设

合作单位提交的数据实体形成了黄河流域基础科学数据实体。其数据类型主要包括矢量数据、属性数据、栅格数据、文本数据等类型。根据某种数据类型的特有组织方式形成统一格式进行建设和共享，提供共享使用数据集。

（1）矢量数据：以坐标或坐标串表示的空间点、线、面等图形数据及与其相联系的有关属性数据的总称^[4]。在矢量数据结构中，不同类型数据的描述方法不同，点数据以坐标值描述，线数据以均匀或不均匀间隔的顺序坐标链来描述，面状数据以边界线来描述^[5]。矢量数据主要包括如SHP、MIF和等EOO在内的全球通用GIS软件的交换格式。黄河流域边界、黄河流域各区县等就属于此类数据。

（2）属性数据：一般指通过对某个数据项的数值或文字的描述而形成的一系列的事务记录。属性数据的格式一般包括ACCESS、Excel和ORACLE等在内的通用关系数据库格式。黄河流域数据集中人文经济统计类数据就是此类数据。

（3）栅格数据：由正方形或者矩形栅格点组成，每个栅格点或者像素的位置由栅格所在的行列号来定义，所对应的数值为栅格所要表达的内容的属性值^[6]。栅格数据的格式主要包括通用的图像、影像数据格式，如TIF、JPEG、MPEG和BIL等。黄河流域各地形因子数据就是此类数据。

（4）文本数据：用文本形式记录的黄河流域基础科学数据，大多采用的格式有Word格式、HTML格式和纯文本格式等。黄河流域城乡规划相关数据就是Word版的文本数据。

1.2.3 元数据与数据标识

元数据又被称为“数据的数据”，是关于数据的内容、质量、状况和其他特性的信息^[7]，用

来描述数据的内容、质量状况、时空范围、管理方式、所有者、共享引用方式等有关信息。黄河流域科学数据元数据定义了所有数据集的共性信息，提供数据集的名称、标识、内容摘要、基本信息（包括采集时间、采集地点、数据量、数据格式、坐标系等）、数据集缩略图、数据来源、数据加工方法、数据质量控制方法、项目支持信息和引用方式等信息，支持数据共享、编目、元数据交换与检索以及数据导航服务等。

通过元数据为黄河流域数据资源管理、数据理解、访问和共享模式提供支持，并通过元数据技术实现数据的汇交、审核发布、共享、管理等功能。结合中文标识CSTR和英文标识DOI，支持用户能快速准确找到数据并获取，通过元数据与CSTR标识符保证各参与单位自身的权益。数据的可获取性是所有数据都可以通过网络来获取，并通过不同的用户账号来控制用户对数据的访问。为了更好地让用户了解已经形成的数据集，中心利用数据出版方式介绍所形成的数据集。数据提供者的权益保护主要是通过数据的出版和引用体现^[8]，规定了数据集引用格式，要求所有数据使用用户，要在正式发表的成果中进行数据引用或致谢。

2 服务方式

黄河流域生态保护与高质量发展体系化科学数据通过国家冰川冻土沙漠科学数据中心进行开放共享。在数据中心建立了黄河流域数据服务平台，以专题数据集提供服务。目前已经形成了黄河流域高质量发展数据专题、水土保持数据专题、祁连山生态环境要素专题、渭河流域专题等开展服务，并为水土保持建立了水土保持专题服务云平台。

2.1 黄河流域高质量发展数据专题

中心联合河南大学、河南省时空大数据产业研究院、河南师范大学收集整编，包括黄河流域基础地理，多源遥感影像，冰川、冻土、积雪、湖泊、沙漠、湿地、森林、草地、农田等专题数据，黄河流域气候气象，黄河流域河流水资源数

据,黄河流域洪水、旱灾、泥石流、滑坡等自然灾害数据,黄河流域经济社会、工农业产业、旅游文化数据资源等数据。近期还发布了黄河流域最大分支渭河流域体系化数据资源,包括渭河流域基础地理、遥感专题、土壤、自然状况统计、水环境情况统计、大气环境统计、固体废物统计、自然生态统计、土地利用情况统计、林业资源状况统计、自然灾害及突发事件统计、环境与投资状况统计、城市环境状况统计、农村环境状况统计、气象站点监测(日)、空气质量监测(日)等数据集。截至2021年12月,该专题数据已发布114个数据集,文件数据量为17.2 TB,数据访问量为78 880人次。

2.2 黄河流域水土保持数据专题

中心联合黄河水利委员会黄河上中游管理局、水利部黄土高原水土保持野外科学观测研究站(天水站、西峰站、绥德站),系统梳理了3个水土保持科学试验站观测数据,形成了长时间序列的水土保持科学数据资料,时间范围为1942—2018年,共62年。从工作之初的建立工作规范、系统人员培训,到形成工作流程,直到完成数据整理,整个梳理工作历时约3年。形成的高质量数据主要包括绥德、天水 and 西峰3个空间跨度大的区域(总面积约为566.23 km²)共15个典型小流域的观测数据。空间组成具体为:由绥德的王茂沟、韭园沟、桥沟、裴家峁、桥沟一支沟、桥沟二支沟和辛店沟等7个小流域组成的黄土丘陵沟壑区第一副区;由天水的吕二沟、罗玉沟、桥子东沟和桥子西沟等4条小流域组成的黄土丘陵沟壑区第三副区;由西峰的杨家沟、南小河沟、砚瓦川和董庄沟等4个小流域组成的黄土高原沟壑区。从数据内容方面来看,包含流域内基础地理信息、植被覆盖情况、土地利用方式、水土保持措施类型、坡度、整地方式等,也包括径流小区的降水量、泥沙量、径流量、土壤含水率等气象水文土壤综合实地观测数据^[9]。截至2021年12月,该专题数据已发布386个数据集,文件数据量为11.8 GB,数据访问量为234 228人次。具有区域代表性强、数据时空跨

度大序列长、体系设置科学完整的资源优势。

2.3 祁连山生态环境指数专题

中心联合甘肃省祁连山生态环境研究中心、甘肃省祁连山水源涵养林研究院、青海省环境监测厅等单位,利用遥感监测、站点实测、采样分析、模式模拟等手段,分析集成了祁连山地区基础地理背景数据,气候植被相关遥感数据,典型流域气象、水文参数实测数据,不同海拔梯度林、灌、草样地调查数据,祁连山及其周边地区关键土壤属性实验分析数据,祁连山地区水—土—气—生动态监测和模拟长时间序列时空数据,建立了祁连山地区系列生态指数及环境要素专题数据,揭示了祁连山生态环境演变规律,探究了祁连山地区生态安全评估,支持祁连山生态保护与可持续发展。截至2021年12月,中心整编发布了40个数据集,数据量为80.4 GB,数据访问量为44 539人次。

2.4 水土保持专题服务云平台

水土保持专题服务云平台(<http://www.ncdc.ac.cn/portal/soil-and-water-conservation>)采用Python语言开发,通过黄河流域3个典型区域水土保持数据库、动态监测管理体系的建设,形成典型水土流失观测区GIS空间数据管理能力,根据流域规划设计指标,可快速准确完成小流域水土保持规划,有效提高了观测区资源信息利用率,实现对黄河上游水土保持试验区的观测数据的综合管理和空间分析展示(图3)。水土保持专题服务云平台包含黄河流域地级市、干流、湖泊、省会城市、直流、重点支流以及粗泥沙集中来源区界、粗沙区界、多沙粗沙区界、黄河流域地区界、黄河流域界、黄河流域省界、黄河跨流域县界、黄河流域支流界、黄土高原地区界、水土流失类型区界、河龙区间多沙粗沙国家级重点治理区、黑河绿洲国家级重点预防区、湟水洮河中下游国家级重点治理区、泾河北洛河上游国家级重点治理区、六盘山国家级重点预防区、伊洛河三门峡库区国家级重点治理区、子午岭国家级重点预防区、祖厉河渭河上游国家级重点治理区、粗泥沙集中来源区界、粗泥沙区界、多沙粗沙区县界、

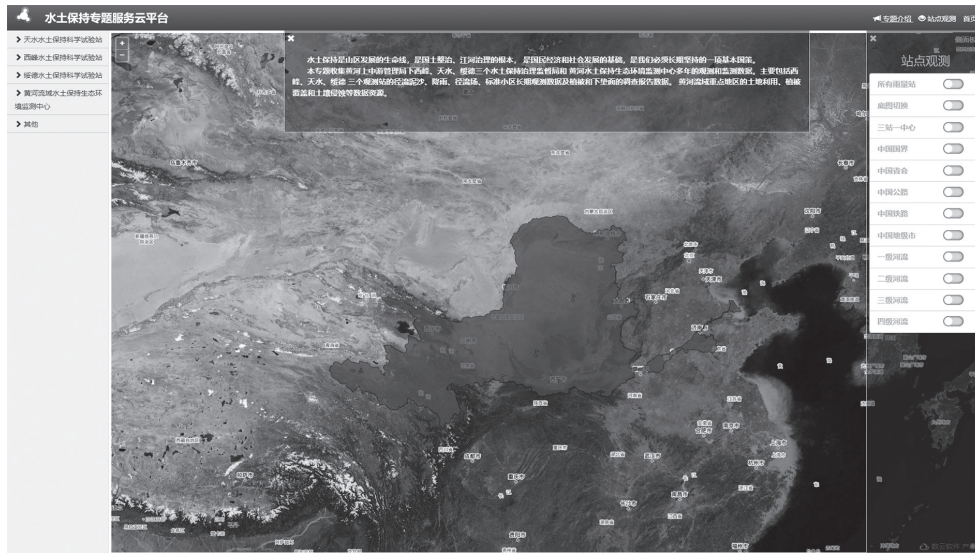


图3 水土保持专题服务平台

多沙粗沙区小流域界、多沙粗沙区支流界、中游多沙粗沙区界等信息。

3 数据服务与成效

国家冰川冻土沙漠科学数据中心面向科技计划项目、科研团体及个人提供冰川、积雪、冻土、沙漠及其作用区相关各类科学数据的汇交、发布、共享服务，通过规范化整理和评审机制提升科学数据质量，在保护数据生产者权益的基础上促进、深化数据再利用，扩大数据影响力，提供数据长期维护、管理及存储服务，确保数据安全及永久访问。数据共享平台的建设是构建黄河流域生态保护与高质量发展科学数据资源发布及共享服务的重要支撑条件。中心在已有数据共享平台的基础上，扩展数据管理、制备、可视化及分析服务功能。基于定义的时空基准、存储格式规范标准，构建数据存储和管理系统，集成关键环境因子在线制备注册功能，实现空间数据、社会经济等数据的可视化展示等功能。中心设置了3个专题数据和1个专题数据云平台，以帮助用户快速查找、浏览和使用黄河流域基础数据。黄河流域数据产品采用全方位的数据知识产权措施，设置数据保护期，确保数据作者的权益，遵循FAIR（可发现性、可公开获取性、可互操作性、可重用性）原则，主要以数据提供者的要求

为准，采用在线、离线服务的共享模式。

3.1 数据服务

（1）数据资源共享方式。为用户提供了完全共享、在线下载和离线申请3种数据共享方式，根据不同的数据共享权限，为用户提供服务。为数据集提供CSTR和DOI注册服务，体现数据的价值，这也在一定程度上体现了数据的质量和用户的认同度。

（2）数据提交和审批服务。在数据整合过程中，以数据资源为主体，规范数据资源提交、发布、共享流程，使数据提交者自主管理数据，充分体现数据提交者的权益。数据提交者必须填写完整的元数据信息，由中心组织专业人员对提交的元数据进行审核，审查通过后发布。为了促进研究人员个人的数据汇集，设计数据作者自主进行数据提交和对其提交数据共享服务的审批权，在保障数据作者权益的前提下，推进数据汇聚和共享应用。

（3）数据下载流程。通过构建元数据、数据、用户之间的对应关系，以实现数据的快速共享服务功能。用户通过数据搜索、导航等途径找到感兴趣的数据并查看数据的详细信息即元数据后，就可以决定是否要申请该数据。黄河流域科学数据采用离线申请和在线申请两种服务方式。离线申请方式的数据申请需要经过数据作者人工

审核后才能对外提供服务,而在线申请方式的数据申请时,只需简单填写用途就可以转入下载或者使用FTP下载,从而获得数据。数据服务流程如图4所示。

3.2 数据服务的成效

黄河流域近600个数据集143个相关研究模型上线发布后,截至2021年12月,数据访问量超过了40万人次,为高等院校、科研院所、企业、政府机构、个人以及国外开展黄河流域研究者提供了数据支持,并形成了基于数据驱动、模型支持融合人工智能的数据反演与生产系统,该

系统已广泛用于黄河流域生态灾害、地质灾害、气候灾害的早期发现与识别分析中,为自然灾害的预测预警奠定了基础。同时,采用基于网络、文献、期刊中数据提取的人工智能挖掘原型技术,初步形成了基于“空—天—地—河”数据驱动的地质灾害发现识别和分析原型技术,为黄河流域生态保护和高质量发展的数据赋能支持奠定了良好的基础。

根据国际《湿地公约》等规范,结合黄河流域地理特征及湿地类型,中心开展了黄河湿地分析服务(图5)。黄河流域冰川、积雪主要分布

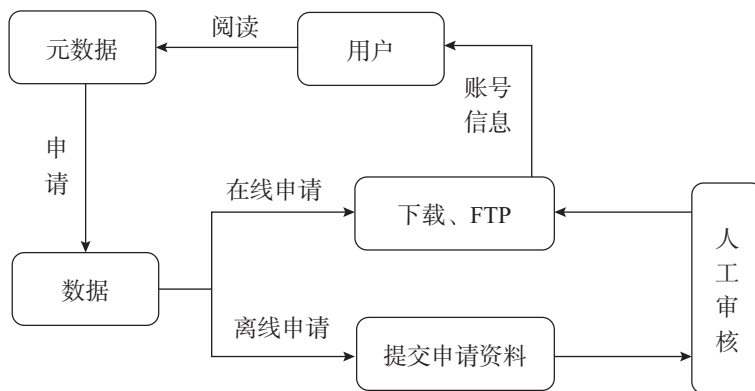


图4 专题数据下载服务流程

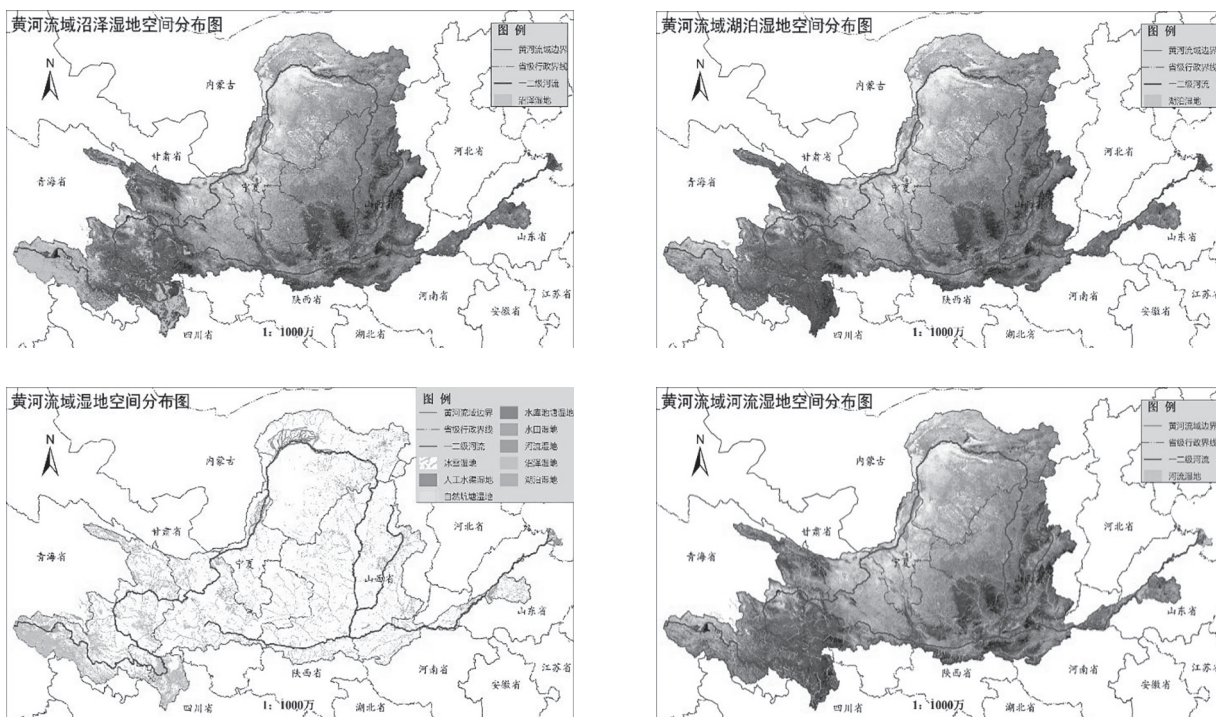


图5 黄河流域河流、沼泽、湖泊等多种类型湿地空间分布

于三江源地区,进入 21 世纪以来,黄河流域阿尼玛卿、祁连山等冰川退缩剧烈。黄河流域河流湿地主要分布于黄河上、中游地区的三级以上支流,青海、甘肃地区河网密度最大、下游地区支流较少。黄河流域湖泊湿地主要分布于黄河上、中游地区的三江源区,以冰雪冻融型湖泊为主,中游地区有零星分布。黄河流域水库池塘湿地以人工建造为主,主要分布于黄河中游地区的黄河干流及其一级支流。黄河流域沼泽湿地主要分布于黄河上游地区,以高寒草甸、河湖沼泽为主,中游和下游地区主要沿黄河及其支流河道分布的林木和草地沼泽。黄河流域自然坑塘湿地主要分布于黄河上、中游地区,在高分影像上 18 m² 的融水坑塘也能清晰地表现出来。黄河流域水田湿地主要分布于黄河中、下游地区,以引用黄河水灌溉的水稻田为主,河套平原、黄河下游一带都有分布。黄河流域人工水渠湿地以人造灌溉设施为主,主要分布于黄河中、下游地区,以引用黄河水灌溉、养殖业为主。

4 结论与展望

黄河流域生态保护和高质量发展数据建设完成了第一阶段目标任务,形成了从数据资源体系梳理、数据收集、规范化整理、集成挖掘到数据服务的机制,建成了数据服务平台,集成了一批黄河流域基础科学数据,为众多科研用户及政府决策部门提供了数据服务。但同时也反映出一些客观问题,如数据体系性不完整、数据时间序列不够长、数据覆盖度还不完全、数据集成程度还不高、要素覆盖面还有缺失、“空一天一地一河”数据一致性还需要改进、数据集产品质量还有待提高、不同部门机构之间的数据引接还存在机制上的障碍、沿河流域机构数据共享仍然是一难题等。数据的不系统、不完整、难整合、难共享等不足,致使数据驱动的科学研究的、赋能规划

发展、支持企业应用等数据支持力不足,并极大地限制了生态环境保护与高质量发展研究与决策的提升,造成了科技资源极大浪费。中心近期将在提高数据集成的广度和深度上继续开展工作,尝试与其他黄河流域数据生产单位合作。在 2021 年 6 月在甘肃省天水市举行的“科学数据与数字经济”研讨会上,王家耀院士代表相关参会单位宣读了形成《黄河流域基础科学数据开放共享倡议书》,号召更多的黄河流域数据生产单位加入“黄河流域科学数据共享联盟”,形成黄河流域分布式互联数据中心来形成全部覆盖的科学数据资源体系,提升科学数据应用保障能力,支持黄河流域生态保护和高质量发展。

参考文献

- [1] 张敏,吕艳荷,刘磊.黄河流域水生态保护的主要问题与对策建议[J].四川环境,2021,40(5):157-161.
- [2] 宋杨.打好黄河流域环境问题整治攻坚战[N].中国环境报,2021-10-29(3).
- [3] 方琳娜,尹昌斌,方正,等.黄河流域农业高质量发展推进路径[J/OL].中国农业资源与区划:1-8[2021-11-02].<http://kns.cnki.net/kcms/detail/11.3513.S.20211020.1508.021.html>.
- [4] 程鹏飞,成英燕,文汉江,等.2000 国家大地坐标系实用宝典[M].北京:测绘出版社 2008.
- [5] 蔡佳男,耿庆斋.水利科学数据共享汇交体系探索与构建[J].中国水利水电科学研究院学报,2006(3):31-35.
- [6] 王存刚,姚文琳.基于 Ontology 的智能信息检索系统研究[D].青岛:中国海洋大学,2006
- [7] 朱星明,张行南,白婧怡,等.水利科学数据共享元数据理论的应用探讨[J].水利学报,2005,36(8):56-59.
- [8] 王亮绪,吴丽宗,李红星,等.面向黑河流域生态水文过程集成研究的科学数据汇交与管理[J].遥感技术与应用,2013(6):362-369.
- [9] 张耀南.《黄土高原水土保持专题》卷首语[J].中国科学数据(中英文网络版),2021,6(3):6-7.