

2020年度报告

CSTR:16666.11.nbsdc.report2020



2021年度报告

CSTR:16666.11.nbsdc.report2021



2022年度报告

CSTR:16666.11.nbsdc.report2022

微信公众号

官方网址

中国科学院
计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

公共事务邮箱
nbsdc@cnic.cn

数据汇交邮箱
datarchive@cnic.cn

联系电话
010-58813784

地址
北京市海淀区东升南路2号院
中国科学院信息化大厦



www.nbsdc.cn

国家基础学科公共科学数据中心

National Basic Science Data Center

NBSDC 2022年度报告

目录

中心总体情况

中心概况 01
目标定位 02
参建单位 03
组织体系 04
发展历程 05
数据资源 07

年度工作进展

年度亮点 09
国际权威数据库 11
国家重点研发计划科学数据汇交 13
项目汇交数据资源特征图 · 学科领域 15
项目汇交数据资源特征图 · 专项 17
创新应用课题一：“新能源汽车”数据分析与共享方案研究 19
创新应用课题二：“纳米科学”数据标准与共享应用策略研究 21
政策和标准体系 23
特色服务与数据软件 25
领域特色工具 27
服务体系 29
合作推广 31

数据资源与服务

科学数据库概况 33
交通运输分中心 35
信息学科分中心 37
世界山茶属植物品种数据库 39
图像理解机器学习数据库 41
环境健康主题数据库 43
中国海岸带植物资源数据库 45
医学物理数据库 47
中国动物主题数据库 49
中国植物主题数据库 51
中国植物DNA条形码数据库 53
化学主题数据库 55
原子分子数据库 57
光学技术数据库 59
核物理主题数据库 61
宇宙学数值模拟数据库 63
VLBI月球与深空探测数据库 65
中国病毒资源基础数据库 67
载人航天空间科学与应用数据库 69
材料科学基础数据库 71
干细胞与代谢疾病数据库 73
中国活体人脑成像共享数据库 75
地理空间数据库 77
干旱区药用资源数据库 79

附录

年度大事记 81
2022年度代表性论文 83
2022年度数据标引代表性论文 84
典型数据库/集 85
运行保障 93

中心总体情况

中心概况

国家基础学科公共科学数据中心（简称“国家基础数据中心”）是中国科学院计算机网络信息中心联合中国科学院、教育部、交通运输部、工信部、国防科工局、国家林草局、黑龙江省等下属研究所40余家单位共同建设的国家科技资源共享服务平台，旨在汇集管理物理、化学、材料、动物、植物、交通、信息科学等基础学科领域，青海湖、黑龙江、新疆等典型区域长期科研活动积累的科学数据，以及相关基础领域政府预算资金支持项目汇交的科学数据。经过多年的持续发展，国家基础数据中心已经形成了完善的基础学科数据资源体系，以及支持分布式科学数据资源统一管理、集成融合、分析挖掘和应用服务的技术体系、标准体系和服务体系，有力支撑了数据驱动的基础学科原始创新、多学科交叉融合创新、国民经济发展、人民生命健康和国际交流。

目标定位

基础学科数据资源汇聚中心

科学数据自主软件研发基地

交叉学科融合研究分析平台

数据科学理论研究实践基地

参建单位



中国科学院计算机网络信息中心
交通运输部公路科学研究院
交通运输部天津水运工程科学研究院
北京大学
北京师范大学
吉林大学
深圳大学
西南大学
烟台大学
浙江海洋大学
东北农业大学
北京航空航天大学
黑龙省科技资源共享服务中心
黑龙江省科信科技信息中心
青海湖国家级自然保护区管理局
北京应用物理与计算数学研究所
中国原子能科学研究院
吉林省肿瘤医院
中国科学院植物研究所
中国科学院动物研究所
中国科学院过程工程研究所
中国科学院化学研究所
中国科学院金属研究所
中国科学院昆明动物研究所
中国科学院上海天文台
中国科学院生物物理研究所
中国科学院文献情报中心
中国科学院心理研究所
中国科学院烟台海岸带研究所
中国科学院长春应用化学研究所
国家纳米科学中心
中国科学院上海生命科学研究院
中国科学院成都生物研究所
中国科学院城市环境研究所
中国科学院广州生物医药与健康研究院
中国科学院合肥物质科学研究院
中国科学院紫金山天文台
中国科学院空间应用工程与技术中心
中国科学院昆明植物研究所
中国科学院上海有机化学研究所
中国科学院水生生物研究所
中国科学院武汉病毒研究所
中国科学院新疆理化技术研究所
中国科学院长春光学精密机械与物理研究所

组织体系

科技部

中国科学院

国家科技基础
条件平台中心

中国科学院
计算机网络信息中心

专家
委员会

用户
委员会

国家基础数据中心

数据资源发展组

技术与安全组

共享与服务组

数据汇交组

数据标准组

国家基础学科公共科学数据中心
National Basic Science Data Center

发展历程

基础学科数据是夯实基础科学研究“地基”、构筑国家核心竞争力“大厦”的重要资源。国家基础数据中心是我国科学数据事业的主要建设者之一，立足44个一级学科数据资源，聚合基础数据、制定通用标准、研发共用技术、创新服务模式，推动基础学科数据开放共享与应用服务，引领大数据驱动的科研模式创新。

2022
服务1500多个国家重点研发计划项目数据汇交

2022
入选“中国科学院科学数据工作十大进展”

2021
信息学科分中心启动建设

2021
入选国家“十三五”科技创新成就展
同年，入选中国科学院“十三五”信息化成果展

2020
启动科技项目数据汇交
服务我国基础学科领域国家级项目数据汇交工作，促进科技资源标识和科学数据开放共享

2019
国家基础学科公共科学数据中心
《国家科技资源共享服务平台优化调整名单的通知》（国科发基〔2019〕194号）

2017
国家基础科学数据共享服务平台
《国家科技资源共享服务平台绩效考核与评估结果的通知》（国科发基〔2017〕24号）

2010
科技数据资源整合与共享工程
中国科学院“十二五”信息化专项

2009
基础科学数据共享网 – 理化天文空间生物（一期）
国家科技基础条件平台建设项目

2003
基础科学数据中心
“国家科学数据共享工程”重点项目

2001
科学数据库及其应用系统
中国科学院“十五”信息化建设重大专项

2022
设立年度数据应用创新课题
“‘新能源汽车’数据分析与共享方案研究”、“‘纳米科学’数据标准与共享应用策略研究”入选国家基础数据中心2022年度数据应用创新课题

2022
信息学科分中心成立

2021
服务800多个国家重点研发计划项目数据汇交

2021
与中国人工智能协会共建多语种信息资源
联合人工智能学会多语种智能信息处理专业委员会和《中国科学数据》，推出多语种智能信息处理数据专刊

2021
交通运输分中心成立
交通运输部公路科学研究院、交通运输部天津水运工程科学研究院等单位联合建设

2019
中国科学院基础学科公共科学数据中心
《中国科学院科学数据中心体系第一批科学数据中心名单》

2017
GB/T 35294–2017 信息技术 科学数据引用
GB/T 34945–2017 信息技术 数据溯源描述模型
牵头发布两项国家标准

2016
科学大数据工程
中国科学院“十三五”信息化专项

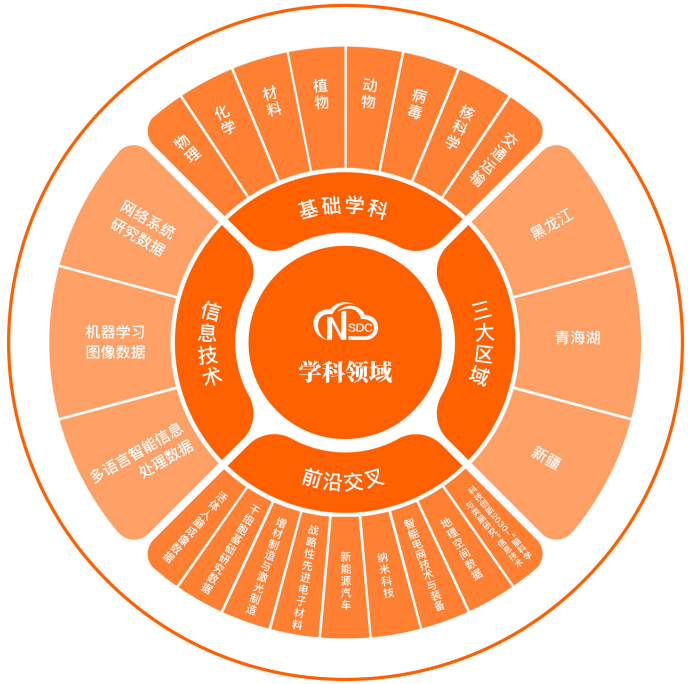
2013
基础科学数据共享网 – 理化天文空间生物（二期）
持续发展资源运行服务

2006
数据应用环境建设与服务
中国科学院“十一五”信息化专项

2004
基础科学数据中心及共享服务网
国家科技基础条件平台重点项目

2002
CODATA中国理化数据库
国家科技基础性工作专项

数据资源



国家基础数据中心系统整合了物理、化学、材料、动物、植物、交通和信息等学科领域，以及青海湖、黑龙江和中亚地区等典型区域长期科研活动中积累的基础科学数据。2022年度持续汇聚科学数据，数据资源总量达2.71PB，覆盖44类一级学科。其中自建数据量累计达1.25PB，项目数据汇交数据量累计达1.46PB。在完善的技术体系、标准体系支撑下构建了基于云门户的“数据服务–信息服务–知识服务”多元化服务体系，有效促进了基础科学数据的资源汇聚、开放共享、多学科交叉融合分析和应用。

一级学科 44类

主题数据库 23个

数据集 10,227个

数据总量 2.71PB

自建数据资源总量 1.25PB

项目汇交数据量 1.46PB

数据记录 10亿条+

参建单位 44个

分中心和数据库列表

序号	分中心/数据库	承建单位	参建单位
1	交通运输分中心	交通运输部公路科学研究院	交通运输部天津水运工程科学研究院
2	信息学科分中心	中国科学院计算机网络信息中心	
3	世界山茶属植物品种数据库	中国科学院昆明植物研究所	
4	图像理解机器学习数据库	中国科学院合肥物质科学研究院	
5	环境健康主题数据库	中国科学院城市环境研究所	
6	中国海岸带植物资源数据库	中国科学院烟台海岸带研究所	烟台大学　浙江海洋大学
7	医学物理数据库	吉林大学	吉林省肿瘤医院　北京航空航天大学 深圳大学
8	中国动物主题数据库	中国科学院动物研究所	中国科学院昆明动物研究所　中国科学院上海生命科学研究院 中国科学院成都生物研究所　中国科学院武汉水生生物研究所
9	中国植物主题数据库	中国科学院植物研究所	中国科学院文献情报中心
10	中国植物DNA条形码数据库	中国科学院昆明植物研究所	
11	化学主题数据库	中国科学院过程工程研究所	中国科学院上海有机化学研究所 中国科学院长春应用化学研究所
12	原子分子数据库	北京应用物理与计算数学研究所	
13	光学技术数据库	中国科学院长春光学精密机械与物理研究所	
14	核物理主题数据库	中国原子能科学研究院	
15	宇宙学数值模拟数据库	中国科学院紫金山天文台	
16	VLBI月球与深空探测数据库	中国科学院上海天文台	
17	中国病毒资源基础数据库	中国科学院武汉病毒研究所	
18	载人航天空间科学与应用数据库	中国科学院空间应用工程与技术中心	
19	材料科学基础数据库	中国科学院金属研究所	中国科学院化学研究所　国家纳米科学中心
20	干细胞与代谢疾病数据库	中国科学院广州生物医药与健康研究院	
21	中国活体人脑成像共享数据库	北京师范大学	中国科学院心理研究所　北京大学 中国科学院生物物理所　西南大学
22	地理空间数据库	中国科学院计算机网络信息中心	
23	干旱区药用资源数据库	中国科学院新疆理化技术研究所	

年度亮点

科学数据作为国家战略资源，能够有力支撑科学研究、科技创新。2022年度，国家基础数据中心持续推动基础学科领域科学数据宏观管理与发展，为交通运输分中心和信息学科分中心授牌、承建中国科学院基础学科公共科学数据中心、入选中国科学院科学数据工作十大进展，联合学术组织、期刊、数据平台构建数据出版社区等。

1. 入选“中国科学院科学数据工作十大进展”

国家基础数据中心联合中国科学院昆明植物研究所申请的“世界山茶属植物品种数据库荣获第十三届中国茶花博览会科技创新类金奖”入选2022年度中国科学院科学数据工作十大进展。这是国家基础数据中心连续第二年入选“中国科学院科学数据工作十大进展”。



2. 交通运输分中心和信息学科分中心正式授牌

2022年3月，国家基础数据中心年度工作会上，科技部基础司、国家科技基础条件平台中心、交通运输部等相关领导出席，为国家基础数据中心交通运输分中心和信息学科分中心授牌。





3. 承建的中国科学院学科中心通过认定

2022年3月，中国科学院基础学科公共科学数据中心启动会顺利召开。中国科学院科学传播局领导宣读了中国科学院科学数据中心认定通知并授牌，中国科学院基础学科公共科学数据中心正式成立，致力于全院基础学科领域科学数据的管理与服务。



4. 设立“新能源汽车”、“纳米科学”2项数据应用创新课题

2022年7月，国家基础数据中心自筹经费启动数据应用创新课题征选，目标是鼓励中青年科技工作者充分利用国家基础数据中心汇交的“十三五”国家重点研发计划重点专项数据资源开展高水平学术研究。经过面向全国公开发布征选指南、初选和专家评审答辩会，最终设立“‘新能源汽车’数据分析与共享方案研究”、“‘纳米科学’数据标准与共享应用策略研究”2个课题，取得积极研究进展。



5. 创新资源发展模式，共建数据社区并出版数据专刊

·联合科学数据银行共建数据社区

2022年6月，国家基础数据中心与科学数据银行共建论文关联数据社区正式上线，基础学科领域科研工作者可通过该社区实现国内外各类论文关联数据的标识获取、永久存储、共享及发布。



·联合中国人工智能学会推出数据论文专刊

2022年6月，国家基础数据中心联合人工智能学会多语种智能信息处理专业委员会推出的《多语种智能信息处理》数据论文专刊正式在《中国科学数据》发布。本专题遴选了10余篇高质量数据论文，包括多语种的词法分析、机器翻译、阅读理解、日常问答、语音翻译、语音识别、语音合成、翻译校正、文本剽窃检测等多个领域的数据集，力争打造具有国际影响力的人工智能多语种数据集，提升我国多语种智能信息领域的国际影响力。



世界山茶属植物品种数据库

https://camellia.iflora.cn

Thanks to that dictionary, a user can search a cultivar by any name (Latin, Japanese, Chinese). Another section includes the albums of the illustrations and pictures uploaded by collaborators. Each illustration is also described by metadata concerning copyright, location, etc. Thanks to these features, ICR surpasses by far other camellia databases, not only in terms of quality, but also in terms of the number of illustrations and descriptions.

— Prof. Gianmario Motta,President ofInternational Camellia Society

As Director of the ICS-Benelux division and the editor of international Camellia Journal (ICJ), I was aware of the work to create a Camellia DB. During the ICS Congress in Nantes/France, the additional information about frequently used synonyms is very handy, to find the correct scientific name of a Camellia. Beside the search options for Camellias, it's also nice to see recently uploaded News and Statistics about the World Wide Camellia Cultivars. Having a multilingual Register is an enormous benefit too. It is very helpful for my work on editing the ICJ.

— Frieda Delvaux,Editor of International Camellia

I have found the Database of the International Camellia Register(DICR, https://camellia.iflora.cn) to be indispensable in my work in the nursery trade. I use it as a reference for spelling of names, descriptions of flowers, history of cultivars and pictures of blooms. Having all this information in one location that is easy to access makes writing my online catalog possible. Having an authority on Camellia cultivars make communication with customers easier too. I sincerely hope that it continues to be available for use worldwide.

— David Parkes, Camellia Forest Nursery

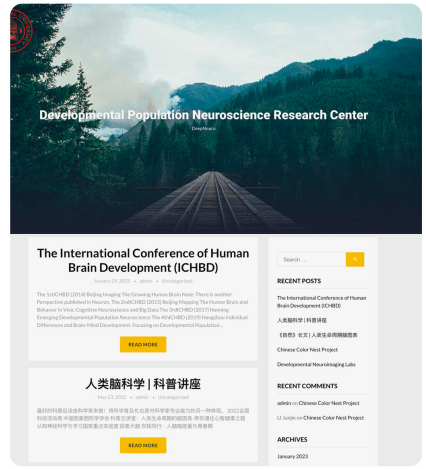


中国活体脑成像共享数据库

http://deepneuro.bnu.edu.cn

BRAIN projects will follow the two notable big brain data projects (HCP and CoRR) to make tools and data avaliable to the community.

— Thomas Insel,,Founder of Mind-strong Health,Former Director of The National Institute of Mental Health



核物理主题数据库

http://www.nuclear.csdb.cn

The new updated CENDL-3.2 is remarkable in those troubled times. I have had no processing errors in the making of such application library forms from CENDL-3.2, unique in my books.

—Jean-Christophe Sublet, unit Head (Data services unit), International Atomic Energy Agency



国家重点研发计划科学数据汇交

国家基础数据中心充分利用依托单位中国科学院计算机网络信息中心基础设施软硬件和数据工作团队的综合优势，积极开展“政府预算资金资助的各级科技计划（专项、基金等）项目所形成的科学数据”的汇交，形成一套面向科研项目全生命周期的完整科学数据汇交管理解决方案。全面支持各级科技计划项目数据汇交，促进科学数据的开放共享与应用服务。

汇交机制及流程



服务专业机构分布信息

项目管理机构	所属专项	项目数
中国21世纪议程管理中心（209个）	公共安全风险防控与应急技术装备	48
	水资源高效开发利用	29
	绿色建筑及建筑工业化	28
	科技冬奥	27
	大气污染成因与控制技术研究	18
	固废资源化	19
	深地资源勘查开采	19
	重大自然灾害监测预警与防范	16
	…	
科学技术部高技术研究发展中心（935个）	重点基础材料技术提升与产业化	86
	纳米科技	76
	智能机器人	75
	新能源汽车	69
	煤炭清洁高效利用和新型节能技术	57
	战略性先进电子材料	56
	增材制造与激光制造	56
	现代服务业共性关键技术研发及应用示范	54
	量子调控与量子信息	54
	重大科学仪器设备开发	51
	云计算和大数据	47
	可再生能源与氢能技术	42
	蛋白质机器与生命过程调控	41
	网络协同制造和智能工厂	41
	高性能计算	39
	综合交通运输与智能交通	19
	光电子与微电子器件及集成	17
	先进轨道交通	14
	大科学装置前沿研究	14
	科技创新2030–“新一代人工智能”重大项目	11
	交通基础设施	8
	…	
工业和信息化部产业发展促进中心（172个）	制造基础技术与关键部件	55
	智能电网技术与装备	58
	网络空间安全	29
	物联网与智慧城市关键技术及示范	24
	…	
中国生物技术发展中心（87个）	科技创新2030–“脑科学与类脑研究”重大项目	45
	合成生物学	11
	数字诊疗装备研发	9
	生物与信息融合（BT 与 IT 融合）	8
	诊疗装备与生物医用材料	6
	…	
中国科学技术交流中心（147个）	政府间国际科技创新合作重点专项	113
	战略性国际科技创新合作重点专项	34

总体成效

截止2022年12月，国家基础数据中心为5家国家重点研发计划项目管理专业机构，累计63个重点专项1550个项目提供科学数据汇交支撑服务，目前汇交科学数据总量超过1.46PB。

2022年度，国家基础数据中心通过线上/线下的方式组织和参加了涉及“川藏铁路”、“可再生能源与氢能技术”、“煤炭清洁高效利用和新型节能技术”等多个国家重点研发计划项目重点专项共10余次科技计划项目汇交工作培训会，并成功举办首次面向国家重点研发计划专项项目科学数据汇交工作的公开培训讲座，累计受众超万人。

项目汇交数据资源特征图·学科领域

新一代
信息技术
>>



新能源
>>



新材料
>>



人工智能
<<



量子科技
<<



公共安全
<<



项目汇交数据资源特征图·专项



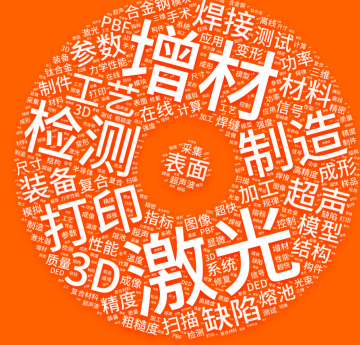


















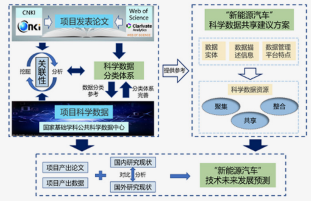
创新应用课题一：

“新能源汽车”数据分析与共享方案研究

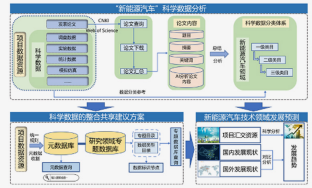
课题通过对“新能源汽车”国家重点研发计划重点专项汇交的科学数据集、关联论文成果等开展综合分析，建立“新能源汽车”重点专项的科学数据分类体系，提出“新能源汽车”重点专项科学数据整合共享建议方案，并对新能源汽车技术未来发展趋势进行预测。

负责人
耿莉敏　genglimin@chd.edu.cn

承担单位
长安大学



项目主要研究内容



项目技术路线图

“新能源汽车”重点专项科学数据分类体系

课题分析了54个“新能源汽车”国家重点研发项目的数据汇交计划文本，以及在线发布的1595个数据集元数据信息。聚类提取关键词，建立了由三级节点构成的“新能源汽车”重点专项的科学数据分类体系，共包含6个一级节点，42个二级节点，641个三级节点。其中将不同技术领域设置为一级节点。同时，进一步统计分析了该专项高频词，共计157个。

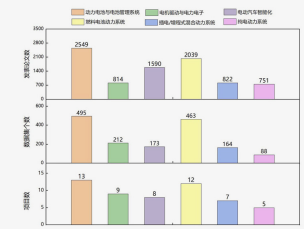
一级节点名称	项目数	数据集个数	二级节点数	三级节点数	高频词数
动力电池与电池管理系统	13	495	7	103	26
电机驱动与电力电子总成	9	212	6	92	21
电动汽车智能化技术	8	173	8	47	21
燃料电池与动力系统	12	463	12	260	46
插电/增程式混合动力系统	7	164	5	69	24
纯电动力系统	5	88	4	70	19

科学数据与论文关联性分析

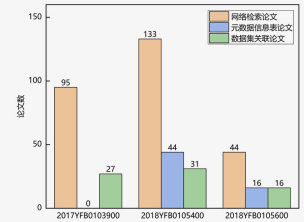
课题对项目汇交的科学数据集名称和公开的学术论文进行关联文本分析。在显性关联方面，54个项目共汇交1595个数据集、发表8565篇论文，其中动力电池与电池管理系统、燃料电池动力系统是重点研发项目投入较大、且产出数据集和论文成果较多的技术领域。隐形关联方面，采用相似度算法对编号为“2018YFB0105400”项目进行代表性量化分析。分析发现已发表的133篇论文中27篇综述类文章未关联数据集，其余106篇论文与数据集的平均综合相关度为50.96%，最高相关度为68.43%，最低相关度为18.58%。总体汇交的科学数据集在项目研究实施过程中发挥了重要作用。

科学数据整合共享建议方案

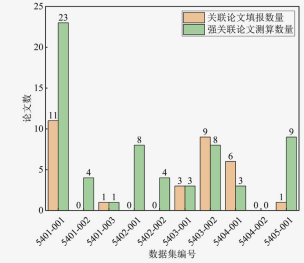
为推动“新能源汽车”重点专项科学数据更好地服务国家科技创新和社会经济发展，提出科学数据整合共享建议方案：一是从元数据及关联成果填报、数据格式标准等方面加强规范管理。二是“新能源汽车”专项应按照项目所属技术领域，对数据集与测试报告、论文、专利等关联成果进行整合共享，强化项目关联成果的整合分析，并建立多维度的科学数据检索发现条件，提升数据共享服务能力。三是建议加强科学数据的全生命周期管理，可在项目执行阶段同步提交阶段性成果数据，推动项目参建单位研究过程中的数据管理和自查，做到中间环节数据与结题汇交数据之间的有效衔接，确保数据的可用性和完整性。



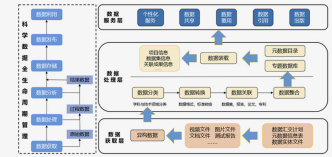
“新能源汽车”专项不同技术领域成果统计



三个典型项目的关联论文分析



数据集关联论文填报数据与测算数据对比



科学数据共享平台建议架构

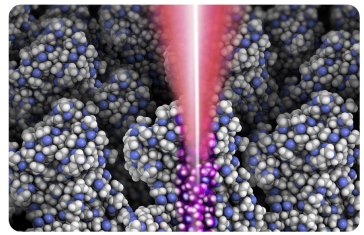
创新应用课题二：

“纳米科学”数据标准与共享应用策略研究

纳米科学与技术经过近年的迅猛发展，已成为和芯片、制药、精密仪器等应用结合紧密的重要学科。我国已成为纳米研究论文发表与专利申请数量全球第一的大国，纳米基础研究原始数据达到海量规模，数据分类和标准需求迫切。本课题对其分级分类和数据汇交标准制定展开研究，提出纳米数据全流程管理方案及循环利用发展路径，以推动纳米数据精准有效的汇交与管理。

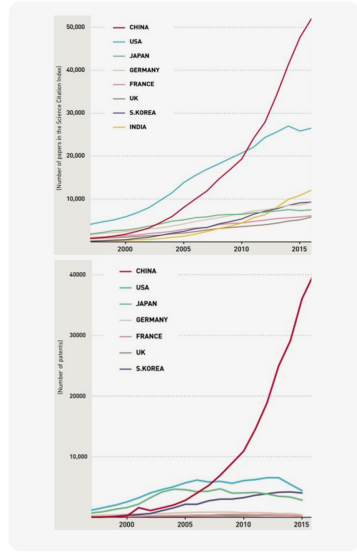
负责人
郭令举　guolj@nanoctr.cn

承担单位
国家纳米科学中心



纳米项目数据的分类体系

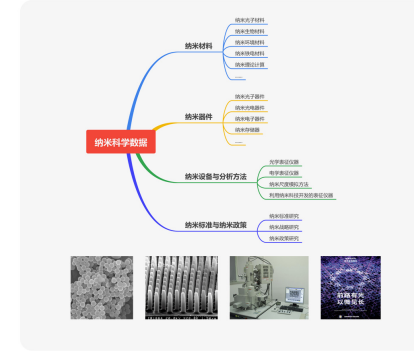
纳米数据通常可依据纳米材料的组分、功能和结构特点进行分类。依据81个项目信息，分析30个项目数据集信息，总结有四类特征：一是以纳米材料为核心，面向某一应用方向或者面向纳米科学的某一物理化学特性，注重微观机理，通过自顶而上方法制备。二是以纳米器件为核心，具有鲜明的应用出口特点，更加关注器件的宏观性能，通过自上而下制备。三是基于表征分析纳米材料的方法，或者运用纳米技术分析其他的方法与手段。四是产自纳米标准与政策的研究，不涉及具体材料。因此建立以纳米材料、纳米器件、纳米设备与分析方法，纳米标准与政策为核心的二级分类方法，有利于数据的汇交管理和重复使用。



1996–2016年间全球各主要国家纳米科学与技术研究论文产出情况与专利申请情况

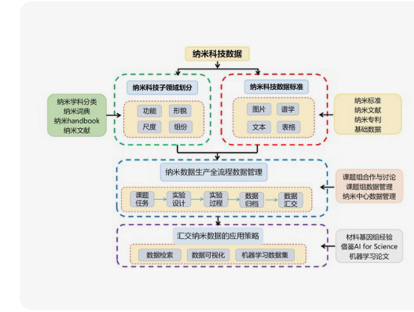
纳米项目数据汇交策略与标准

在纳米领域项目数据汇交策略方面，建议采用“项目–课题–目标–研究成果–佐证数据集”的分级分类方法。对照任务书中的考核指标，佐证研究任务，描述完整加工和质控过程，便于用户快速的浏览、检索数据，获取和重用原始的表征数据。数据标准方面，建议遵循材料和器件两类通用标准，其中材料类数据分为样品–表征数据–性能数据模块，器件类数据分为器件组成–器件构建方法–表征数据–性能数据模块。



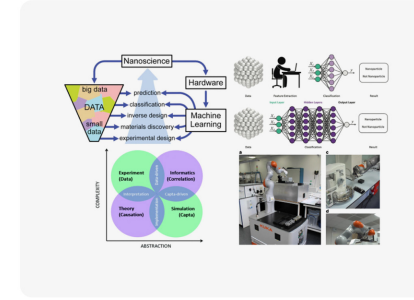
纳米科学数据的全流程管理方案

为解决当前相关科技项目数据制度不明确、层级管理复杂、参与单位较多和原始数据缺失等问题，课题根据纳米学科的特点，提出了纳米科学数据的全流程管理建议方案：一是项目立项后应制定数据汇交计划，确定数据范围并建立数据目录。二是开展数据汇交管理培训，提高研究人员的数据汇交和管理意识。三是提供数据汇交的数据标准或数据模板，可将支撑学术成果的数据分为核心数据和试错数据等。四是开发面向领域用户的科学数据全流程管理的支撑软件，实现数据实时上传和利用。



纳米数据循环利用发展路径

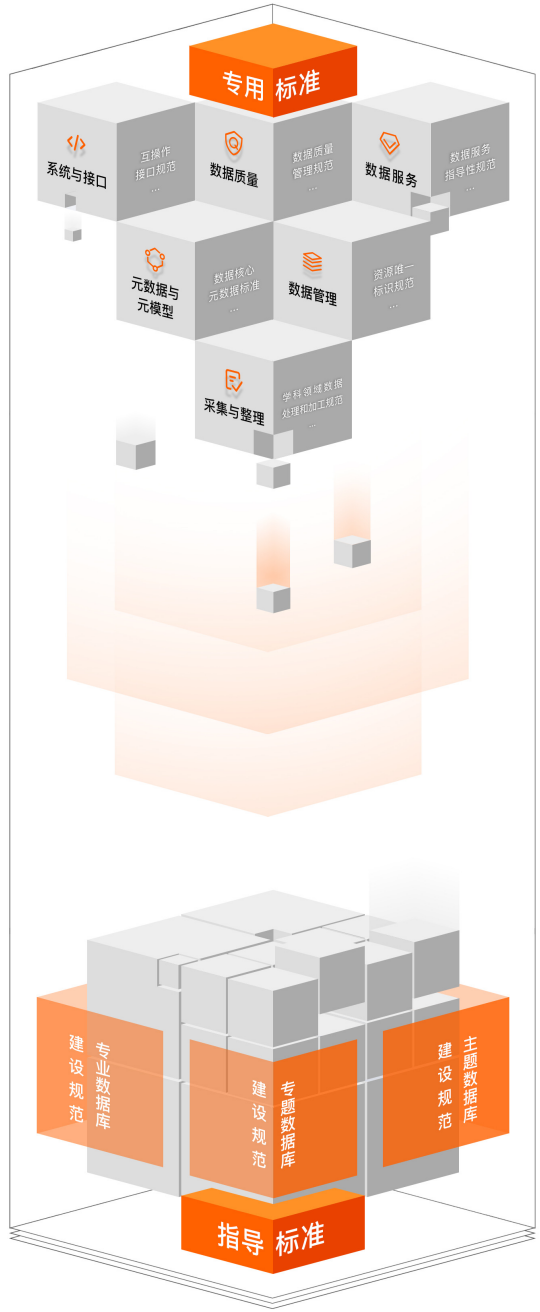
课题研究分析纳米领域学科特点和已汇交项目的数据内容，提出了纳米数据循环利用的发展路径建议：一是完善纳米数据分类分级和数据标准制定工作。二是融合计算多种来源数据，特别是结合实验数据和理论数据，成为理论计算数据与实验数据交流的枢纽。三是面向人工智能技术应用提供基础标准化的纳米科学数据资源，推动AI for Science的快速发展。



政策和标准体系

为落实《科学数据管理办法》、《国家科技资源共享服务平台管理办法》、《中国科学院科学数据管理与开放共享办法（试行）》等政策文件要求，国家基础数据中心积极推动政策制度和标准体系的研究与制定。

在政策制度方面，2022年继续完善国家基础数据中心政策体系，同时结合国家重点研发计划项目科学数据汇交工作实践，形成《国家基础学科公共科学数据中心科技计划项目科学数据汇交工作指南V2.0》等文档。在标准规范方面，利用国家重点研发计划“科学数据安全技术及基础技术标准研究”项目成果，进一步完善中心平台和自建数据库安全相关数据标准。年度牵头和参与研制6项国家标准。



主导和参研标准清单

序号	标准名称	标准号	中心排名
1	生态科学数据元数据	GB/T 20533–2006	第1单位
2	信息技术 科学数据引用	GB/T 35294–2017	第1单位
3	信息技术 数据溯源描述模型	GB/T 34945–2017	第1单位
4	科学数据 安全传输技术要求	T/CIIA 020–2022	第1单位
5	科学数据 安全防护技术要求	T/CIIA 021–2022	第1单位
6	科学数据 云存储环境运维流程与服务要求	T/CIIA 022–2022	第1单位
7	科学数据 云平台运维流程与要求	T/CIIA 023–2022	第1单位
8	科学数据托管存储要求	T/STSI 41–2023	第1单位
9	虾青素旋光异构体含量的测定——液相色谱法	GB/T 38478–2021	第1单位
10	中国动物分类代码 第1部分：脊椎动物	GB/T 15628.1–2021	第1单位
11	土壤科学数据元数据	GB/T 32739–2016	第2单位
12	土壤质量 土壤相关数据的数字交换	GB/T 41224–2021	第2单位
13	科学数据 安全标准体系	T/CIIA 017–2022	第2单位
14	科学数据 数据安全分级程序	T/CIIA 024–2022	第2单位
15	科学数据 数据安全分类质量评价指标	T/CIIA 025–2022	第2单位
16	信息技术 数据交易服务平台 通用功能要求	GB/T 37728–2019	第3单位
17	科技平台 资源核心元数据	GB/T 30523–2014	第4单位
18	科技平台 一致性测试的原则与方法	GB/T 31071–2014	第4单位
19	科技平台 服务核心元数据	GB/T 31073–2014	第4单位
20	信息技术 数据质量评价指标	GB/T 36344–2018	第4单位
21	科技平台 用户元数据	GB/T 39913–2021	第4单位
22	科学数据 安全能力成熟度模型	T/CIIA 019–2022	第4单位
23	生态系统长期观测数据产品分类分级	T/CIIA 027–2022	第4单位
24	生态系统长期观测数据产品规范	T/CIIA 028–2022	第4单位
25	海洋科学数据共享安全管理要求	T/CIIA 029–2022	第4单位
26	微生物数据库安全体系设计要求	T/CIIA 030–2022	第4单位
27	科学数据责、权、利参考类型	T/MSITISA 02–010–2022	第4单位
28	科学数据责、权、利界定指南	T/MSITISA 02–011–2022	第4单位
29	科学数据权益保护技术实施指南	T/MSITISA 02–012–2022	第4单位
30	科学数据责任追究技术实施指南	T/MSITISA 02–013–2022	第4单位
31	信息技术 数据交易服务平台 交易数据描述	GB/T 36343–2018	第5单位
32	科技平台 元数据注册与管理	GB/T 30524–2014	第5单位
33	科学数据 安全管理指南	T/CIIA 018–2022	第5单位
34	中文新闻信息内容 第1部分：概念模型	GB/T 20092.1–2022	第6单位
35	中文新闻信息内容 第2部分：新闻元数据	GB/T 20092.2–2022	第6单位
36	中文新闻信息内容 第3部分：数据交换的XML格式	GB/T 20092.3–2022	第6单位
37	科技计划形成的科学数据汇交 通用代码集	GB/T 39908–2021	第9单位
38	科技计划形成的科学数据汇交通用数据元	GB/T 39909–2021	第9单位
39	科技计划形成的科学数据汇交 技术与管理规范	GB/T 39912–2021	第9单位
40	数据质量 第61部分：数据质量管理：过程参考模型	GB/T 42381.61–2023	第10单位
41	信息技术 大数据 术语	GB/T 35295–2017	第16单位
42	数据论文出版元数据	计划号20205082–T–306	研制中
43	科学数据溯源元数据	计划号20221229–T–306	研制中
44	科学数据安全审计要求	计划号20221227–T–306	研制中
45	科学数据安全要求通则	计划号20221067–T–306	研制中
46	科学数据安全分类分级指南	计划号20221063–T–306	研制中
47	科学数据权益保护基本要求	计划号20221065–T–306	研制中

特色服务与数据软件

国家基础数据中心面向科学数据中心体系化、系统化建设管理的需求研发系列工具，面向不同类型数据用户和应用场景提供分布式数据资源自助管理与服务、数据资源协同管理与外延服务。同时提供特色数据服务，强化数据产品加工和深度分析挖掘、服务国家重大项目和产业升级、充分发挥基础科学数据在我国双创战略实施中的基础支撑作用，提升我国基础科学数据资源的权威汇聚、长期保存、开放共享、分析挖掘、应用服务和驱动创新的整体能力和国际影响力。

数据服务

国家基础数据中心通过提供数据集下载、数据众包、数据汇交、数据出版和数据标准等服务，推动跨学科、跨地域的数据融合和合作研究。截至2022年底，国家基础数据中心可共享数据量2.71PB，同比2021年数据量增长18%，累计注册用户15.5万人，在线访问人次19867.04万次，下载量2163.56TB；其中2022年在线访问人次3356.34万次，下载量521.45TB，全力保障社会各界对中心数据资源的多元化使用需求。





*服务统计数据自基础科学数据共享网至国家基础学科数据云

专题服务

国家基础数据中心充分发挥多学科的优势，全面支持国家重大发展战略与国家区域发展，特别面向京津冀、长江经济带、港粤澳大湾区、一带一路等重点区域提供多元化的数据服务、专题服务、信息服务等。为科研机构、高等院校、企事业单位与相关政府部门提供线上、线下相结合的离线服务、跟踪服务和主动服务，取得良好的经济成效。

自主关键技术服务　　大湾区服务

一带一路服务　　交叉学科服务

重大战略服务　　基础地理服务

科学数据协同管理平台

https://ds.nbsdc.cn

面向国家重点研发计划项目数据汇交用户分配专属项目空间，可实现多人跨地域协同管理、持续汇聚数据，大幅缩减项目数据汇交中的上传时间，有效提升了项目数据汇交管理效率。

- 注册用户2345人
- 访问量24万人次
- 项目专属空间603个



数据存储发布服务

https://idb.nbsdc.cn

提供国家重点研发计划项目数据汇交审核支持，允许用户按照自定的权限与许可公开科学数据，免费分配CSTR永久唯一标识，为用户项目数据汇交、发布数据、数据确权等提供支持。

- 支持审核元数据及实体数据29225条
- 支持审核项目603个
- 已分配CSTR资源数10227个



科学数据汇交凭证查验服务

https://www.scichain.cn/nbsdc

提供国家重点研发计划项目数据汇交凭证查验服务，利用区块链技术实现数据汇交凭证唯一可信、防篡改，确保国家基础数据中心出具的国家级科技项目汇交凭证完整性、真实性和可追溯性。

- 上链数据资源10227个
- 存证项目1550个

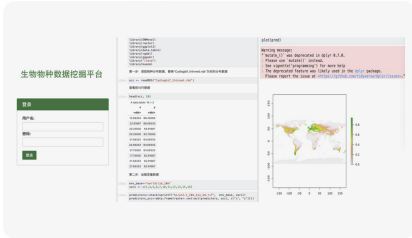


领域特色工具

生物物种数据挖掘平台

完成方：中国动物主题数据库

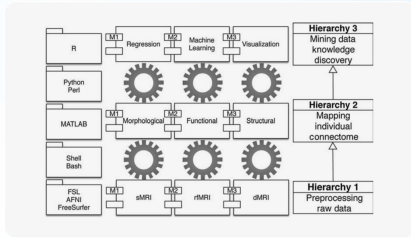
依托中国动物主题数据库搭建的生物物种数据挖掘平台，可用于构建物种分布模型。支持用户对接动物主题库中的数据，利用 R 语言或者 Python 进行在线编程，实现在线运算和展示，该平台已用于入侵昆虫潜在分布的研究。



人脑影像数据处理软件

完成方：中国活体人脑成像共享数据库

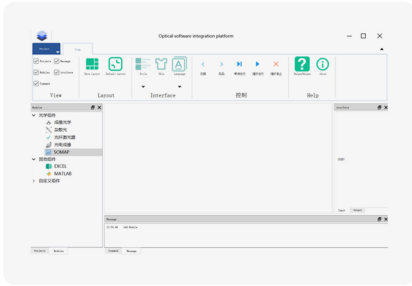
人脑影像数据分析挖掘工具CCS（Connectome Computation System）面向用户终端，以命令行程序设定参数配置文件，保留运行过程日志，为检查、调试和备份数据处理过程做详细记录，实现结构图像(sMRI)预、静息态功能磁共振图像(RFMRI)、扩散张量图像(dMRI)的预处理、后处理、知识挖掘及个体功能图谱计算等功能。



人工智能通用光学系统仿真优化软件

完成方：光学技术数据库

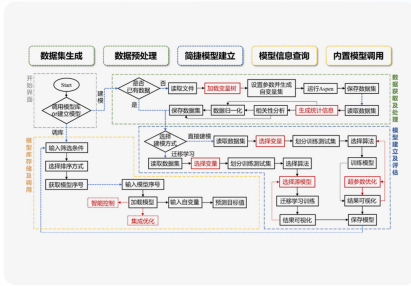
基于智能算法自动调用计算资源，将光学技术数据库数据导入光学软件总体集成平台，对公用基础组件、光机热耦合分析模块、光机拟合模块、云端数据库及软件远程升级维护中心提供数据支撑，根据用户设定的设计目标值寻找最优方案，整个优化过程不需要任何人工参与。



数据驱动工业生产单元建模工具

完成方：化学主题数据库

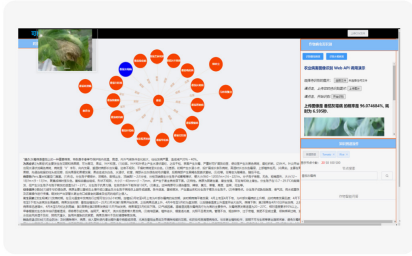
数据驱动工业生产单元建模工具Modelfactory是面向化学专业研究人员的无/少代码领域专业数据加工、分析及建模工具。内置多种经典机器学习算法用于模型开发，同时支持用户自定义算法；支持迁移学习，方便用户在现有模型基础上进行参数调整，提升模型开发效率和质量；具备数据集及模型管理功能，便于数据管理和共享。



跨媒体知识图谱交互平台

完成方：图像理解机器学习数据库

构建基于图像理解机器学习数据集的跨媒体知识图谱交互平台，包括传统文本知识图谱的视觉模态增强、用于后台数据管理的可视化交互、图文联合检索与问答系统示范应用等内容。该平台基于跨媒体知识图谱的病害图文联合检索，实现作物病害图像识别与跨媒体知识图谱的实时关联。



更多软件工具列表

序号	软件名称	完成方
1	生物分类树工具	中国动物主题数据库
2	QuickView “VLBI快视”软件	VLBI月球与深空探测数据库
3	图像处理光滑点扩散函数技术	载人航天空间科学与应用数据库
4	DEM在线切割模型	地理空间数据库
5	植物DNA条形码管理系统	中国植物DNA条形码数据库
6	DREAM人脑功能振荡信号多频段分解软件	中国活体人脑成像共享数据库

服务体系

总体成效

国家基础数据中心基于G-FAIR原则，构建安全可管理的一体化数据基础设施，形成了完善的服务体系。通过门户网站“国家基础学科数据云”（www.nbsdc.cn）提供包括数据采集、数据发现、数据出版、技术服务、数据分析挖掘、图谱关联等数据全生命周期管理服务。截止2022年底，实现对科学数据集、论文、专著、报告、软著、软件、标准等多种类型数据资源的在线管理和关联发布，有力支持了基础领域多学科数据的汇聚、融合和共享应用。

G-FAIR

可管理 Governable
数据汇交
数据安全
分级分类

可发现 Findability
唯一标识
数据检索
关联推荐

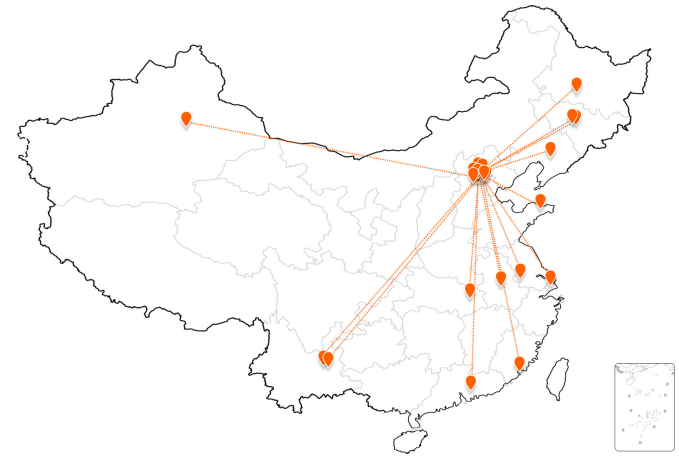
可访问 Accessibility
数据下载
数据服务
访问控制

可操作 Interoperability
数据服务
数据接口

可重用 Reuse
数据出版
数据引用
版权声明

国家基础数据中心一体化基础设施

2022年度，国家基础数据中心依托中国科技云建成数据专有环境，超5PB存储空间



高可靠、高稳定的基础设施运行环境

7×24小时实时监控
进口大容量UPS系统
国际最高标准机房
专业化的科学数据存储基础设施

高性能、高可靠、弹性扩展云存储环境

在线盘阵、近线盘阵和近线磁带库三级存储框架
服务器处理节点达到128个
正常运行率达99.9%

高速稳定的网络设施

基于IPv4、IPv6协议数据存储服务
中国科技网
网络总带宽达到10GB/S

合作推广

“数字经济时代的科学数据伦理”专题研讨会

2022年1月，国家基础数据中心联合国际科学理事会数据委员会（CODATA）中国委员会组织召开“数字经济时代的科学数据伦理”专题研讨会，来自人口健康、农业食品、地球观测、非物质文化遗产保护、大数据等领域伦理研究专家、学者等三十余人参加了此次专题研讨会。专家们深度探讨我国各领域伦理进展、成就及其与科学数据伦理的关系，一致认为数字经济对数据伦理的要求越来越高，平衡数据开发利用、开放共享、自由流动和数据安全之间的关系，降低共享利用可能引发的伦理风险，是推动数据要素市场化配置的必答命题和重大挑战。



“开放科学与科学数据”专题研讨会

2022年3月，为深度探讨国内外开放科学发展总体情况，推动科学数据开放共享实践，国家基础数据中心联合国际科学理事会数据委员会（CODATA）中国全国委员会组织召开“开放科学与科学数据”专题研讨会，来自北京市科学技术研究院、中国科学院、高校的十余位专家和学者参加此次研讨会。专家们认为开放科学的发展将会对全球学术界产生变革性影响，在国际相关政策驱动下，如何平衡数据伦理、数据安全、数据标准之间的关系是推动数据开放共享的重要课题，建议增强开放科学教育和技能培训、增加高校等教育机构的参与度，保护数据产权、开展示范工程，结合国家发展实际需求，推动开放科学在中国落地。



“我国科学数据开放共享的问题与挑战”专题研讨会

2022年6月，国家基础数据中心组织召开“我国科学数据开放共享的问题与挑战”专题研讨会，来自中国科学院空天信息创新研究院、中国科协创新战略研究院、中国科学院科技战略咨询研究院、中国科学院地理科学与资源研究所、中国科学技术信息研究所、中国科学院文献情报中心等十余位研究专家、学者参加了本次研讨会。与会专家围绕我国科学数据开放共享问题与挑战展开深入的讨论，从文化建设、科学数据全生命周期管理、科学数据管理机构发展模式、数据治理机制等方面提出了发展建议，为推动我国科学数据的发展和未来工作实践提供参考。



“可计量的科学数据影响力智能评价方法研究”专题研讨会

2022年8月，国家基础数据中心组织召开“可计量的科学数据影响力智能评价方法研究”专题研讨会，来自国家科技基础条件平台中心、北京师范大学、中国科学院高能物理研究所、中国科学院地理科学与资源研究所、中国科学院国家天文台、中国科学院过程工程研究所、中国科学院空天研究院、中国科学院动物研究所的领域专家参加本次研讨会。与会专家从研究对象、计算方法、评价维度、应用场景和领域实证等方面提出了具体建议，指出相关研究工作对促进我国科学数据共享生态建设具有重要意义。



组织参加“共享杯”科技资源共享服务创新大赛

2022年5月，国家基础数据中心积极组织和参加国家科技基础条件平台中心主办的第十届“共享杯”科技资源共享服务创新大赛，提供基础学科领域数据应用创新赛题及相关科学数据集，荣获“优秀组织奖”。



宣传推广及媒体策略

以国家基础数据中心门户和官方公众号为核心构建宣传矩阵，促进前沿学术交流，开展科普教育，以数字化赋能推动科学数据开放共享。

年度推送科学数据
支撑成果、动态500篇次+

追踪国际前沿，推动数据伦理、
数据互操作、数据应用创新
等理论研究

20+媒体联动宣传
服务国内科研用户80万人+



科学数据库概况

在2002年科技部科技基础性工作专项“CODATA中国理化数据库”和2003年国家科技基础条件重点项目“基础科学数据中心”的基础上，经过近20余年的持续建设和发展，国家基础数据中心聚焦物理、化学、材料、动物、植物、信息科学、交通运输、重点区域、前沿学科等学科方向，以我国科学家在科学研究活动中获取的观测数据、探测数据、试验数据、实验数据、调查数据、考察数据、统计数据、模拟数据等为核心，建设自主数据库，目前共建有23个主题数据库，数据总量1.25PB，在支持科研创新、多学科交叉融合、国家战略、产业经济发展、全民健康等方面发挥了积极作用。





交通运输分中心

信息学科分中心

世界山茶属植物品种数据库

图像理解机器学习数据库

环境健康主题数据库

中国海岸带植物资源数据库

医学物理数据库

中国动物主题数据库

中国植物主题数据库

中国植物DNA条形码数据库

化学主题数据库

原子分子数据库

光学技术数据库

核物理主题数据库

宇宙学数值模拟数据库

VLBI月球与深空探测数据库

中国病毒资源基础数据库

载人航天空间科学与应用数据库

材料科学基础数据库

干细胞与代谢疾病数据库

中国活体人脑成像共享数据库

地理空间数据库

干旱区药用资源数据库

交通运输分中心

交通运输分中心依托交通运输部公路科学研究院、交通运输部天津水运工程科学研究院等单位联合建设，主要建设公路科学数据中心和水运工程科学研究与应用数据库。

分中心汇聚交通行业的海量数据，覆盖了足尺环道、道路桥隧、路网监测、智能交通、客运货运等各个领域，形成了沿海主要港口设施维护管理数据、近岸水环境数据、环保监测数据、地下水监测数据等特色数据集，建设了一带一路、京津冀一体化等八大专题资源库。公路科学研究院建设的足尺环道，是目前除美国以外唯一具备实车加速加载能力的足尺路面试验研究设施，基于该设施获取了7000万次实车加载作用下的沥青路面全寿命、全时空、全环境下服役性能观测数据，数据具有全面性和权威性，是国家野外观测站主要数据来源。相关数据在道路服役性能研究、出行服务与疫情防控、货运监测与冷链运输、车路协同与风险防控等实际应用中发挥了重要作用，支撑服务多项国家重点研发计划项目，为国家重大战略实施和重大工程建设提供了有效的数据支持。

联系人

郑晓妍　xy.zheng@rioh.cn

尚东方　shangdf@tiwte.ac.cn

参建单位

交通运输部公路科学研究院

交通运输部天津水运工程科学研究院

2022年新增数据资源清单（节选）

序号	数据库/集名称	年度增量
1	2022年中心站沥青路面结构响应原始数据	188.9TB
2	2022年内蒙古海拉尔观测点结构响应数据	438GB
3	2022年中心站落锤式弯沉仪响应数据	145GB
4	2022年广东观测点结构响应数据	90GB
5	2022年河北观测点结构响应数据	47.1GB
6	水运工程科学研究与应用数据库	26.1GB
7	2022年中心站结构温湿度数据	219.2MB

支撑国家重点研发计划项目
19个

支撑论文
15篇

团体标准
3项

专利
14项

专著
5部

软件著作权
6项

荣获科学技术奖项
14个

典型案例



“足尺环道数据集”为国家野外观测站研究工作提供重要数据支持

交通运输分中心足尺环道数据集为“国家队”北京大杜社公路材料腐蚀与工程安全野外科学观测研究站的研究工作提供重要数据支持，基于大数据的道路服役性能衰变模型，提供了基础数据服务，针对公路工程材料腐蚀和工程安全的共性规律问题，开展野外观测和试验观测研究，通过数据、模型揭示公路工程材料腐蚀机理，掌握公路工程全寿命周期性能演化规律，创新设计体系和模型，为智能道路路面养护的最佳时机确定方法，对我国公路路网建养决策提供了数据支持。

“公路低碳环保绿色养护数据集”支撑了苏北地区高速公路养护绿色廊道示范工程建设

公路网建设是影响路域生态的重要因素，交通运输分中心科研团队围绕生态化公路网建设的关键难点，构建了“协同降噪——水质净化——低碳减排”多功能公路低影响开发组合技术公路低碳环保绿色养护数据集，并结合苏北路网扩建、改造等工程，开展了组合技术的集成应用。修筑的综合降噪试验段，综合降噪效果达到12dB~15dB，极大减弱了交通噪声对物种种群的影响；利用多孔路面雨水径流回收及滞纳功能，提升了路域水体质量30%以上；实施多孔路面就地热再生和双层同步就地热再生工程，路面材料固体废弃物回收率100%，避免了废弃物大量侵占土地，提高降碳减排效益，助力苏北地区高速公路养护绿色廊道示范工程建设。





“水运工程科学研究与应用数据库”助力渤海湾港口航道长期性能野外科学观测站建设

渤海湾港口航道长期性能野外科学观测站，对保障渤海湾内重大港口航道基础设施的安全和畅通、增强渤海湾交通基础设施的防灾减灾能力具有重要的战略意义。基于水运工程科学数据，形成了淤泥和粉砂质海岸泥沙运动机理及航道淤积规律研究、波浪—结构—地基耦合效应及结构破坏机理研究、港口水工建筑物冻融破损机理与长期性能研究等系列成果，建立了两级数据中心与模拟分析系统嵌套耦合的“渤海湾港口航道长期性能大数据服务平台”，以高精度实时模拟预报成果为切入点，形成了海上快速搜救辅助应用、黄骅港航道淤积智能预报等一系列智能化场景应用。

“水运工程科学研究与应用数据库”助力乌江水情监测预测系统建设

乌江下游受流域暴降大雨和电站下泄非恒定流影响，带来山区河流水位骤涨、流速激增、流态复杂等安全隐患。为解决这一问题，基于水运工程科学研究与应用数据库，通过先进的物联网技术手段实时监测乌江航道水文信息，研发了基于深度学习的水文预测分析模型，对航道水位、流速、流量等关键要素进行计算、分析和预测，实现下游水位的中短期预测。系统成功应用于多次洪峰过境，建设成果为乌江航运安全提供了全天候、实时性、动态化的水上安全监测与科学预警手段。



信息学科分中心

信息学科分中心依托中国科学院计算机网络信息中心建设，主要面向图像识别、自然语言处理、机器视觉等信息科学研究领域，以及信息学科与其他学科的交叉研究领域，汇聚海量信息数据，建设信息学科相关科学研究与领域应用数据库，形成图像分割、文本分类、文本生成、视频捕捉、领域分析等特色数据集。同时提供了多模态信息抽取工具，通过领域标注和实体抽取，尝试构建了高质量的AI领域数据集和光电催化领域数据集，支持了铜基CO2还原催化剂知识推荐研究，推荐的材料法拉第效率预测精度与效果，均高于领域已有推荐方法，成果支撑多项国家重点研发计划项目，为领域应用研究和国家重点战略实施提供了数据支持。

支撑国家重点研发计划

项目编号	项目名称
2022YFF0712200	基于领域知识图谱的光电催化材料挖掘软件
2022YFF0711900	科技文献内容深度挖掘及智能分析关键技术和软件
2021YFE0108400	基于大数据原位可视化的高精度大气污染仿真

专著
1部

支撑论文
15篇

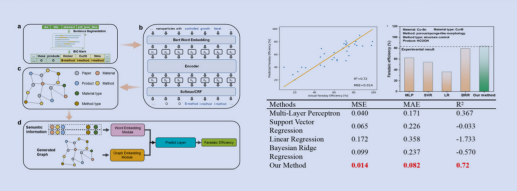
软件著作权
6项

专利
5项

支撑国家重点研发计划项目
3个

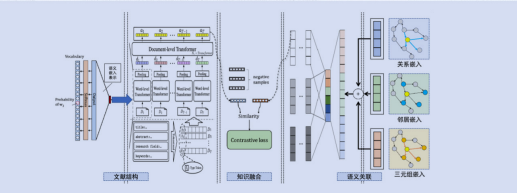
联系人
杜一　duyi@cnic.cn

承建单位
中国科学院计算机网络信息中心



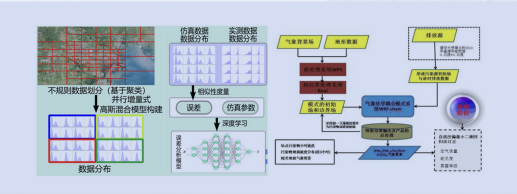
支撑光电催化材料推荐研究，大幅度提高材料法拉第效率预测精度

面向铜基CO2还原催化剂推荐，对小样本下的知识抽取方法进行了初步探索，知识抽取效果超出Bert、Sci-Bert等通用领域知识抽取方法，推荐出的材料法拉第效率预测精度与效果，均高于领域已有推荐方法。成果发表于科学数据领域期刊《Scientific Data》,为国家重点研发计划项目“基于领域知识图谱的光电催化材料挖掘软件2022YFF0712200”研究工作提供数据支持。



支撑面向科技文献内容深度挖掘及智能分析，助力科研人员知识发现与创新

针对文献情报机构迫切需要提升智能技术方法手段，将掩藏在海量科技文献中的深度知识内容有效挖掘并加以充分利用，突破智能化关键技术，研发知识服务与情报分析自主软件。支撑了国家重点研发计划项目“科技文献内容深度挖掘及智能分析关键技术和软件2022YFF0711900”的研究工作。



支撑高精度大气污染仿真中的大数据特征分析，解决实测数据与仿真数据的误差分析难题

针对大气仿真数据分析科学问题，重点分析污染源以及研究海量实测数据与仿真数据间的误差来源。提出大数据分析驱动的污染源分析方法，突破大气污染大数据特征提取瓶颈，构建深度学习驱动的特征挖掘与发现方法，实现大气污染高精度仿真，辅助领域专家进行污染源交互式分析。支撑了国家重点研发计划项目“基于大数据原位可视化的高精度大气污染仿真2021YFE0108400”的研究工作。

世界山茶属植物品种数据库

世界山茶属植物品种数据库涵盖了山茶属中所有国家的茶花、茶叶、茶油品种，发表和使用的品种名称超过5万条，其中接受名2.6万条，图片数6.1万张。数据库以中国科学院昆明植物所成员为核心，组建国际团队，建立了全世界主要国家和地区协同更新数据的工作机制，实现数据持续更新。匿名用户可以浏览、搜索、下载图片等；注册用户可以上传照片，建立云相册等；世界多地的权威专家能够在线实时更新数据内容、完成鉴定和认证工作，保证了数据库可持续性、权威性、完整性及可用性。目前，世界山茶属植物品种数据库已成为世界范围内信息全面，领域内权威的数据库之一，为提升世界茶花、茶油、茶叶产业的健康发展提供积极的数据支持。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	“2022年新登录茶花品种”数据集	26,088个	101个
2	新增山茶花照片	61,818张	5569张

收录品种
26,088个

收录品种名称
51,329个

山茶属原种（含变种、变型）
333个

茶花影像资料
61,818张

总访问量
2,000万人次+

支撑论文
7篇

联系人
王仲朗　wang@mail.kib.ac.cn

承建单位
中国科学院昆明植物研究所

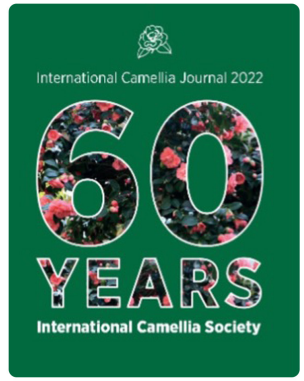


获奖荣誉

荣获第十三届中国茶花博览会金奖

2022年3月，“世界山茶品种数据库的开发与应用”荣获第十三届茶花博览会（科技创新类）金奖。

典型案例



山茶品种规范学名获«国际山茶会志»现任主编高度认可

«国际山茶会志»（International Camellia Journal）是国际山茶协会的官方正式会刊，至今已建刊60周年。其宗旨是：促进全世界对山茶的热爱、保持和提升茶花的知名度和影响力；开展与山茶有关的历史、科学和园艺研究；开展与各国和地区山茶协会和其它园艺协会的合作；传播与山茶品种相关的各类信息以及鼓励全球山茶爱好者进行友好交流。现任主编Frieda Delvaux表示：“我经常使用世界山茶属植物品种数据库来搜索品种历史，或确认特定山茶品种的原产国。比如可以很方便地查询所有比利时原产的山茶品种，非常有用。此外，有关常用异名的附加信息也非常方便，可以找到其正确的学名。这对我编辑《国际山茶会志》的工作很有帮助。”



助力美国著名山茶专业苗圃网络销售决策和品种培育

美国北卡罗莱纳大学教授Clifford Parks先生创立著名的“山茶林苗圃”（https://camforest.com），经营山茶苗木40多年，培育了大量特色山茶品种，同时保持稀有、古老和传统品种，为全世界山茶专业人士和山茶爱好者提供服务。2022年2月苗圃负责人David Parks来信致谢，他表示经常使用世界山茶属植物品种数据库，查询正确的山茶花品种名称、性状描述、品种的历史和开花的照片等信息。数据库中丰富全面的山茶花信息，为他编写在线销售目录提供了极大便利，基于权威数据库也更易于与客户沟通，他认为，世界山茶属植物品种数据库已成为苗圃不可或缺的“好帮手”。

国家基础学科公共科学数据中心
National Basic Science Data Center

图像理解机器学习数据库

图像理解机器学习数据库由中国科学院合肥物质科学研究院承担建设，包含图像、文本、视频等语义相关联的多模态数据，旨在推动人工智能机器学习理论方法及图像理解、视觉问答、知识表示与计算推理等关键技术研究。目前，该数据库已建成农业病虫害研究图库（IDADP）和农业图像理解数据集（ICDA）两个主题库，主要包括近20种大田作物和设施蔬菜，记录数达到7万余条，总数据量超过1.2TB。该数据库为机器学习建模相关的理论方法、关键技术与评价标准等研究提供基准数据，促进各个行业的智能化应用落地。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	黄瓜图像理解数据集	80GB	80GB
2	玉米图像理解数据集	120GB	120GB

联系人

袁媛　yuanyuan@iim.ac.cn

陈雷　chenlei@iim.ac.cn

承建单位

中国科学院合肥物质科学研究院

主题库
2个

资源总量
1.2TB

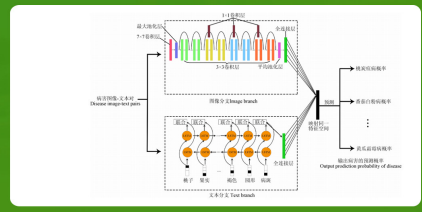
数据集服务科研院所
16家

支撑论文
5篇

软件著作权
3项

支撑国家自然科学基金面上项目
5个

典型案例



IDADP数据为农业图像理解研究提供支撑

基于深度卷积神经网络建立的视觉识别模型，通常只利用了图像模态的数据，导致模型的识别准确率和鲁棒性都依赖训练数据集的规模和标注的质量。构建开放环境下大规模的病害数据集并完成高质量的标注，需要付出巨大的经济和技术代价。基于图像理解机器学习数据库，提出了一种基于图像与文本双模态联合表征学习的开放环境下作物病害识别模型（bimodalNet）。该模型在病害图像模态的基础上，嵌入病害文本模态信息，利用两种模态病害信息间的相关性和互补性，实现了病害特征的联合表征学习。最终bimodalNet在较小的数据集上，取得了优于单纯的图像模态模型和文本模态模型的效果。目前，该成果服务于国家重点研发计划、北京市科技计划项目、河北省重点研发计划等项目。



葡萄病害图像数据集已累计服务16家科研机构

为促进机器学习中图像识别、图像理解等领域的理论和方法研究，满足具有高附加值经济作物的病虫害防治需求，面向非商业机构公开了IDADP-葡萄病害图像数据集，包含7种葡萄病害，共3622张图像，可通过在线下载并签署“IDADP-putao20220328数据集使用承诺”的方式获取数据集。截至目前，已累计服务浙江大学、东北农业大学、塔里木大学、香港物流及供应链管理应用技术研发中心等16家单位。

环境健康主题数据库

环境健康主题数据库由中国科学院城市环境研究所承建，面向“美丽中国”和“绿色丝路”国家建设需求，纳入生物体负荷数据、生物体响应胁迫数据、环境胁迫数据等，构建环境健康大数据平台。其中，生物体污染物数据库整合我国和“一带一路”沿线国家的生物体污染物数据，覆盖污染物种类超过600种，物种数目超过1400种。生物体污染物负荷水平数据记录数超过15万条。抗性基因数据库集成了环境样本类型，地理信息，抗性基因种类和丰度等信息，积累超过40万条环境ARGs分布数据。为环境保护问题提供数据支撑，为目标区域的污染防治和人群健康提供科学依据。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	生物体污染物数据库	15万条	1万条
2	抗性基因数据库	40万条	14万条

联系人

黄乾生　qshuang@iue.ac.cn

承建单位

中国科学院城市环境研究所

生物体污染物负荷记录

15万条+

环境抗生素抗性基因丰度记录

40万条+

支撑科研项目

4个

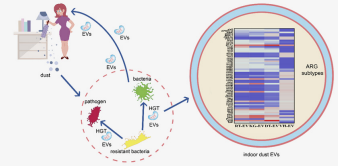
支撑论文

5篇

软件著作权

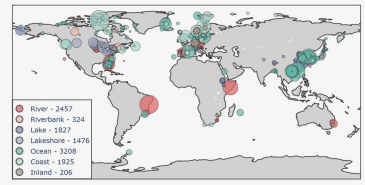
1项

典型案例



抗性基因数据库为抗性基因存在形式、分布和风险评估提供数据支撑

抗生素抗性基因（ARGs）加剧细菌耐药问题，长期严重威胁人群健康。抗性基因数据库以环境中ARGs的时空分布为主题，集成了环境样本类型、地理位置信息、环境因子信息、抗生素含量、ARGs种类、绝对丰度和相对丰度等，形成了具有优势的中国环境抗性基因分布数据库。该数据库提供全国抗性基因种类、组成和丰度信息、揭示抗性基因在水、土、气等环境介质中的新存在形式，支撑病原菌中抗性基因类型和丰度的识别，通过提供抗性基因种类和分布清单，为抗性基因数据比对、筛选提供数据支持。



生物体污染物数据库为多家医院提供污染数据

环境污染与人群健康息息相关。生物体污染物数据库整合、汇聚各类污染物的环境分布、浓度等系统数据，为重庆医科大学附属第一医院、厦门大学附属第一医院、山西医科大学公共卫生学院等多家医疗机构提供空气污染物、生物体污染物负荷、人群健康等数据。通过与健康人群数据进行耦合分析，支持解析污染对疾病发生的贡献率，从宏观尺度揭示环境污染导致相关疾病的成因。发表SCI论文2篇，获福建省科技奖项1个。

中国海岸带植物资源数据库

中国海岸带植物资源数据库是一个以海岸带蓝碳为导向，主张海岸带植物资源保护与利用一体化，兼顾生态环境保护的公众科普为特色的数据库。该数据库已涵盖海岸带的食用植物资源、药用植物资源、工业用植物资源、保护和改善海岸带生态的植物资源、海岸带植物资源的开发与利用，以及海岸带植物资源保护与管理等方面。至2022年底数据量已超过1.1TB，并形成了“线上数据+线下服务”的服务模式。从海藻资源数据支持到数据服务系统引导，数据库为建设鲁、浙海藻场，促进传统肥料和种植企业“引陆下海”，形成陆、海“双生态”产业模式作出了显著贡献，取得了良好的经济效益和生态效益，在海岸带蓝碳研究和生态环保科普方面发挥了积极作用。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	藻类培养基成分数据集	329KB	329KB
2	中国海岸带硅藻分布数据	3.22MB	3.22MB

在线子库
6个

数据资源量
1.1TB

数据系统及服务模式创新与示范
1个

经济效益超
24亿元

联系人
王寅初　ycwang@yic.ac.cn

承建单位
中国科学院烟台海岸带研究所

参建单位
烟台大学
浙江海洋大学

服务成效

支撑科技项目
1个
山东省重点研发计划 2019GHZ026 海洋藻类的资源调查、示范养殖和开发应用

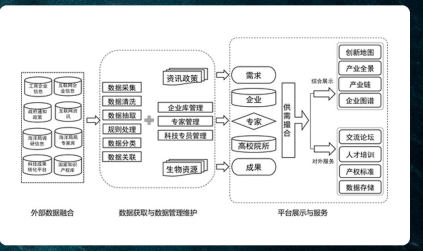
支撑论文
1篇
[1] 王寅初,李文军,任承钢,刘正一,秦松.海洋生物产品服务平台的构建与应用[J].中国科技资源导刊,2022,54(01):40-46+75.

支撑国家标准
1项
《虾青素旋光异构体含量的测定——液相色谱法》 GB/T 38478-2021

典型案例

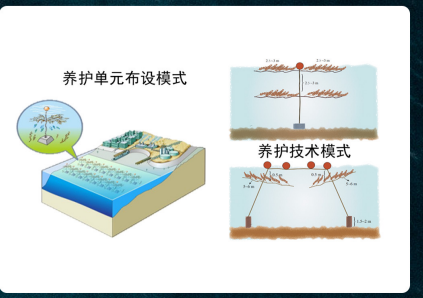
面向海岸带生物资源养护提供线上线下相结合的数据服务

中国海岸带植物资源数据库与海洋生物产品服务平台深度整合，形成了“线上数据+线下服务”的数据服务模式。从海藻资源数据支持到数据服务系统引导，数据库为建设鲁、浙海藻场，促进传统肥料和种植企业“引陆下海”，形成陆、海“双生态”产业模式，做出了显著贡献，取得了良好的经济效益和生态效益。2022年度，数据库团队开辟“海岸带农业大讲堂”专栏，共举办14期专家讲座，总关注人数达9599人次。



为长岛和黄河口区域的海藻场养护与多重生态系统构建提供数据支撑

自2009年起通过对黄渤海海藻资源的变迁进行连续调查分析，获取了中国海岸带藻类数据，为开展海藻场的保护与利用研究提供重要支撑。基于数据指引，聚焦天然海藻资源的保护与利用，对长岛和黄河口区域的海藻场开展养护与多重生态系统构建工作，形成了4900亩示范区。同时在海藻植物资源产业化方面，数据支持了新型海藻肥研发与应用工作。截至目前，基于海藻发酵技术形成的海藻肥生产线已达到30万吨以上年产能。基于数据库服务已发表论文16篇，授权专利3项，撰写企业标准3项。



医学物理数据库

医学物理数据库是我国首个且唯一公开的医学物理交叉学科数据库，由吉林大学承担建设。旨在为我国的医学物理产业发展提供必需的数据研究和服务。该数据库全面覆盖核医学、放疗、医学辐射防护等方向所需的从原子核到器官尺度的微观物理、宏观剂量及各类医学技术数据，包括重离子肿瘤治疗、光子放疗、同位素药物和医疗辐射防护等技术方向所需的相关数据。服务用户及应用方向包括：医院的临床医疗技术研究；药品生产企业的同位素药物生产和研发；软件公司的CT成像和放疗计划设计等AI软件开发；相关高校和研究所的基础研究等，为我国的医学物理产业发展提供基础研究和服务支撑。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	光子肿瘤治疗数据集	205GB	205GB
2	重离子肿瘤治疗数据集	237GB	110GB
3	同位素药物数据集	548GB	220GB
4	医疗辐射防护数据集	352GB	100GB

联系人

杨东　dyang@jlu.edu.cn

承建单位

吉林大学

参建单位

吉林省肿瘤医院

北京航空航天大学

深圳大学

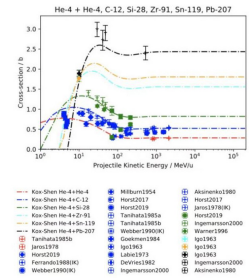
数据集 4个

支撑论文 3篇

资源量 1342GB

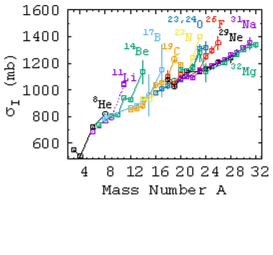
支持科技项目 1个

典型案例



为西南科技大学的重离子反应微观数据评价研究提供实验数据服务

重离子反应微观数据主要服务医学物理的重离子放疗和同位素药物生产技术，航天产业宇航员的辐射防护等问题同样与此类数据相关。目前国际上现有数据计算得到的剂量数据偏差了将近6%。西南科技大学研究团队在重离子反应微观数据库的实验数据基础上针对相关数据开展了评价研究，得出结果表明：与业内领先的德国GSI实验室与美国NASA得出重离子反应微观数据数据相比较，双方数量和质量接近。这个结论对于重离子治疗、同位素药物生产和航天辐射防护将具有重大意义，也是我国首次和国际同步进行数据研究，将为数据理论计算提供基准依据。



为长春工业大学的同位素药物生产方法研究提供实验数据服务

同位素药物生产所需的微观反应截面数据一直是国际上基础数据的研究热点。而我国医用同位素产业刚刚起步，医用同位素几乎完全依赖进口。长春工业大学研究团队基于医学物理数据库提供的重离子的微观反应截面的实验数据，利用Talys等计算程序，研究相关的实验方法。目前国际上已有的相关同位素生产数据库存在两个共性问题，一是只有数据收集，不涉及评价和计算；二是数据不完整。重离子的微观反应截面数据将弥补重要同位素和缺少的数据，该项研究预计为我国医用同位素稳定自主供给体系的建设提供必要方法研究。

国家基础学科公共科学数据中心
National Basic Science Data Center

中国动物主题数据库

中国动物主题数据库以动物形态学、动物生理学、动物分类学、动物生态学、动物地理学、动物疫病学等多个动物界分支学科数据为主体，依据国内外相关研究成果进行建设，是目前国内动物学领域覆盖全面、内容科学的权威数据库。2022年度完成整合地方动物志、中国动物志电子化的卷册数据、鸟类声音语谱图数据库、濒危动物遗传资源数据库及动物图像标注数据库等。截至2022年底，总计在线服务子库38个，增加数据量15GB左右，总服务数据量达1697GB，总记录数约920万条，在线下载量约2TB，全年访问量约42万人次。数据库被浙江大学、南京农业大学、安徽大学、广西大学、大理大学等12家高校和机构图书馆收录，为国家生物安全、动物保护决策、动物学科研究、公民科学普及提供重要数据支撑。

被国内外期刊引用
29次

典型应用案例
6个

国家标准
1项

在线子库
38个

支撑SCI论文
2篇

资源量
1697GB

年度下载量
2TB

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	鸟类声音语谱图数据库	9GB	5GB
2	濒危动物遗传资源数据库	12GB	6GB
3	动物志数据库	8.84GB	1GB
4	地方志数据库	1.4GB	0.2GB
5	动物图像标注数据库	1583GB	3GB

联系人
林聪田　linct@ioz.ac.cn

承建单位
中国科学院动物研究所

服务成效

支撑论文
2篇

论文： [1] Xue, Y.; Lin, C.; Wang, Y.;Liu, W.; Wan, F.; Zhang, Y.; Ji, L.Predicting Climate Change Effects on the Potential Distribution of Two Invasive Cryptic Species of the Bemisia tabaci Species Complex in China. Insects 2022, 13, 1081.

数据论文： [2] 李文杰,王江宁,卜翠萍,葛斯琴,林聪田,韩艳,纪力强.基于动物志的粉蝶形态特征数据集[J].中国科学数据(中英文网络版),2022,7(01):91-98.

支撑国家标准
1项

《中国动物分类代码 第1部分：脊椎动物》国家标准已于2021年10月11日发布，并于2022年5月1日实施

支撑科技项目
2个

1 2022YFB3901103重要生态空间人类干扰活动智能识别技术与应用研究 国家重点研发计划重点专项课题

2 XDA19050200 物种多样性信息平台 中国科学院A类先导科技专项课题

典型案例

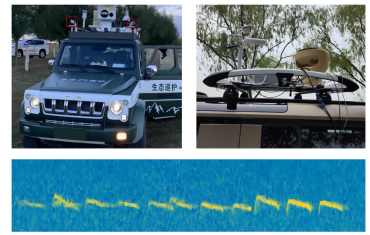
为《2022年北京市密云区陆生野生动物名录-鸟篇》提供权威数据

北京市密云区具有良好的生态环境，生物多样性非常丰富。密云区园林绿化局长期在该区开展生物多样性调查和监测工作，每年动态发布密云区生物物种名录。中国动物主题数据库为北京市密云区园林绿化局提供鸟类名录及分布数据，并通过对密云鸟类名录分类系统与名称进行比对分析，验证鸟类信息准确性，提出修订建议，保障县域生物多样性调查的准确性与科学性。



为生态环境部卫星环境应用中心提供数据及技术支持

生物多样性是生态环境状况与生态系统修复成效的重要指标。如何对生物多样性尤其野生动物进行监测，是生态环境监测与保护的难点之一。2022年中国动物主题库与生态环境部环境卫星应用中心联合开展生物多样性智能监测研究，基于动物声音数据库，利用深度学习训练动物声纹智能监测模型，研发车载形态野生动物监测系统，并在北京延庆区、密云区、朝阳区、内蒙古通辽地区等开展示范应用，在实际场景应用中取得良好效果。该系统作为生态环境部“五基”监测体系组成部分，2022年入选“奋进新时代”主题成就展。



国家基础学科公共科学数据中心
National Basic Science Data Center

中国植物主题数据库

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	中国植物图片数据库	5009624张	20000张
2	古植物标本照片库	14682条	5682条
3	省级植物名录库	14个	4个
4	分子材料库	684种	184种
5	文献志书库	103011本	100本
6	植物物候照片专题库	5230条	5230条

联系人
肖翠　xiaocui@ibcas.ac.cn

承建单位
中国科学院植物研究所

参建单位
中国科学院文献情报中心

中国植物主题数据库是以中国生物物种名录（2022版）为基础建设而成的综合型数据库。该数据库系统整合了植物彩色照片、植物志文献记录、化石植物名录、标本以及药用植物数据，建成33个数据子库及7个特色软件工具，通过免费共享的方式为公众、科研人员、企事业单位等各行业人士提供具有特色的中国植物数据服务。

2022年，中国植物主题数据库新增数据资源54846条，总记录数达到1019.06万条。本年度网站浏览量超过1933万次，在线服务152个国家的59.74万人次用户。

在线子库
33个

总记录数
1019.06万条

特色软件工具
7个

支持科技项目
9个

支撑论文
21篇

支撑专著
1部《海岛花开》

服务成效

支撑学术论文
19篇

[1] 龚治,彭正强,马光昌,等. 自由贸易港背景下的海南外来入侵生物防控探讨[J]. 植物保护,2022,48(06):221-231.
[2] 邵良玉. 基于迁移学习的花类图像分类方法研究[J]. 农业装备与车辆工程,2022,60(07):62-64+99.
[3] 王益和. 福建长汀典型水土流失区木本植物区系分析[J]. 福建林业科技,2022,49(02):94-99.
[4] 席少阳,曹后康,郭延秀,等.甘肃省珍稀濒危药用植物资源种类空间分布差异研究[J].中国中药杂志,2022,47(16):4322-4332.
[5] 刘祥东,杨庆华,付陈龙.未来气候情景下基于MaxEnt模型的微毛樱桃潜在适生区预测[J].江苏林业科技,2022,49(02):11-16.
[6] 杨守生,邝粉良,余诗瑞,等.禄丰樟木箐州级自然保护区药用维管植物资源多样性分析[J].中国野生植物资源,2022,41(02):63-68.
……

支撑数据论文
2篇

[1] 林秦文,肖翠,马金双.中国外来植物数据集[J].生物多样性,2022,30(05):110-117.
[2] 肖翠,刘冰,林秦文,等.北京维管植物编目和分布数据集[J].生物多样性,2022,30(06):5-13.

支撑国家或社会需求
2个

1 服务国家重大需求——北京市外来入侵植物调查
2 支持经济社会发展——北京市温榆河公园植物本底调查

典型案例

“国家重点保护野生植物数据库”为福建长汀县木本植物区系分析提供数据支撑

水土流失严重制约着生态和经济发展。福建长汀县是我国南方红壤地区的典型水土流失区，探明该地区木本种子植物的组成及地理分布，可以为当地生态恢复与植物资源合理开发利用，以及水土流失区的整治与改善提供参考。中国植物主题数据库“国家重点保护野生植物数据库”为福建省龙岩市林业科学研究所提供数据，支撑福建省龙岩市林业科学研究所的研究人员，对福建长汀县的木本植物进行系统分析。相关研究成果以题为《福建长汀典型水土流失区木本植物区系分析》的学术论文发表。



第40卷 第2期　福 建 林 业 科 技　Vol.49 No.2
2022年6月　Jour of Fujian Forestry Sci and Tech　Jun., 2022

doi: 10.13428/j.cnki.fjlk.2022.02.015

福建长汀典型水土流失区木本植物区系分析

王益和

（福建省龙岩市林业科学研究所，福建 龙岩 364000）

摘要：于2019年5—7月、2020年9—11月，采用野外实地调查和样地法相结合的方法，对福建长汀典型水土流失区的木本植物进行统计分析，结果表明：典型水土流失区共有木本种子植物116科、327属、671种（其中：裸子植物3科5属6种，被子植物113科322属665种），优势科植物为壳斗科、松科、樟科、山茶科、杜鹃花科、蝶形花科、禾本科、桑科、冬青科、金缕梅科、忍冬科、蔷薇科、漆树科、芸香科、大戟科、茜草科、紫金牛科、安息香科、木犀科、马鞭草科、唇形科、茄科、玄参科、紫葳科、苦苣苔科、爵床科、茜草科、葫芦科、菊科、百合科、天门冬科、薯蓣科、兰科、棕榈科、禾本科、莎草科、灯心草科、鸭跖草科、姜科、芭蕉科、天南星科、石蒜科、鸢尾科、百部科、仙茅科、水玉簪科、露兜树科、菝葜科、黄杨科、胡颓子科、五加科、伞形科、杜英科、椴树科、锦葵科、梧桐科、瑞香科、千屈菜科、桃金娘科、野牡丹科、柳叶菜科、八角科、五味子科、木兰科、番荔枝科、樟科、毛茛科、小檗科、防己科、木通科、罂粟科、十字花科、景天科、虎耳草科、海桐科、金缕梅科、杜仲科、悬铃木科、蔷薇科、豆科、酢浆草科、牻牛儿苗科、芸香科、苦木科、楝科、远志科、大戟科、漆树科、无患子科、槭树科、七叶树科、清风藤科、凤仙花科、鼠李科、葡萄科、椴树科、锦葵科、梧桐科、瑞香科、胡颓子科、千屈菜科、桃金娘科、野牡丹科、柳叶菜科、五加科、伞形科、山茱萸科、杜英

国家基础学科公共科学数据中心
National Basic Science Data Center

中国植物DNA条形码数据库

中国植物DNA条形码数据库是我国植物DNA条形码科学数据汇聚和共享服务平台。其依托中国西南野生生物种质资源库的持续积累，开展以中国植物资源为核心的DNA条形码加工、采集和整合，共有中国植物DNA条形码数据26.1万条，覆盖中国植物2584属11911种物种，2022年度新增重要经济作物和经济作物近缘种蔓菁、白菜、竹类、沙生槐植物，寄生/半寄生植物模式物种向日葵列当、埃及列当、马先蒿、独脚金等植物全基因组数据8份，所有植物DNA条形码数据资源面向公众开放共享。平台提供植物DNA条形码检索和在线对比服务，基于DNA条形码推动新一代智慧植物志（iFlora）建设，为国家战略生物资源的精准鉴定、综合管理和相关科学数据服务提供重要支撑。

资源总量
69.5GB

在线访问浏览量
285万人次+

资源下载量
112.87GB

支撑科技项目
3个

支撑论文
1篇

软件著作权
1项

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	中国植物DNA条形码数据库	26.1万条	3.5万条

联系人
庄会富　zhuanghuifu@mail.kib.ac.cn

承建单位
中国科学院昆明植物研究所

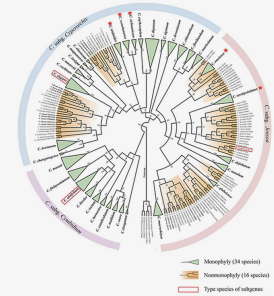
服务成效

中国植物DNA条形码数据库建立了植物各大类群的DNA条形码技术规范体系，构建了中国最大的野生植物DNA条形码参考数据库和物种鉴定平台，相关工作显著提高了物种鉴定功能准确率，提升了中国西南野生生物种质资源库科学研究和公共服务的能力。共支撑发表论文138篇，相关成果受到国内外同行的广泛关注和正面评价。先后培养DNA条形码专业技术人员30余人，持续为我国植物DNA条形码研究输送和培养骨干力量。

典型案例

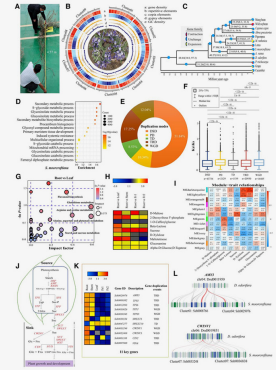
兰属植物DNA条形码研究取得新进展

兰属（Cymbidium）植物具有较高的经济、文化、保护和研究价值，而物种准确鉴定是开展生物多样性保护、资源可持续利用和系统演化研究的重要前提和保障。中国植物DNA条形码研究团队通过合作研究对兰属植物进行了广泛取样，运用二代基因组测序技术获得了50个物种237个个体的基因组浅层测序数据，并基于序列分析（Barcoding gaps和ABGD）、建树分析（ML、BI和mPTP）和基因组遗传距离分析（Skmer）等方法对兰属DNA条形码深入研究，积累了完整的DNA条形码参考数据，并提出了兰属物种优化鉴定策略，实验表明Skmer方法具有最高的物种识别率。相关研究成果以《DNA barcoding of Cymbidium by genome skimming: call for next-generation nuclear barcodes》为题发表在该领域国际期刊《Molecular Ecology Resources》上。



砂生槐基因组为西藏高原生态修复适宜树种提供数据支持

砂生槐（Sophora moorcroftiana）是西藏地区分布极广的一种多年生的豆科类灌木，其根系发达，具有极强的耐干旱、耐贫瘠、抗风沙的生态特征，然而对砂生槐基因组学以及抗旱性的分子机制的研究仍相对较少。中国植物DNA条形码研究团队通过二三代测序技术，组装和注释了染色体水平高质量的砂生槐基因组，基因组大小为737.35Mb，包括9条染色体、contig N50为22.5Mb、注释了53,640个蛋白编码基因、BUSCO为98.3%。比较基因组学分析显示，来自于中高海拔的豆科植物如砂生槐基因组中扩张的基因家族以及重复基因主要在代谢过程中发挥作用，可能与砂生槐的高原适应相关，分析揭示蔗糖代谢对其干旱适应性的作用，砂生槐基因的发现可以作为培育耐旱性高原植物的候选基因。相关研究成果以《Sophora moorcroftiana genome analysis suggests an association between sucrose metabolism and drought adaptation》为题发表在期刊《Plant Physiology》上。



化学主题数据库

化学主题数据库是国内目前规模最大、稳定在线服务的化学数据库，汇聚了中国科学院化学化工学科领域80%以上的基础数据资源，在线服务数据子库40余个，数据覆盖化学化工研究主要的二级学科。自主开发建设国内唯一的化合物登录系统，截至2022年底累计登录化合物870万余种，并以此为基础支撑基于化合物唯一标识的异地异构化学数据资源整合共享、数据一致性核查以及专题数据库的构建，保证数据库的可持续发展；面向全球提供化合物全景式的化学化工科学数据检索展示、化学数据加工及面向项目的专题数据服务。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	化合物登录数据库	870万种化合物	5万种化合物
2	化合物生物活性数据库	50万种化合物	5.6万种化合物
3	工业污染防控数据库	4000条记录	1000条记录
4	金属离子形态数据集	750条记录	750条记录
5	中草药单一化学成分质谱数据库	2800条记录	100条记录

联系人

赵月红　yhzhao@home.ipe.ac.cn

承建单位

中国科学院过程工程研究所

参建单位

中国科学院上海有机化学研究所

中国科学院长春应用化学研究所

服务成效

化学主题数据库聚焦化学化工领域科学数据在线服务以及离线专题服务，资源总量超200GB，2022年度累计资源访问量约为35万人次。共计为100余个课题/项目提供数据支撑，其中涉及10余个国家重点研发计划/重大科技专项等国家级科研项目，40多项国家自然科学基金项目、中国科学院项目、省部级科研项目和企业委托项目；共支撑发表论文60余篇，申请专利2项、获省部级奖1项、设计开发领域数据分析软件3项。

资源总量
200GB+

资源访问量
35万人次

数据支撑课题/项目
100个+

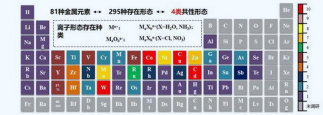
支撑论文
60篇+

专利
2项

省部级奖
1个

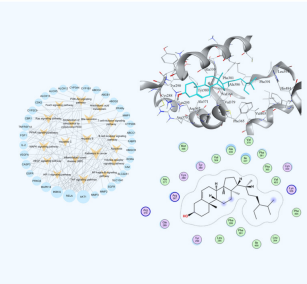
设计开发领域数据分析软件
3个

典型案例



为工业污染控制研究提供数据支撑

废水中关键金属离子高效分离是实现废水无害化处理、资源化利用的关键，也是国际化学界长期关注的难题。基于化学主题数据库、文献资料及实验数据，科研团队收集金属离子形态相关的结构、物化性质等多维度分子描述信息，构建了包含V、Cr、W、Mo、Ni、Co、Mn等多种金属离子的专题数据集，通过结合量化计算，建立金属离子–萃取剂构效关系预测模型，支撑萃取剂的智能设计筛选。相关工作为国家自然科学基金杰出青年基金（No.51425405）、重点基金（No.51934006）、面上基金（No.51178446）、青年基金（No.21206168）、国家科技支撑计划（No.2015BAB02B05）等项目研究提供数据支持。



支撑乐草石有效成分与治疗乙肝机制研究

乐草石(LCS)是一种治疗乙型肝炎的民间方剂，LCS配方水提物的抗乙肝作用已经通过HepG2.2.15体外模型和shelduck乙型肝炎体内模型进行了验证，但其抗乙肝机制尚未明确。科研团队通过对化合物登录库、植物化学成分数据库和化合物生物活性数据库跨库检索，获得乐草石方剂213种化学成分以及12种关键化合物的基本性质数据和生物活性数据。为研究人员通过网络药理学和肠道菌群分析技术，开展乐草石方剂及其单药抗乙肝作用的潜在机制基础研究提供数据支撑。研究成果发表在《Biomedicine&Pharmacotherapy》，DOI: 10.1016/j.biopha.2022.112692.

国家基础学科公共科学数据中心
National Basic Science Data Center

原子分子数据库

原子分子数据库由北京应用物理与计算数学研究所承担建设，是目前国内唯一的原子分子领域数据库。原子分子数据在天体物理、武器物理、聚变能源（ICF、Tokomak）和工业生产等领域都有重要应用，与国际上的IAEA原子分子数据组、俄罗斯spectr-w3数据库建有长期合作关系。该数据库现有原子光谱、能级、电离势、电子与原子分子碰撞等多个方面的科学数据200余万条。可全部在线检索，实现数据免费共享服务。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	原子光谱数据库	189万条记录	2.0万条记录

在线子库
7个

支撑发表论文
2篇

资源量
183GB

年度下载量
10GB

联系人
刘春雷　liu_chunlei@iapcm.ac.cn

承建单位
北京应用物理与计算数学研究所

服务成效

支撑论文
2篇

[1] Chensheng Wu, Yong Wu, Jun Yan, T. N. Chang, and Xiang Gao, Transition energies and oscillator strengths for the intrashell and intershell transitions of the C-like ions in a thermodynamic equilibrium plasma environment, Phys. Rev. E 105, 015206 (2022)

[2] Meng, Ju; Li, Wen-Xian; Li, Ji-Guang; Wu, Ze-Qing; Yan, Jun; Wu, Yong; Wang, Ji-an-Guo,M1 transition energy and rate in the ground configuration of Ag-like ions with 62 <= Z <= Chin Phys B 94,31,13101(2022)

支撑科技项目
2个

1 2017YFA0403200 高温高压高密度物质的压缩性质、不透明度和输运系数研究 国家重点研发计划

2 TA28 非平衡核反应动力学模型研究　中国工程物理研究院重大课题

典型案例

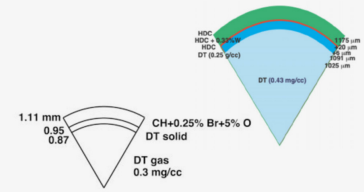
为国家重点研发计划“高温高压高密度物质的压缩性质、不透明度和输运系数研究”提供数据支持

高温高压高密度物质的不透明度研究，需要Fe等元素的原子能级和光谱的系统数据。这些数据广泛被应用于高温稠密等离子体实验谱的理论模拟与分析中，也被应用于宽温度、密度状态区间等离子体辐射不透明度参数的理论研制中。原子分子物理主题数据库为其了解实验测量中高温稠密碳等离子体靶的状态分布、能量沉积实验测量结果的理论模拟和解释做出了贡献。

电子数	能级数目	谱线数目
1	98	802
2	9,058	104,782
3	6,993	34,302
4	50,980	958,762
5	186,924	6,894,758
6	435,419	20,829,355
7	640,687	34,490,044
8	682,086	31,061,859
9	480,343	17,687,903
10	232,617	5,243,799
11	23,875	54,479
12	302,901	4,144,240
13	1,467,530	45,710,289
14	5,902,019	194,883,463

为中国物理工程研究院“非平衡核反应动力学模型研究”TA28课题组提供数据支撑

DT聚变中，α粒子在DT等离子体中的能量沉积是聚变研究的主要参数，聚变靶设计中必需掺有其它如含碳的材料。在聚变中这些材料会不可避免地混入DT燃料区域，也会受到α粒子的加热，所以须研究α粒子在混合物等离子体中的能量沉积或阻滞本领参数。数据库提供了能量沉积或阻滞本领参数，应用于非平衡核反应动力学模型研究，为课题组得到更高精度的聚反应理论模拟、更可靠的等离子体状态诊断结果做出了贡献。



国家基础学科公共科学数据中心
National Basic Science Data Center

光学技术数据库

光学技术数据库是目前国内唯一的光学领域数据库，包含有光学镜头数据、光学材料数据、光学系统常用国家标准和光电元器件库等公开数据。该数据库始于1987年中国科学院长春光学精密机械与物理研究所牵头建立的光学专业库。目前已建成5个主题库，21个子库、数据资源总量858GB，网站访问量17万人次。通过集成国产化自主可控光学软件的高质量仿真数据，加强数据库的应用效能。目前已建成数据、软件、应用三位一体管理模式，为国内外上千家行业单位、数十万用户提供大量的科学数据，帮助科研人员取得了多项重要科研成果。

联系人

娄洪伟　louhw@ciomp.ac.cn

承建单位

中国科学院长春光学精密机械与物理研究所

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	光学玻璃数据库	15GB	5GB
2	光学晶体数据库	5GB	10MB
3	光源数据库	2GB	100MB
4	光电探测器数据库	5GB	205MB
5	光学镜头数据库	11GB	1GB
6	光学标准库	3GB	2GB
7	光学成像数据库	381GB	1GB
8	光学检测数据库	90GB	10GB
9	特色光谱信息数据库	160GB	40GB
10	激光器件数据库	16GB	14GB
11	激光传输特性数据库	15GB	14GB
12	激光光学系统数据库	23GB	22GB
13	短波光学数据库	2GB	1GB

成像光学仿真与案例数据
58GB

杂散辐射仿真与案例数据
77GB

激光仿真与案例数据
32GB

可见光和红外成像领域数据
45GB

支撑论文
20篇

支撑国家重大任务
6个

数据挖掘分析工具
3个

服务成效

支撑论文
2篇

[1] Sheng, L.; Zhao, W.; Zhou, Y.; Lin, W.; Du, C.; Lou, H. A Surrogate Model Based
[2] Multi-Objective Optimization Method for Optical Imaging System. Appl. Sci. 2022, 12, 6810. https://doi.org/10.3390/app12136810

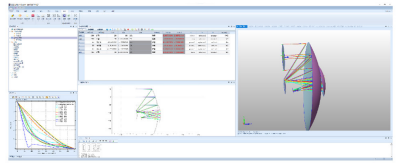
软件著作权
3项

1 基于GPU并行计算的资源弹性调度软件
2 基于单机加云一体化的光学仿真平台
3 基于CPU+GPU异构并行加速的多目标优化软件

典型案例

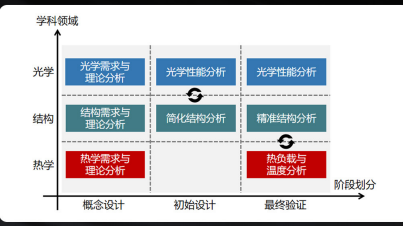
为国产化光学设计分析软件的研制提供光学基础数据

光学技术数据库依托光学镜头国产光学设计软数据、光学材料数据、膜系数据、样板数据、光学加工数据、装配工差数据、光源数据、散射数据等资源，成功研制并发布了国产光学设计软件内测版。同时，光学技术数据库为软件提供了光学相关数学计算模型，解决了软件使用初始值问题。这项工作将我国独有且先进的光学技术数据，从独立数据结构的数据形式转换为可被光学设计软件读取的格式，极大提高了光学技术数据库的使用率，对我国核心软件自主化生产和生态形成做出了重要贡献。



为人工智能通用光学系统仿真优化软件研究提供基础数据

光学技术数据库数据为光学智能引擎软件研发，提供光学玻璃、光学晶体、光学塑料、光源、光电探测、光学传感等基础数据，目前已应用于光学智能引擎软件测试和AI训练中。通过向用户提供光学传感器阵列试验数据智能化分析，混合现实交互、多维信息可视化、光学仿真求解器等人工智能服务，大大促进该库数据在光学智能引擎软件研发中的利用率。



国家基础学科公共科学数据中心
National Basic Science Data Center

核物理主题数据库

联系人
金永利　jyl@ciae.ac.cn

承建单位
中国原子能科学研究院

核物理主题数据库是核基础数据信息化合集，收录数据来自我国或者国际权威机构，主要包括中子、带电粒子等引起的反应数据、裂变产额、核结构和衰变数据等。这些数据是表征原子核自身特性和原子核与其它粒子发生反应，及核基础研究、核装置设计与运行、核技术应用的重要数据。其子库中国评价核数据库由我国自主研发、也是国际五大主流核数据库之一，具有较高的国际影响力。2022年首次在国内系统化建立了中子活化截面数据库，包含818个核素，已为我国重大核工程如示范快堆等研究提供了数据支撑。

在线子库
11个

资源总量
1.1TB

科研院校申请专用账号
78个

专利
8项

软件著作权
6项

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	评价核数据库	1113.5GB	279GB
2	实验核数据库	992MB	20MB
3	裂变产额数据库	90MB	30MB
4	中子活化反应截面数据库	400MB	400MB

服务成效

支撑论文
8篇

[1] Neutron activation cross section data library，Chin. Phys. B 31, 060102 (2022)
[2] $^{142-146,148,150}$Nd光核反应理论计算，原子能科学技术， Vol. 56（5), 2022
[3] 裂变产额的Zp模型评价方法及燃耗不确定度分析，原子能科学技术， Vol. 56（5), 2022
[4] 宏观-微观模型和朗之万方法在低能核裂变研究中的应用，原子能科学技术， Vol. 56（5), 2022
[5] 核数据评价与中国评价核数据库CENDL，原子能科学技术， Vol. 56（5), 2022
[6] 热中子诱发239Pu裂变初级裂变产物质量分布测量，原子能科学技术， Vol. 56（5), 2022
[7] Impact of nuclear dissipation on the fission dynamics within the Langevin approach，PHYSICAL REVIEW C，105,034614(2022)-1-11
[8] Neutron capture cross section of 169Tm measured at the CSNS Back-n facility in the energy region from 30 keV to 300 keV，Chinese Physics C，Vol. 46, No. 4 (2022) 044002-1-13

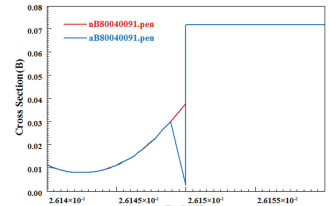
支撑科技项目
2个

1　2022YFA1602403 准单能伽马源的光核反应与关键技术研究 国家重点研发计划重点专项课题
2　WDJC-2019-09 铀钚核数据产生的关键基础技术研究 财政部稳定支持项目

典型案例

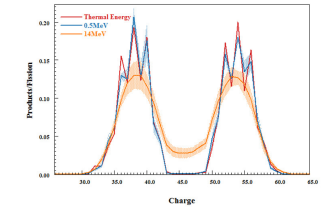
为“液态金属冷却快堆关键设计分析软件工程验证”项目提供数据和软件支持

在与西安交通大学合作开展的“液态金属冷却快堆关键设计分析软件工程验证”项目研究中，需要对核数据处理软件LoongATLAS进行测试。核物理主题数据库研究团队从多个版本的国际五大评价核数据库中筛选了具有不同格式特性的评价核数据，利用LoongATLAS和NJOY2016程序加工选出的评价数据，制作了ACE格式连续能量点截面库和WIMS多群常数库，并通过比对分析发现了LoongATLAS软件的多种缺陷。最终核物理主题数据库被选为项目重要评价数据来源，研发软件NDPlot作为评价核数据绘图比对工具得到应用，有效支撑了研究顺利开展。



为解放军某部提供裂变产额数据

核物理主题数据库研究团队与中国人民解放军某部基于核物理主题数据库平台合作研制了裂变产额数据（约1100个产物核）检索程序。核物理主题数据库为其提供了裂变产额数据库的数据。裂变产额数据库基于ENDF/B-VIII.0、ENDF/B-VII.1、JENDL-4.0、JEFF-3.1、JEFF-3.3以及CENDL-TMSR裂变产额数据库，包含233U、235U、238U、239Pu、240Pu等37个典型裂变系统3个能点（热能点、裂变谱和14MeV）的中子诱发裂变产物核的裂变产额数据。



宇宙学数值模拟数据库

宇宙学数值模拟数据库是中国科学院紫金山天文台承建，以该单位的“宇宙大尺度结构与星系形成高性能计算平台”为硬件依托的天文数据库。该库是目前国内体量大而且种类全的宇宙学模拟数据库之一，科学数据总量1.5PB左右，并以每年约20%的速度增加。其中“FAST观测图像数据库”与“中国空间站工程巡天望远镜CSST的观测数据仿真数据库”是支撑国家重大科学项目“天眼”与被誉为中国的哈勃太空望远镜的“中国空间站工程巡天望远镜”的数据库，已广泛为科研人员和工程人员提供数据服务。截止目前，该数据库已为国内外120余位天文专业注册用户提供数据服务。2022年在PRL、ApJ、MNRAS等国际期刊上发表高质量学术论文21篇。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	银河系星流数值模拟数据集	20GB	20GB
2	孪生宇宙模拟	6300GB	6300GB
3	中国空间站工程巡天望远镜CSST的观测数据仿真7平方度	103010GB	103010GB

联系人

王蕾　leiwang@pmo.ac.cn

承建单位

中国科学院紫金山天文台

在线子库
12个

支撑论文
21篇

资源总量
1.47PB

年度下载量
420TB

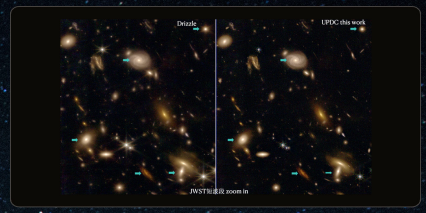
专利
4项

软件著作权
2项

获科技奖项
3个

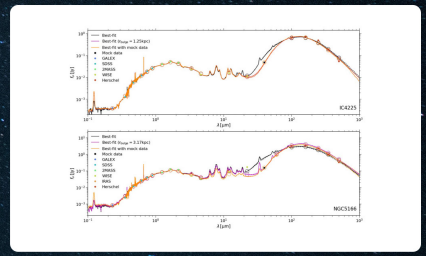
支撑科技项目
10个

典型案例



支持“面向超分辨率的多次曝光图像叠加”算法研究

中国科学院紫金山天文台和浙江大学的研究人员，通过改进软件算法技术，获得了超过传统光学衍射极限的超分辨率图像。该技术不仅能提高分辨率还可以实现超采样、或欠采样图像的反混叠。该算法已经运用于我国正在筹建中的“中国空间站工程巡天望远镜”（被誉为中国的哈勃太空望远镜）的科学数据处理管线，并在詹姆斯韦伯太空望远镜（JWST）数据处理中取得良好效果。该成果以《Towards super-resolution via iterative multi-exposure coaddition》为题发表在天文领域重要期刊《Monthly Notices of the Royal Astronomical Society》上。



支持“基于数值模拟和机器学习的多波段SED模型STARDUSTER”研究

在实际观测中，受限各种吸收效应、滤光片技术、望远镜波段等，星系的能谱（SED），特别是多波段往往是不连续的，这给星系形成与演化模型带来了挑战。近些年，随着数值模拟技术的进步，使得人们可以基于数值模拟和机器学习来预言星系的多波段SED。中国科学院紫金山天文台和浙江大学的研究人员，通过宇宙学数值模拟和改进的机器学习算法成功预言了星系的多波段SED，与观测数据对比，获得了很高的精度。该技术为星系的多波段能谱生成和预言提供了一条可行路径，为将来更高精度的宇宙学数值模拟、星系数值模拟、以及更好地还原星系演化历史打下了基础。成果以《Stardust: A Multiwavelength SED Model Based on Radiative Transfer Simulations and Deep Learning》为题发表在天文领域重要期刊《The Astrophysical Journal》上。

国家基础学科公共科学数据中心
National Basic Science Data Center

VLBI月球与深空探测数据库

VLBI (Very Long Baseline Interferometry)月球与深空探测数据库是基于中国VLBI网利用甚长基线干涉测量，开展国家重大科技专项所属的月球与行星探测任务观测数据建立的公共数据库。该数据库在网站上提供数据公开检索服务，并通过工程内网建立镜像站点，提供便捷的数据服务。近年直接服务于我国探月工程嫦娥系列探测器和首次火星探测天问一号探测器的测轨定位。

探月工程火星探测历次任务VLBI测量数据是国家月球与深空探测重大专项过程中积累的重要数据集，是我国射电天文观测装置服务于国家重大科技专项的科学数据产出，具有唯一性和独有性特征。未来还将支持探月工程四期、天问二号小行星探测和地月VLBI科学试验的顺利开展，也将服务后续的月球与深空探测等各项任务。

2022年度新增数据资源清单

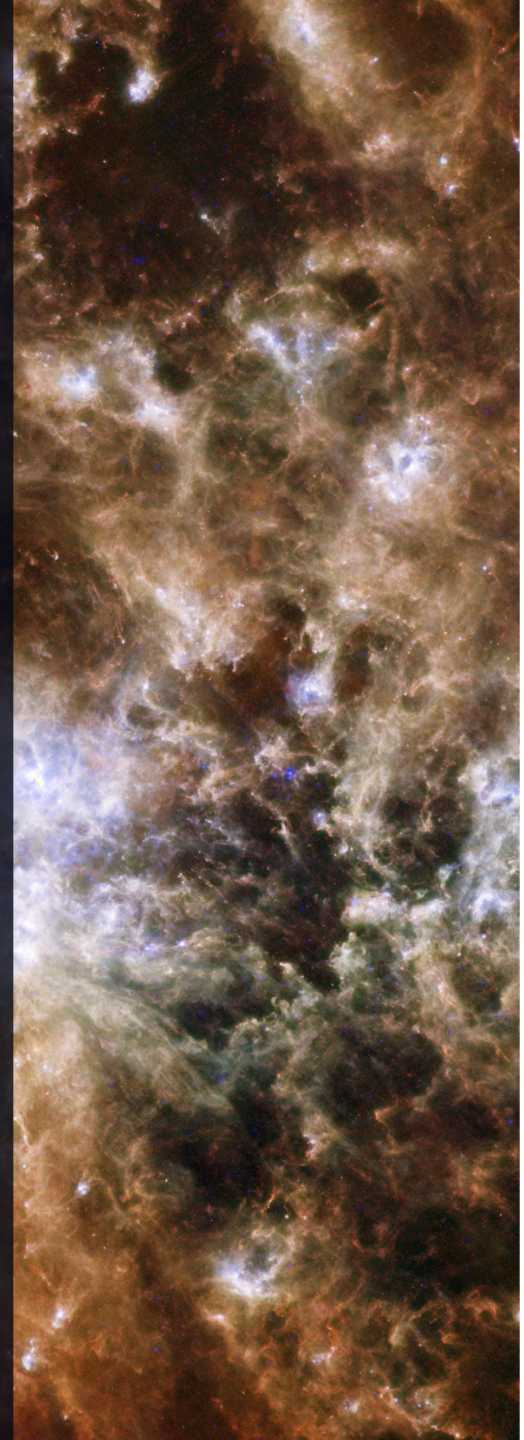
序号	数据库/集名称	数据总量	年度增量
1	嫦娥四号VLBI测量数据集	1.96TB	160GB
2	嫦娥五号VLBI测量数据集	2.58TB	77GB
3	天问一号VLBI测量数据集	3.3TB	290GB

联系人

陈中　zhchen@shao.ac.cn

承建单位

中国科学院上海天文台



服务成效

支撑科技项目 4个

1 2018YFA0404704 超宽带低噪声低功耗致冷接收系统关键技术研究 国家重大研发计划
2 探月工程三期嫦娥五号任务、探月工程嫦娥四号任务 国家科技重大专项
3 探月工程四期与行星探测 国家科技重大专项
4 首次火星探测天问一号任务 国家民用航天工程

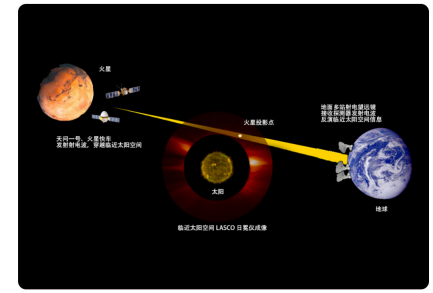
支撑论文 9篇

[1] 李婷,童锋贤,郑为民,张娟. GEO卫星的VLBI快速高精度定位观测. 天文研究与技术, 2022,19(04), 305-316
[2] 陈奥,童力,郑为民, 用于VLBI 测轨的新型Gold 码调制PN-DOR 信标的设计与仿真.天文学进展，2022, 40(2):271-283
[3] 朱晓娜,童力,郑为民,张娟. 一种用于空间VLBI的射电源条纹搜索算法. 天文研究与技术, 2022,19(6)：568-575
[4] Xu Zhao, Wei-Min Zheng, Li Tong, Mao-Li Mao. Dual-target Signal Simulation of Very Long baseline Interferometry for Chang'E-5. PASP，2022，Volume 134, Number 1038
[5] Lei Liu, Weimin Zheng, Jian Fu, and Zhijun Xu. OmniUV: A Multipurpose Simulation Toolkit for VLBI Observation，The Astronomical Journal，2022.164.67
[6] Ting Li, Lei Liu, Wei-Min Zheng, Juan Zhang. The Precision Analysis of the Chinese VLBI Network in Probe Delay Measurement, Research in Astronomy and Astrophysics.2022, Volume 22, Number 3
[7] Kazunori Akiyama, Juan-Carlos Algaba, Tao An, Keiichi Asada, Kitiyanee Asanok, Do-Young Byun, Thanapol Chanapote, Wen Chen, Zhong Chen, Xiaopeng Cheng, et al. Overview of the Observing System and Initial Scientific Accomplishments of the East Asian VLBI Network (EAVN),Galaxies. 2022, 10(6), 113
[8] Maoli Ma , Guifré Molera Calvés, Giuseppe Cimò, Ming Xiong, Peijia Li1, Jing Kong, Peijin Zhang, Jiansen He, Lijia Liu, Pradyumna Kummamuru, Chuanpeng Hou, Jasper Edwards, Qinghui Liu, Zhong Chen, Zhanghu Chu, et al. Detecting the Oscillation and Propagation of the Nascent Dynamic Solar Wind Structure at 2.6 Solar Radii Using Very Long Baseline Interferometry Radio Telescopes. The Astrophysical Journal Letters.2022, 940:L32 (10pp),
[9] Zamri Zainal Abidin, Farah Ramadhani, Noraisyah Mohamed Shah, Zainol Abidin Ibrahim, Chen Zhong, Mahidzal Dahari, Mohd Shazwan Mohd Radzi. Optimal performance of data transmission between Malaysia-China for VLBI-based radio astronomy observation using Jive5ab. Zamri Zainal Abidin,Farah Ramadhani, Noraisyah Mohamed Shah,Zainol Abidin Ibrahim,Chen Zhong,Mahidza Dahari. Experimental Astronomy. 2022. 53:1165—1176

典型案例

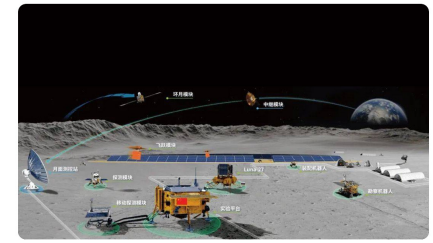
支持火星探测器掩星观测开展太阳风和星际介质科学研究

火星日凌是指地球、火星运行至太阳两侧且三者近乎处于一条直线的自然现象。火星日凌期间，通过研究探测器经过临近太阳空间向地球测控站发射的无线电信号所受太阳电磁辐射干扰导致其信号强度和频率发生变化，可反演研究太阳活动情况。VLBI数据库通过提供探月与探火历史测量数据，支持研究和观测方案设计，服务数据管理和国内外科研共享。利用多个VLBI站开展火星探测器高灵敏度联合观测，获取高时间分辨率和高空间分辨率数据，解决了原位探测器无法进入的临日空间和光学手段无法识别的空间小尺度快速变化的限制，极大的帮助研究人员更加细致地研究临日空间环境及其对深空通信的影响。



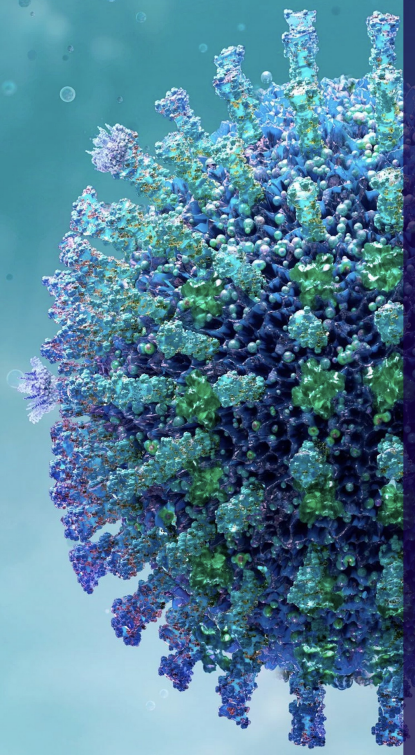
支持月球与深空探测等后续项目论证与立项

VLBI数据库支持月球与深空探测后续重大专项和空间VLBI阵项目的论证，包括探月工程四期、行星探测天问二号小行星采样返回、嫦娥七号中继星月球轨道VLBI试验、天问三号关键技术攻关、近地小行星防御论证、空间VLBI低频射电阵、科技创新2030重大项目等。上述各个空间和地面重大工程项目建设和论证中，均使用VLBI数据库中的探月工程一期~三期和首次火星探测天问一号VLBI测量数据进行项目论证或关键技术研究。相关数据提升了技术方案验证效率，特别是在轨道设计和测控能力分析方面，为各项目总体提供了必要数据支撑。



国家基础学科公共科学数据中心
National Basic Science Data Center

中国病毒资源基础数据库



中国病毒资源基础数据库是由中国科学院武汉病毒研究所牵头建设，服务病毒学领域的专业数据库。该库依托“国家病毒资源库”，是以国家战略生物资源为核心建立融病毒资源保藏、分布、遗传信息及病毒学科研究重点领域信息的共享数据库。按照保藏资源类型下设16个子数据库和13个数据集。为研究者探索和理解病毒的结构、功能、遗传信息和病原性等方面提供了宝贵的数据和信息，为病毒学领域的研究和应用奠定了坚实的基础。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	极端环境宏基因组数据集	40.1GB	40.1GB
2	2022年湖北地区蜱传病毒宏基因组数据集	100GB	100GB
3	2022年海南地区蜱传病毒宏基因组数据集	100GB	100GB

联系人
范兆军　roy@wh.iov.cn

承建单位
中国科学院武汉病毒研究所

服务成效

支撑数据论文
2篇

[1] Zhang Y, Hu B, Agwanda B, et al. Viromes and surveys of RNA viruses in camel-derived ticks revealing transmission patterns of novel tick-borne viral pathogens in Kenya[J]. Emerging Microbes & Infections, 2021, 10(1): 1975-1987.
[2] Wang J, Xiao J, Zhu Z, et al. Diverse viromes in polar regions: A retrospective study of metagenomic data from Antarctic animal feces and Arctic frozen soil in 2012—2014[J]. Virologica Sinica, 2022, 37(6): 883-893.

软件著作权
5项

1. 样本出入库系统V1.0　2022SR0696565
2. 样本检测引物序列处理系统V1.0　2022SR1030993
3. 样本检测系统V1.0　2022SR1030992
4. 微斑图像识别软件V1.0　2022SR0696566
5. 高通量测序数据病毒自动识别软件V1.0　2022SR0260571

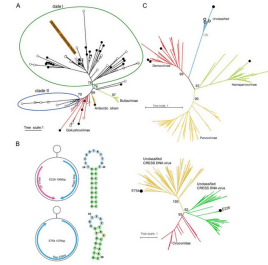
支撑科技项目
3个

1. 1. 31900154 深度挖掘杆状病毒蛋白质互作网络　国家自然科学基金课题
2. 2022YFC2302701 跨境虫媒传染病防控病原监测体系研究与口岸应用　国家重点研发计划
3. KJZD-SW-L11　冰川、冻土等极端生境和水体环境中“休眠”病毒复苏的潜在传播风险 中国科学院重点部署项目

典型案例

“极端环境宏基因组数据集”为极地地区病毒多样性研究提供原始数据支撑

病毒是极地生态系统的重要组成部分，环境原因导致极地地区动物和环境样本难以获得，使得人类对极地地区的病毒了解非常有限。中国病毒资源基础数据库向研究团队提供南北极地区的病毒多样性基础病毒组学数据，科研团队对2012-2014年收集的南极动物粪便和北极冻土宏基因组测序数据进行了回顾性分析，研究结果表明南极动物粪便和北极土壤样本分别含有至少23个科和16个科的DNA和RNA病毒，该项研究揭示北极土壤、北极淡水和南极土壤中南北极病毒科变异之间的遗传相似性和功能相关性。



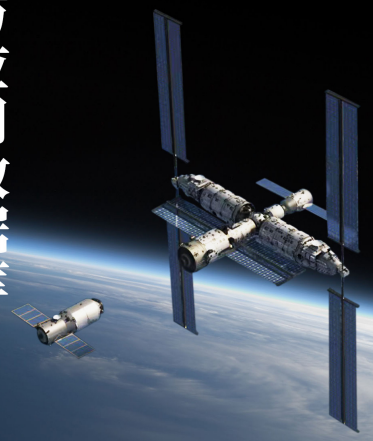
开展微生物知识-病毒主题系列科普活动，提升公民病毒认知

“只要有生命的地方，就有病毒存在”，然而公众对病毒的认知非常有限。作为独具特色的病毒学研究机构，中国病毒资源基础数据库结合自身优势，依托国家病毒资源库、病毒标本馆，2022年度联合湖北省科协、中国科学院武汉病毒研究所、湖北省微生物学会、武汉微生物学会等多家机构，通过“常见的传染病，如何避免‘中招’”、“为什么你总是感冒，科学家帮你找答案”、“P4实验室中的‘宇航服’有多牛，快来涨姿势”、“病毒到底长啥样，放大N倍给你看”等多个有趣话题开展病毒系列主题科普活动，激发公众对病毒学的关注和兴趣，引导公众正确认识病毒。



国家基础学科公共科学数据中心
National Basic Science Data Center

载人航天空间科学与应用数据库



载人航天空间科学与应用数据库2022年度持续为各领域、多用户提供历次任务（天宫一号、天宫二号、天舟一号）空间科学与应用存档数据。同时面向载人空间站工程开展了天和核心舱、问天实验舱、梦天实验舱共计14个科学实验柜和舱外实验载荷的近百种空间科学数据产品的共享服务工作，涵盖空间生命生态、空间微重力基础物理、微重力流体物理等8大学科领域。该数据库高效管理载人航天空间科学与应用各领域、各类型数据产品，为用户提供数据共享、成果管理、专题数据集等服务，助推具有国际影响力的前沿科学发现和应用研究成果产出。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	多模态遥感场景分类数据集MRSSC2.0	211MB	211MB
2	空间站空间科学与应用领域数据库	25TB	25TB

数据总量
400TB+

数据条目
40万条

服务用户
1000家+

累计下载
60万次+

平台访问量
120万次+

联系人
李轩　lixuan@csu.ac.cn

承建单位
中国科学院空间应用工程与技术中心

服务成果

支撑论文
3篇

[1] Liu K, Yang J, Li S. Remote-Sensing Cross-Domain Scene Classification: A Dataset and Benchmark[J]. Remote Sensing, 2022, 14(18): 4635.

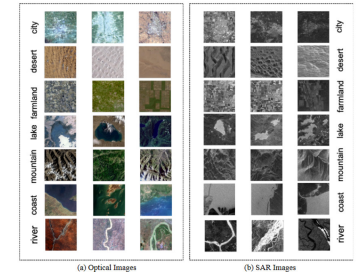
[2] LIU Siqing,SUN Yueqiang,ZHONG Qiuzhen,WEI Fei,LIN Ruilin,LI Zhitao,REN Tingling,LI Yongping,ZHANG Xian-guo,ZHANG Shenyi,LUO Bingxian,CHANG Zhen,WANG Chunqin,PENG Songwu,AI Jiangzhao.Space Environment Measurements and Applications on the Tianhe Core Module of China Space Station[J].空间科学学报,2022,42(04):744-753.

[3] 李永平,孙越强,张贤国,郑晓亮,艾讲赵,李佳,王玉洁,朱光武,陈光明.中国空间站运行轨道上大气密度就位探测[J].科学通报,2023,68(01):128-136.

典型案例

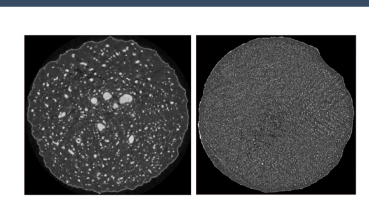
跨域遥感场景发布多模态遥感场景分类数据集

2022年度数据库团队制作并发布了多模态遥感场景分类数据集（MRSSC2.0），该数据集是基于天宫二号宽波段成像仪和三维成像微波高度计获取的高质量对地观测数据，所构建的可见光、短波红外、热红外和SAR共4种模态的跨域遥感场景分类数据集。包含城市、农田、山区、沙漠、海岸、湖泊和河流共7个典型场景的26710幅图像。该数据集资源已通过数据共享服务平台（http://www.msadc.cn/-main/setsubDetail?id=1552963250075734018）对公众放开访问，为国内外学者开展基于人工智能+遥感图像的域迁移学习的应用研究，提供数据源，将激发多模态遥感场景分类的创新研究思路和算法研究。



助力中国科学院金属研究所团队开展相分离合金的凝固实验

依托载人空间站天和核心舱无容器材料实验柜，中国科学院金属研究所赵九洲研究团队，利用中国空间站上的静电悬浮无容器实验装置，开展了相分离合金的凝固实验，获得了预期的实验结果，证明空间站为材料凝固研究提供了理想的条件，数据库累计为该实验项目提供高质量数据产品约500GB，相关成果发表于国际期刊《National Science Review》。



国家基础学科公共科学数据中心
National Basic Science Data Center

材料科学基础数据库

材料科学基础数据库是由中国科学院金属研究所承建、中国科学院化学所和国家纳米科学中心参加建设的科学数据库。该库涵盖金属材料、有机高分子材料及纳米材料数据，数据资源量5.5GB。已建有的金属材料数据子库包含材料牌号1730多个，数据总量约12万条，涵盖金属材料成分、组织、工艺和性能等数据。其中有机高分子数据子库10045种材料（牌号）的6万余条数据，塑料材料7981种，橡胶材料660种，纤维材料102种，胶黏剂1738种，有机涂料1053种，助剂123种。纳米材料数据子库主要包括碳纳米管、石墨烯、磁性纳米颗粒等，涵盖光、热、电、声、磁、力、化学性质及毒性等基本特性数据。目前总数据量超过20万条，为基础研究人员、研发工程师和企业界之间建立联系。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	金属材料节点	11.7万条	2000余条
2	有机高分子材料节点	9.40万条	1500余条
3	纳米材料数据资源节点	1.3GB	0.12 GB
4	高分子材料形态结构	2.2GB	0.2GB

年度数据增量
500MB

支撑重要期刊论文
2篇

累计下载量
166.98GB

访问量
376万次

联系人
崔丽娜　materia@imr.ac.cn

承建单位
中国科学院金属研究所

参建单位
中国科学院化学研究所
国家纳米科学中心

服务成效

支撑论文
2篇

[1] 于英杰, 李辉, 张重远，多元谱线拟合-ICP-AES法测定钢中微量铌 [J]. 冶金分析, 2022,42(05):74
[2] 于英杰,李辉,张重远，基体匹配-ICP-AES法测定新型抗菌不锈钢中镓[J].化学分析计量，2022,31(03):27

支撑科技项目
2个

1 2016YFB0700500 材料基因工程专用数据库架构、标准规范与数据库 国家重点研发计划
2 2017YFB0702100 材料环境失效过程的高通量计算模拟与高通量实验技术 国家重点研发计划

专利
4项

1 一种实现EBSD系统跨尺度连续自动表征分析测试方法　ZL202110575323.3
2 一种用于测试非均质材料电阻率的样品组合物　ZL202220026155.2
3 一种高通量制备多元合金的激光熔覆装置　ZL202123082965.4
4 Method for Growing large-size crystal by laser assister heating and dedicated device　US113411B2

典型案例

为“祝融号”火星车用铝基复合材料研发提供数据支撑，助力祝融号成功“登火”

“祝融号”火星车车轮材料要求较高，需同时满足轻量化、耐磨、高塑性等指标。材料科学基础数据库团队建立了SiC/Al复合材料热变形的精确本构模型、塑性加工有限元模型和“金属基体+界面+增强相”三相复合材料3D逼真微观结构模型，助力铝基碳化硅复合材料的研发。所研发的材料具有高强、高塑、高稳定性等特点，可应用于火星车行走机构、驱动机构、探测器等50余种460余件零部件制造，让火星车能够在沙石遍地的荒漠地貌和零下90℃至30℃的巨大温差环境中顺利工作。



为超大型盾构机用直径8米轴承研制提供数据支撑

盾构机是国民经济建设的重大装备，承载着穿山越岭、过江跨海的重任。我国已实现了盾构机的国产化，但其核心部件主轴承却一直依赖进口，亟需打通自主可控制造的“最后一公里”。材料科学基础数据库的金属材料数据库为国产盾构机轴承研制团队提供重要轴承材料对选材、材料工艺优化、材料和轴承性能预测等数据，为我国成功研制首台（套）超大型盾构机用主轴承提供了有力数据保障。



干细胞与代谢疾病数据库

干细胞与代谢疾病数据库是以汇集干细胞基础研究数据、代谢疾病临床数据、细胞谱系资源和互联网公共数据为主要目标，基于高性能计算技术，建立的集数据汇集、存储管理、挖掘分析、可视化呈现为一体的创新应用平台。

该数据库以中国科学院广州生物医药与健康研究院、广州再生产医学与健康广东省实验室（生物岛实验室）、人类细胞谱系大科学设施为数据来源，同时汇聚高校、科研单位、医药企业多元异构数据，包括细胞生物学、免疫学、发育生物学、干细胞基础研究数据、代谢疾病临床数据和细胞谱系数据等。相关数据持续服务我国各级研究中心及实验室、科技计划项目和应用平台，为人类健康与“健康中国”战略实施提供数据支撑。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	GIBH论文关联数据	21148.85GB	14352.4GB
2	hECA	156GB	156GB
3	Spatial_Organ_Atlas	952GB	952GB
4	SCXA	1000GB	1000GB
5	singlecellportal	500GB	500GB
6	scRNAseqDB	11480GB	11480GB
7	SpatialDB	6GB	6GB
8	VascularSingleCell	2GB	2GB
9	iSyTE	5GB	5GB
10	CeDR Atlas	1GB	1GB
11	Gut Cell Atlas	30GB	30GB
12	stemformatics	1GB	1GB

联系人

赵娟娟　Zhao_juanjuan@gibh.ac.cn

承建单位

中国科学院广州生物医药与健康研究院

服务成效

在线子库
27个

资源总量
168TB

总记录数
362万条

支撑论文
23篇

研制分析挖掘工具
1个

融合汇集细胞图谱数据库资源
11个

科研项目数据汇交数据集
56个

新增数据量
30TB+

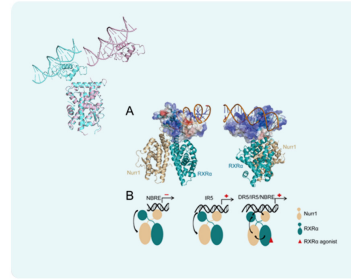
支撑国家重点研发计划
3个

1　2017YFA0105100　神经疾病大动物模型的建立及干细胞治疗评价　国家重点研发计划

2　2017YFA0102600　人多能干细胞分化过程中谱系命运决定的调控及异质性机制研究　国家重点研发计划

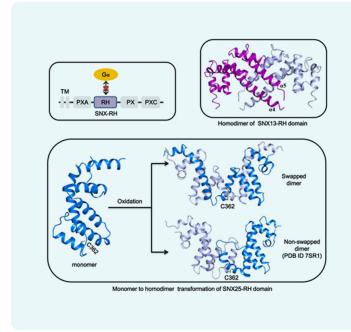
3　2017YFA0504100　细胞编程与重编程相关蛋白质机器研究　国家重点研发计划

典型案例



助力帕金森病靶点Nurr1以及Nurr1–RXRα别构激活的结构基础揭示

干细胞与代谢疾病数据库为该研究提供了数据及计算资源，利用CCP4软件包，完成的单体Nurr1三维晶体结构的解析，同时利用AMBER20软件，模拟分析了Nurr1和Nurr1–RXRα蛋白质复合物构象的动态变化，为其转录调控机制的提出，提供了重要的参考依据。研究团队解析了Nurr1多结构域结合DNA的结构，是核受体领域第一个单体多结构域晶体结构，并进一步利用整合结构生物学方法，研究了Nurr1受DNA和配体RXRα调控的分子机制。



助力新型非典型G蛋白信号转导调节因子同源结构域发现

干细胞与代谢疾病数据库为该研究提供了数据及计算资源，利用部署的软件包完成蛋白结构的解析，为SNX–RH亚家族蛋白的功能机制研究，提供了结构基础。研究团队对SNX–RH蛋白的RH结构域进行了结构生物学研究。课题组获得了SNX13、SNX14和SNX25的RH结构域的晶体，并成功解析了SNX13–RH和SNX25–RH结构域的晶体结构。

中国活体人脑成像共享数据库

中国活体人脑成像共享数据库(Chinese Data-sharing Warehouse for In-vivo Imaging Brain)是北京师范大学主导建立的包含广泛个体差异的人群的脑影像大数据库。该数据库依托于北京师范大学、中国科学院心理所等多个高校和科研院所，含中国儿童彩巢计划、中国重度抑郁症数据库、国际信度与可重复性(CORR)大数据共享、I See your brains (ISYB)等数据资源，以及中国科学院心理研究所独立自主研发的人脑数据处理软件CCS和DPABI，并经过该软件标准化流程处理的超过一万例的多模态人脑活体成像数据。国际人脑数据库自开放共享后，截至2022年底，国内外已有超过万余名科研工作者下载了预处理过后的活体脑影像数据，仅2022年下载量超过4万余次，成为国内外研究者最为常用的共享数据下载平台之一。

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	中国人脑连接组计划	3.69TB	3.69TB
2	2022年中国彩巢计划	900GB	50GB

在线子库
3个

支撑发表论文
3篇

资源量
4TB

年度下载量
1.5TB

联系人
左西年　xinian.zuo@bnu.edu.cn

承建单位
北京师范大学

参建单位
中国科学院心理研究所
北京大学
中国科学院生物物理研究所
西南大学

服务成效

支撑论文
3篇

[1] Bethlehem, R. A. I., Seidlitz, J., White, S. R., Vogel, J. W., Anderson, K. M., Adamson, C., Adler, S., Alexopoulos, G. S., Anagnostou, E., Areces-Gonzalez, A., Astle, D. E., Auyeung, B., Ayub, M., Bae, J., Ball, G., Baron-Cohen, S., Beare, R., Bedford, S. A., Benegal, V., . . . Alexander-Bloch, A. F. (2022). Brain charts for the human lifespan. Nature, 604(7906), 525-533. https://doi.org/10.1038/s41586-022-04554-y
[2] Gao, P., Dong, H.-M., Liu, S.-M., Fan, X.-R., Jiang, C., Wang, Y.-S., Margulies, D., Li, H.-F., & Zuo, X.-N. (2022). A Chinese multi-modal neuroimaging data release for increasing diversity of human brain mapping. Scientific Data, 9(1), 286. https://doi.org/10.1038/s41597-022-01413-3
[3] 张青,王银山,陈丽珍,张家鑫,周子轩,左西年.发展人口神经科学中的规范化建模:量化脑智发展规律与个体分化特征的“显微镜”[J].科学通报,2023,68(16):2086-2100.

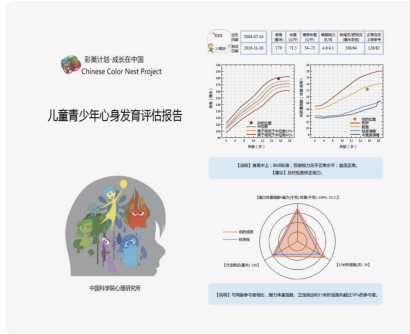
支撑科技项目
1个

20&ZD296　汉语自闭症儿童多模态交际障碍的多学科协同研究 国家社会科学基金重大项目

典型案例

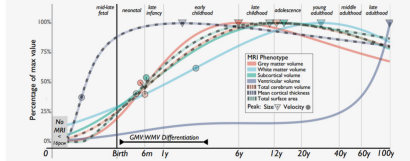
发布中国青年人多模态MRI数据，丰富全球脑影像数据共享多样性

中国活体人脑成像共享数据库团队开发了理想的面部匿名化工具，该工具能够最小化减少干扰，保证优质的脑组织提取和形态学测量，相关成果《A Chinese multi-modal neuroimaging data release for increasing diversity of human brain mapping》于2022年发表于Nature子刊《Scientific Data》(IF:8.501)上 (https://doi.org/10.1038/s41597-022-01413-3)。在此基础上，数据库通过Science Data Bank向公众发布了215个中国青年人多模态MRI数据“I See your brains (ISYB)”，为全球研究不同文化背景对人脑成像的潜在影响提供支持，丰富了全球脑影像数据共享的多样性(https://doi.org/10.11922/sciencedb.00740)。截至2022年底已有国内外超过四万余名科研工作者下载了预处理过后的活体脑影像数据。



参与人类生命周期脑图表研究，提高人脑图表样本多样性

人类生命周期脑图表研究汇集了来自全球200多个研究单位的10万余名志愿者，共12万余次磁共振脑影像数据，跨越了人类从受孕后115天到出生后100年的生命周期，建立首个人类大脑毕生发展常模，揭示了人脑毕生发展的里程碑。生长图表按性别进行分类，克服了不同研究中个体评估和技术手段的差异，可以稳定地检测神经与精神疾病的脑形态变化。中国活体人脑成像共享数据库为该项目提供了1000余例珍贵的非欧美国家样本数据，提高了人脑图表的样本多样性。这一成果于2022年4月6日发表于Nature(IF:43.07)(https://doi.org/10.1038/s41586-022-04554-y)。



国家基础学科公共科学数据中心
National Basic Science Data Center

地理空间数据库

地理空间数据库是先进的“一站式”地学数据综合服务平台，自主研发可靠的大规模地理空间数据实时分析引擎，实现亿级元数据高效管理和PB级实体数据分析服务。该数据库汇聚国内外开放地理空间数据资源，生产优质数据产品，创新模型在线、数据众包及在线FTP服务模式，提供数据和各类服务。2022年新增数据量10TB，总数据量达270TB，新增用户数16万人，总服务用户达到77万，极大地推动了我国数据开放共享，节约大量科研成本，产生显著社会效益。

联系人
林青慧　lqh@cnic.cn

承建单位
中国科学院计算机网络信息中心

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	LANDSAT_45TM_C2_L2 数据产品	10.5TB	3TB
2	Landsat 7 TM C2 L2	13.6TB	3TB
3	Landsat 7 TM C2 L2	8.8TB	4TB
4	GDEMV3 30M 分辨率数字高程数据	313.6GB	300GB

2022年度主要成果列表

序号	成果名称	年度增量
1	支撑CNKI学术期刊论文	1990篇
2	支撑CNKI学位论文	888篇
3	支撑SpringerLink 英文论文	221篇
4	支撑sciencedirect 英文论文	426篇
5	支撑Googlescholar英文论文	2030篇

服务成效

数据量
270TB

在线子库
108个

数据条目数
1000万条

服务用户
77万人

总访问人次
1.69亿次

独立IP数
161.71万+

总在线下载量
418.4TB

在线任务数
3万个

支持科研项目
1000个+

支持学术论文
5000篇+

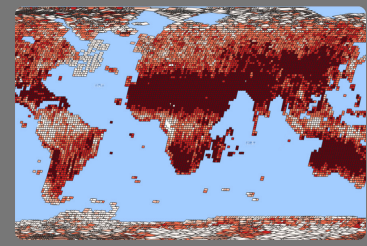
支撑科技项目
3个

1. XDA19020103 对地观测大数据云集成与共享 地球大数据先导专项自子课题
2. XDA19020203 格网数据引擎研制 地球大数据先导专项自子课题
3. 2019QZKK0308生态脆弱性与生态安全 第二次青藏高原综合考察研究专题

专利
1个

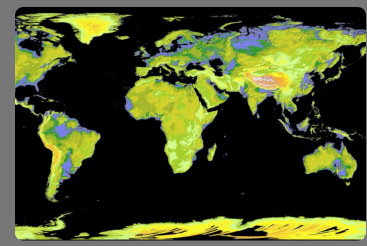
1. 一种基于多分辨率特征融合的遥感图像分类方法和装置CN202210006083.X

典型案例



为“风云02星质量检验评估分系统研制”项目提供全球遥感数据支持

为解决国内缺乏静止卫星定位导航精度评估工具的问题，“风云02星质量检验评估分系统研制”项目启动研发国内首个为风云四号静止气象卫星提供完整导航与配准精度评估功能的工具。在项目实施过程中，需要使用大量Landsat 8全通道，无云的图像作为地标点真值。项目在获取Landsat8数据的过程中遇到了极大的困难，相关研发受到阻碍。地理空间数据库基于数据处理技术与离线服务功能，为项目提供海量全球覆盖的Landsat 8数据资源，数据量超过6TB，对项目实施起到了至关重要的推动作用。



为民航运行安全提供DEM数据支撑

民航的运行安全越来越多地依赖数字地形和障碍物数据的精确性和完整性，湖南航空股份有限公司在研究高原航线运行过程中，面临全球数字高程数据匮乏的难点。地理空间数据库依托长期的技术积累和丰富的数据资源，为湖南航空股份有限公司提供全球DEM数据支撑，解决了相关数据匮乏、可靠性不足、精度不够等问题，对该公司高原航线运行研究、地表形态信息提取及障碍物分析等，提供了十分重要的数据支撑，节约了大量的时间和人力成本，有利保障高原航线研究工作的顺利开展，为民航运行安全做出贡献。

国家基础学科公共科学数据中心
National Basic Science Data Center

干旱区药用资源数据库

干旱区药用资源数据库及实物库是涵盖新疆、中亚等干旱、半干旱地区的药用资源综合数据库，包含药用资源信息、天然产物/标准提取物信息、药材标本、基原植物标本、标准品等实物信息。该数据库实现干旱区药用数据资源规范化线上发布，提供药材基原、有效成分化合物和标准提取物等数据查询和下载服务，成为中亚传统医用药材资源重要数据共享服务平台。

在线子库
4个

资源量
2.5GB

2022年度新增数据资源清单

序号	数据库/集名称	数据总量	年度增量
1	药用物种资源	2180条	500条
2	药材资源	1410条	360条
3	化合物资源	6011条	2610条
4	组分实物资源	2500条	400条

联系人
刘戈宇　liugy@ms.xib.ac.cn

承建单位
中国科学院新疆理化技术研究所

服务成效

支撑科技项目
1个

2022YFE0202900口服抗新冠病毒新药VV116在“一带一路”国家的临床研究　国家重点研发计划重点专项课题

支撑论文
8篇

[1] Isolation, structure elucidation, and biological activity of polysaccharides from Saussurea involucrata, International Journal of Biological Macromolecules 222 (2022) (154–166)
[2] Guangrong Liu,Mayila Kamilijiang,Aytursun Abuduwaili,Deng Zang,Nuermaimaiti Abudukelimu, Geyu Liu, Abulimiti Yili, Haji Akber Aisa,Non-volatile compounds of Hyssopus cuspidatus Boriss and their antioxidant and antimicrobial activities, Food Chemistry, 2022, 374, 131638 Karimu Aihaiti, Jun Li, Saimijiang Yaermaimaiti, Liu Liu, Xuelei Xin, Haji Akber Aisa*
[3] Discovery of a novel highly potent and low-toxic jatrophane derivative enhancing the P-glycoprotein-mediated doxorubicin sensitivity of MCF-7/ADR cells, European Journal of Medicinal Chemistry, 2022, 244, 114822Ayitila Maimaitijiang, Bianlin Wang, Hequn Yang, Dan Tang, Yongqiang Liu, Haji Akber Aisa*
[4] Isolation, structural modification, characterization, and bioactivity of polysaccharides from Folium Isatidis, Industrial Crops & Products, 2022, 176, 114319, Aytursun Abuduwaili, Rehebati Nuerxiati, Paiheerding Mutailifu, Yanhua Gao, Chunfang Lu, Abulimiti Yili*
[5] Diversity and Biological Activities of Endophytic Fungi from the Flowers of the Medicinal Plant Vernonia anthelmintica, International Journal of Molecular Sciences, 2022, 23, 11935 Litao Niu, Nigora Rustamova, Huxia Ning, Paiziliya Paerhati, Chunfang and Abulimiti Yili*
[6] Study of Novel Furocoumarin Derivatives on Anti-Vitiligo Activity, Molecular Docking and Mechanism of Action, International Journal of Molecular Sciences, 2022, 23, 7959, Chao Niu, Deng Zang and Haji Akber Aisa*
[7] α Hydroxylation of α,α-Disubstituted N-tert-Butanesulfinyl Ketimines with Molecular Oxygen: Stereoselective Synthesis of α Tertiary Hydroxyimines, Organic Letters, 202224, 746–751, Nuermaimaiti Yisimayili, Hui Liu, Yun Yao, and Chong-Dao Lu*
[8] Shabyar Ameliorates High Glucose Induced Retinal Pigment Epithelium Injury Through Suppressing Aldose Reductase and AMPK/mTOR/ULK1 Autophagy Pathway, Frontiers in Pharmacology, 2022, 13, 852945, Xiao Yan Liu, Jun Peng,Fei He, Xirali Tursun, Shu Ping Li, Xue Lei Xin* and Haji Akber Aisa*

支撑专利
15项

1 一种天山雪莲多糖提取物的制备方法与应用 202210676641.3
2 一种天山雪莲多糖的制备方法及其用途 202210676705.X
3 一种从刺糖中海藻糖的制备技术及其应用 202210648896.9
4 一种乙酰糖苷类橙酮及抗肿瘤用途 202210597495.5
5 从鹰嘴豆芽中分离纯化紫檀烷类化合物高丽槐素和美迪紫檀素的方法 202210202755.4
6 一种从野生条斑钳蝎中分离纯化透明质酸酶的方法及应用 202210203340.9
7 一种三环吡咯并[2,3-d]嘧啶酮类衍生物及用途 202210171535.X
8 一种吡唑并[3,4-d]吡咯(啶)并[1,2-a]嘧啶酮亚芳基类衍生物及其用途 202210149626.3
9 一种驱虫斑鸠菊地上部位组织中内生真菌的分离纯化方法及其用途 202210048437.7
10 一种含杂环橙酮类衍生物及抗菌用途 202210045243.1
11 一种电化学法由甲醇制备甲缩醛的方法 202210042151.8
12 一种从蟾蜍毒素中制备蟾蜍二烯内酯的方法及其用途 202210009684.6
13 202210009684.6
14 一种罗布麻花有效部位的制备方法及用途 202210009006.X
15 14.蒙古蒿中的倍半萜类化合物及制备方法和用途 202210009546.8
15.伊犁绢蒿中倍半萜类化合物的制备方法和用途 202210010136.5

典型案例

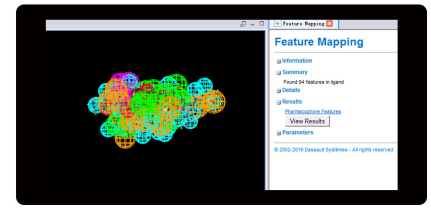
推动中药民族药研究，入选国家首批“科创中国”创新基地

为响应“健康中国”和“一带一路”国家战略需求，干旱区药用资源数据库以新疆中药民族药为抓手，通过整合规范化数据资源，分离鉴定化合物2000余个，发现了源于民族药的3个先导化合物及20余种新骨架类型和600余个新化合物。形成基础研究、成果转化、国际科技合作的研发创新链，目前已研发上市1个有效部位新药和2个复方新药。2022年，团队入选国家首批“科创中国”创新基地，极大推动新疆医药产业高质量发展，加快中药民族药融入中亚国家医药体系进程。



支持建成抗白癜风化合物数据集

干旱区药用资源数据库提供30余种新骨架化合物和600余个新化合物，为抗白癜风等疾病的先导化合物和多个候选药物研发，提供了数据支撑。助力7个活性部位和1个单体的作用机制研究，建立500多个小分子抗白癜风化合物库，发现抗白癜风先导化合物7个，并建立JAK/STAT信号通路抑制剂的活性化合物筛选平台。研制的19个国家标准样品已投入应用，包括2款中药新药转让企业进入 II 期临床、1款中药新药获批临床、2款新药在乌兹别克斯坦获批生产、3款新药获批临床等。



附录·年度大事记

1月

召开“数字经济时代的科学数据理论”专题研讨会

国家基础数据中心联合国际科学理事会数据委员会（CODATA）中国全国委员会组织召开“数字经济时代的科学数据伦理”专题研讨会。

2月

开展“数据使用情况及数据价值评估”问卷调查

国家基础数据中心面向科研数据工作者开展“数据使用情况及数据价值评估”问卷调查，启动可计量的开放科学数据集影响力评价体系研究。

3月

中国科学院基础学科公共科学数据中心正式启动运行

国家基础数据中心承建的院级学科数据中心通过中国科学院网络安全和信息化领导小组办公室认定，中国科学院基础学科公共科学数据中心正式启动运行。

3月

授牌成立分中心

国家基础数据中心召开年度工作会议，正式授牌成立交通运输分中心和信息学科分中心。

3月

顺利召开“开放科学与科学数据”专题研讨会

国家基础数据中心联合国际科学理事会数据委员会（CODATA）中国全国委员会组织召开“开放科学与科学数据”专题研讨会。

4月

成功获批中国科协科技期刊支撑服务类项目

国家基础数据中心获批中国科协科技期刊支撑服务类项目：“科学数据开放共享标准研究”（编号2022KJQK005-1）。

6月

顺利召开2021年度内部考核会议

国家基础数据中心2021年度内部考核会议顺利举行，经过专家评审，形成了数据库后续支持方案。

6月

国家标准《数据论文出版元数据》专家研讨会顺利召开

国家基础数据中心牵头的国家标准《数据论文出版元数据》专家研讨会顺利召开，该项标准对于规范科学数据出版、推动科学数据共享生态建设具有重要意义。

7月

数据库负责人左西年应邀出席Nature数据期刊编委

国家基础数据中心“中国活体人脑影像共享数据库”负责人左西年应邀出席Nature数据期刊《Scientific Data》编委。

7月-8月

发布2022年度数据应用创新课题征选指南

国家基础数据中心发布2022年度数据应用创新课题征选指南，面向全国科研人员完成“新能源汽车”、“纳米科技”2个领域方向的应用创新课题遴选。

11月

《科学数据安全标准体系》等15项团体标准正式发布

国家基础数据中心牵头的国家重点研发计划“科学数据安全技术及基础技术标准研究”发布15项团体标准作为该项目研究成果发布。

12

《中国科学数据资源发展研究报告2022》正式出版

国家基础数据中心依托牵头的国家基础科技条件平台中心课题“国内外科学数据资源发展动态跟踪与科学数据发展报告编写”（项目编号2022WT06），联合多家单位共同出版《中国科学数据资源发展研究报告2022》。

2022年度代表性论文

（科学数据管理与技术）

[1] Dharejo F A, Zawish M, Zhou Y, et al. Fuzzyact: A fuzzy-based framework for temporal activity recognition in iot applications using rnn and 3d-dwt[J]. IEEE Transactions on Fuzzy Systems, 2022, 30(11): 4578-4592.

[2] Du Y, Guo Y. Machine learning techniques and research framework in foodborne disease surveillance system[J]. Food Control, 2022, 131: 108448.

[3] Wang Z, Zhao J, Zhang R, et al. UATNet: U-shape attention-based transformer net for meteorological satellite cloud recognition[J]. Remote Sensing, 2022, 14(1): 104.

[4] Qiao Z, Fu Y, Wang P, et al. RPT: toward transferable model on heterogeneous researcher data via pre-training[J]. IEEE Transactions on Big Data, 2022, 9(1): 186-199.

[5] Wang P, Liu K, Zhou Y, et al. Unifying human mobility forecasting and trajectory semantics augmentation via hawkes process based lstm[C]//Proceedings of the 2022 SIAM International Conference on Data Mining (SDM). Society for Industrial and Applied Mathematics, 2022: 711-719.

[6] Deeba F, Dharejo F A, Zawish M, et al. A novel image dehazing framework for robust vision-based intelligent systems[J]. International Journal of Intelligent Systems, 2022, 37(12): 10495-10513.

[7] Li C, Zhou Y, Zheng X, et al. Tracing the footsteps of open research data in China[J]. Learned Publishing, 2022, 35(1): 46-55.

[8] 童昭,王露笛,朱小杰,杜一.基于预训练模型的军事领域命名实体识别研究[J].数据与计算发展前沿,'2022,4(05):120-128.

[9] 胡良霖,朱艳华,李坤,胡泊,王璐,高瑜蔚,李国庆.科学数据伦理关键问题研究[J].中国科技资源导刊,2022,54(01):11-20.

[10]Shi W, Fan G, Shen Z, et al. gcCov: Linked open data for global coronavirus studies[J]. Mlife, 2022, 1(1): 92-95.

[11] Song C, Wang L, Du Y, et al. A Knowledge Graph Based Approach to Operational Coordination Recognition in Wargame[C]//Methods and Applications for Modeling and Simulation of Complex Systems: 21st Asia Simulation Conference, AsiaSim 2022, Changsha, China, December 9–11, 2022, Proceedings, Part I. Singapore: Springer Nature Singapore, 2022: 510-521.

2022年度数据标引代表性论文

（作者单位为国家基础数据中心）

[1] 黄海宁,黄乾生.细菌胞外囊泡的研究进展[J].微生物学报,2022,62(05):1613-1628.DOI:10.13343/j.cnki.wsxb.20210582.

[2] 袁媛,陈雷.IDADP-葡萄病害识别研究图像数据集[J].中国科学数据(中英文网络版),2022,7(01):86-90.

[3] 李英勇.中药材化学成分的晶体结构数据集[J].中国科学数据(中英文网络版),2022,7(01):74-85.

[4] 赵玉娟,康飞,王海娟等. 废旧锂离子电池界面解离技术研究进展[C]//中国环境科学学会环境工程分会.中国环境科学学会2022年科学技术年会——环境工程技术创新与应用分会场论文集(三).《工业建筑》杂志社有限公司,2022:541-550.-DOI:10.26914/c.cnkihy.2022.027874.

[5] 李婷,童锋贤,郑为民,张娟.GEO卫星的VLBI快速高精度定位观测[J].天文研究与技术,2022,19(04):305-316.-DOI:10.14005/j.cnki.issn1672-7673.20210916.003.

[6] 王寅初,李文军,任承钢,刘正一,秦松.海洋生物产品服务平台的构建与应用[J].中国科技资源导刊,2022,54(01):40-46+75.

[7] 李文杰,王江宁,卜翠萍,葛斯琴,林聪田,韩艳,纪力强.基于动物志的粉蝶形态特征数据集[J].中国科学数据(中英文网络版),2022,7(01):91-98.

[8] 陈中,郑为民,陈肖,薛岩松.2007—2020中国探月工程VLBI测量数据集[J].中国科学数据(中英文网络版),2022,7(01):45-56.

[9] 肖翠,金冬梅,李颖超,李晓京,张路杨,郑柏岩,吉小冬,林秦文.云南漾濞石门关景区动植物资源数据集[J].中国科学数据(中英文网络版),2022,7(01):168-178.

[10] 朱晓娜,童力,郑为民,张娟.一种用于空间VLBI的射电源条纹搜索算法[J].天文研究与技术,2022,19(06):568-575.-DOI:10.14005/j.cnki.issn1672-7673.20220718.001.

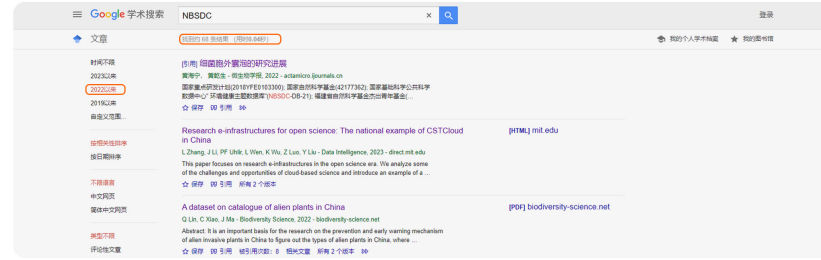
[11] 任庆敏,林燕妮,王寅初,秦松.不同铁浓度对一种颗石藻Emiliania huxleyi的生长及光合作用的影响[J].海洋科学,2022,46(12):50-62.

[12] 赵旭,郑为民,童力,马茂莉.VLBI探测器数传信号仿真技术研究[J].天文学进展,2022,40(03):441-452.

[13] 陈奥,童力,郑为民.用于VLBI测轨的新型Gold码调制PN-DOR信标的设计与仿真[J].天文学进展,2022,40(02):271-283.

Google Scholar标引论文

国家基础数据中心2022年度支撑发表学术论文超过5000篇。其中通过Google Scholar精确查询发现在2022年1月1日至2022年12月31日期间，已经落实明确标引（作者单位为国家基础数据中心或致谢国家基础数据中心基金支持）的论文数量为68篇，标引方式得到了大家的支持和认可。



典型数据库/集

序号	主题库名称	数据库/集名称	数据总量
1	交通运输分中心	足尺环道数据库	600TB
2	交通运输分中心	桥隧服役性能数据库	81GB
3	交通运输分中心	出行云数据库	12TB
4	交通运输分中心	高寒高海拔公路数据库	1GB
5	交通运输分中心	交通安全驾驶模拟数据库	30GB
6	交通运输分中心	城市综合客运协同服务评估计算结果数据库	20.5GB
7	交通运输分中心	潮流泥沙数据库	12.7GB
8	交通运输分中心	地下水数据库	7GB
9	交通运输分中心	水质监测数据库	1.6GB
10	交通运输分中心	结构监测数据库	4.8GB
11	世界山茶属植物品种数据库	“2022年新登录茶花品种”数据集	26,088个
12	世界山茶属植物品种数据库	新增山茶花照片	61,818张
13	图像理解机器学习数据库	水稻病虫害图像数据集	217GB
14	图像理解机器学习数据库	小麦病虫害图像数据集	101GB
15	图像理解机器学习数据库	玉米病虫害图像数据集	80GB
16	图像理解机器学习数据库	棉花病虫害图像数据集	40GB
17	图像理解机器学习数据库	大豆病虫害图像数据集	32GB
18	图像理解机器学习数据库	黄瓜病虫害图像数据集	59GB
19	图像理解机器学习数据库	葡萄病虫害图像数据集	64GB
20	图像理解机器学习数据库	豇豆病虫害图像数据集	9GB
21	图像理解机器学习数据库	番茄病虫害图像数据集	52GB
22	图像理解机器学习数据库	西瓜病虫害图像数据集	6GB
23	图像理解机器学习数据库	梨树病虫害图像数据集	43GB
24	图像理解机器学习数据库	大蒜病虫害图像数据集	9GB
25	图像理解机器学习数据库	马铃薯病虫害图像数据集	19GB
26	图像理解机器学习数据库	甜瓜病虫害图像数据集	5GB
27	图像理解机器学习数据库	林木病虫害图像数据集	2GB
28	图像理解机器学习数据库	鼠害病虫害图像数据集	3GB
29	图像理解机器学习数据库	蚕豆病虫害图像数据集	9GB
30	图像理解机器学习数据库	水稻图像理解数据集	217GB
31	图像理解机器学习数据库	番茄图像理解数据集	51GB
32	图像理解机器学习数据库	黄瓜图像理解数据集	80GB
33	图像理解机器学习数据库	玉米图像理解数据集	120GB
34	环境健康主题数据库	生物体污染物数据库	15万条
35	环境健康主题数据库	抗性基因数据库	40万条
36	中国海岸带植物资源数据库	海岸带陆生高等植物分类信息与图片数据集	5GB
37	中国海岸带植物资源数据库	海岸带底栖大型藻类图片与代表物种基因组数据集	580GB
38	中国海岸带植物资源数据库	海岸带底栖微藻分类信息与图片数据集	15GB
39	中国海岸带植物资源数据库	藻类培养基成分查询系统	329KB
40	中国海岸带植物资源数据库	中国海岸带硅藻分布查询系统	3.22MB

序号	主题库名称	数据库/集名称	数据总量
41	医学物理数据库	光子肿瘤治疗数据集	205GB
42	医学物理数据库	重离子肿瘤治疗数据集	237GB
43	医学物理数据库	同位素药物数据集	548GB
44	医学物理数据库	医疗辐射防护数据集	352GB
45	中国动物主题数据库	中国动物志数据库	8.84GB
46	中国动物主题数据库	中国动物图谱数据库	0.632GB
47	中国动物主题数据库	中国经济动物数据库	1.315GB
48	中国动物主题数据库	中国鸟类数据库	60GB
49	中国动物主题数据库	中国哺乳动物数据库	0.86GB
50	中国动物主题数据库	中国蝴蝶数据库	2.76GB
51	中国动物主题数据库	中国蜜蜂数据库	0.1GB
52	中国动物主题数据库	中国内陆水体鱼类数据库	1GB
53	中国动物主题数据库	中国两栖动物	0.47GB
54	中国动物主题数据库	中国爬行动物数据库	0.3GB
55	中国动物主题数据库	中国直翅目与革翅目昆虫数据库	398.77MB
56	中国动物主题数据库	中国蜚蠊目数据库	1MB
57	中国动物主题数据库	中国双尾纲与原尾纲数据库	53MB
58	中国动物主题数据库	中国甲虫名录数据库	2.26MB
59	中国动物主题数据库	中国动物物种编目数据库	111.54MB
60	中国动物主题数据库	中国灵长类物种文献及使用数据库	3325 MB
61	中国动物主题数据库	中国有毒动物数据库	200MB
62	中国动物主题数据库	西南县级脊椎动物分布名录	25MB
63	中国动物主题数据库	西南保护区脊椎动物分布名录	15MB
64	中国动物主题数据库	云南鸟类数据库	500MB
65	中国动物主题数据库	云南蝴蝶分布名录	15MB
66	中国动物主题数据库	云南森林昆虫分布名录	20MB
67	中国动物主题数据库	中国脊椎动物分类代码数据库	10MB
68	中国动物主题数据库	中国昆虫模式标本数据库	1.8GB
69	中国动物主题数据库	中国昆虫新种数据库	8MB
70	中国动物主题数据库	地方动物志数据库	1.4GB
71	中国动物主题数据库	实验动物数据库	100MB
72	中国动物主题数据库	动物图像标注数据库	1583GB
73	中国动物主题数据库	中国脊椎动物红色名录	30MB
74	中国动物主题数据库	昆虫模式标本数据库v2	500MB
75	中国动物主题数据库	云南两栖类数据库	230MB
76	中国动物主题数据库	西南地区保护、濒危、特有种编目数据库	12MB
77	中国动物主题数据库	生物评估与保护	50MB
78	中国动物主题数据库	动物研究专家数据库	1MB
79	中国动物主题数据库	鸟类声音语谱图数据库	9GB
80	中国动物主题数据库	濒危动物遗传资源数据库	12GB

序号	主题库名称	数据库/集名称	数据总量
81	中国动物主题数据库	中国动物名录数据库	0.25GB
82	中国动物主题数据库	中国动物志出版信息数据库	10MB
83	中国植物主题数据库	中国植物物种名录（2022版）	117330条
84	中国植物主题数据库	中国植物物种名录（2021版）	115788条
85	中国植物主题数据库	植物文献电子书库	103011本
86	中国植物主题数据库	雄安新区植物名录数据库	641条
87	中国植物主题数据库	国家重点保护野生植物2021	495条
88	中国植物主题数据库	京津冀常见野生植物数据库	7126条
89	中国植物主题数据库	中国芳香植物数据库	1377条
90	中国植物主题数据库	食用植物资源数据库	1133条
91	中国植物主题数据库	药用植物资源数据库	1954条
92	中国植物主题数据库	木材植物资源数据库	2534条
93	中国植物主题数据库	中国油脂植物数据库	974条
94	中国植物主题数据库	栲胶植物资源数据库	62条
95	中国植物主题数据库	中国有毒植物物种库	1137条
96	中国植物主题数据库	外来归化及入侵植物数据库	737条
97	中国植物主题数据库	省级植物名录	61876条
98	中国植物主题数据库	图解植物学词典数据库	3981条
99	中国植物主题数据库	中国植物名称数据库	262920条
100	中国植物主题数据库	中国植物图片数据库	5009624条
101	中国植物主题数据库	中国植物文献库	3652312条
102	中国植物主题数据库	中国水生植物数据库	739条
103	中国植物主题数据库	城市绿化植物数据库	208条
104	中国植物主题数据库	四季时令性花卉数据库	509条
105	中国植物主题数据库	植物与昆虫相互关系数据库	12185条
106	中国植物主题数据库	濒危保护物种名录共享专区	10847条
107	中国植物主题数据库	保护区物种名录库	17737条
108	中国植物主题数据库	中国珍稀濒危植物名录	1929条
109	中国植物主题数据库	国家重点保护野生植物名录（第一批和第二批）	495条
110	中国植物主题数据库	民族药用植物、综合系统	6225条
111	中国植物主题数据库	西南野生生物种质资源库	73607条
112	中国植物主题数据库	植物园：综合、名录、定植	85415条
113	中国植物主题数据库	古植物：名录、化石	8072条
114	中国植物主题数据库	标本库（昆明植物所）	622436条
115	中国植物主题数据库	植物物候照片专题库	5230条
116	中国植物DNA条形码数据库	中国植物DNA条形码数据集	26.1万条
117	中国植物DNA条形码数据库	竹类植物资源数据库	10.1GB
118	中国植物DNA条形码数据库	板蓝基因组数据平台	203.3GB
119	中国植物DNA条形码数据库	蓖麻属植物组学数据库（RcDB）	13.5GB
120	中国植物DNA条形码数据库	杜鹃花属组学数据库RPGD	1.2GB

序号	主题库名称	数据库/集名称	数据总量
121	化学主题数据库	化合物基本信息库	7,290MB
122	化学主题数据库	物性及热化学数据库	168MB
123	化学主题数据库	纯化合物相变数据库	18MB
124	化学主题数据库	化合物结构数据库	4,166MB
125	化学主题数据库	化学危险品安全与处置数据库	154MB
126	化学主题数据库	非电解质体系汽液相平衡数据库	92MB
127	化学主题数据库	参考文献数据库	36MB
128	化学主题数据库	离子簇热化学数据库	44.5MB
129	化学主题数据库	化合物估算性质数据	50MB
130	化学主题数据库	化学核心期刊数据库	6,000MB
131	化学主题数据库	中国化学文献数据库	1,800MB
132	化学主题数据库	中国化学专利数据库	6,390MB
133	化学主题数据库	质谱谱图数据库	150MB
134	化学主题数据库	化合物毒性数据库	213.9MB
135	化学主题数据库	MSDS数据库	6MB
136	化学主题数据库	药物与天然产物数据库	623.1MB
137	化学主题数据库	化学试剂数据库	1,400MB
138	化学主题数据库	精细化工产品数据库	16.6MB
139	化学主题数据库	化合物活性数据库(ciac)	1,130MB
140	化学主题数据库	C-13核磁数据库	2,330MB
141	化学主题数据库	化合物登录数据库	8,313MB
142	化学主题数据库	化学反应数据库	1,970MB
143	化学主题数据库	红外谱图数据库	3,780MB
144	化学主题数据库	化工产品数据库	28.6MB
145	化学主题数据库	中药与化学成分数据库	647.5MB
146	化学主题数据库	专业情报数据库	3.4MB
147	化学主题数据库	量化计算结果数据库	4,600MB
148	化学主题数据库	稀土物性数据库	140MB
149	化学主题数据库	稀土萃取数据库	140MB
150	化学主题数据库	化学物质分析方法	10.15MB
151	化学主题数据库	药品名称数据库	4.67MB
152	化学主题数据库	化学配方数据库	46.85MB
153	化学主题数据库	工程塑料数据库	6.39MB
154	化学主题数据库	化学品纯化数据库	20MB
155	化学主题数据库	化学物质毒性中文文献数据库	5.6MB
156	化学主题数据库	毒性化合物数据库	300MB
157	化学主题数据库	化学品安全特性数据库	5MB
158	化学主题数据库	离子簇反应热化学数据库	0.98MB
159	化学主题数据库	植物化学成分数据库	5,735MB
160	化学主题数据库	中草药单一化学成分质谱数据库	630MB

序号	主题库名称	数据库/集名称	数据总量
161	化学主题数据库	化合物生物活性数据库(sioc)	25MB
162	化学主题数据库	中药方剂药物代谢组学数据库	270MB
163	化学主题数据库	工业污染防控数据库	40MB
164	化学主题数据库	金属离子形态数据集	0.1MB
165	原子分子数据库	原子光谱数据库	191万
166	原子分子数据库	原子能级数据库	9万
167	原子分子数据库	原子电离势数据库	0.2万
168	原子分子数据库	电子原子碰撞激发数据库	36万
169	原子分子数据库	电子原子碰撞电离数据库	1.4万
170	原子分子数据库	离子分子碰撞数据库	8.2万
171	原子分子数据库	电子分子碰撞数据	1.4万
172	光学技术数据库	光学玻璃库	15GB
173	光学技术数据库	光学晶体库	5GB
174	光学技术数据库	光学塑料库	7GB
175	光学技术数据库	光源库	2GB
176	光学技术数据库	光电探测器	5GB
177	光学技术数据库	光学传感器	25GB
178	光学技术数据库	光学镜头库	11GB
179	光学技术数据库	光学棱镜库	18GB
180	光学技术数据库	光学标准库	3GB
181	光学技术数据库	镀膜材料及膜系数据库	10GB
182	光学技术数据库	光学成像库	381GB
183	光学技术数据库	光学检测库	90GB
184	光学技术数据库	特色光谱信息	160GB
185	光学技术数据库	激光器件库	16GB
186	光学技术数据库	激光传输特性库	15GB
187	光学技术数据库	激光光学系统库	23GB
188	光学技术数据库	光纤光学库	21GB
189	光学技术数据库	微光与红外器件库	30GB
190	光学技术数据库	红外传输特性库	8GB
191	光学技术数据库	微光与红外光学系统库	11GB
192	光学技术数据库	短波光学库	2GB
193	核物理主题数据库	评价核数据库	1113.5G
194	核物理主题数据库	原子核特性数据库	3MB
195	核物理主题数据库	实验核数据库	992MB
196	核物理主题数据库	常用核衰变数据库	24MB
197	核物理主题数据库	核天体数据库	24MB
198	核物理主题数据库	钍铀循环专用核数据库	800MB
199	核物理主题数据库	医用同位素数据库	7MB
200	核物理主题数据库	中国评价核数据库	2GB

序号	主题库名称	数据库/集名称	数据总量
201	核物理主题数据库	光核反应数据库	504MB
202	核物理主题数据库	裂变产额数据库	90MB
203	核物理主题数据库	中子活化反应截面数据库	400MB
204	宇宙学数值模拟数据库	SDSS,WISE,ROSAT星系，星系团的星表库观测数据	10GB
205	宇宙学数值模拟数据库	NIHAO流体数值模拟	16032GB
206	宇宙学数值模拟数据库	弱引力透镜全天球光线追踪数值模拟	28012GB
207	宇宙学数值模拟数据库	ELUCID宇宙学数值模拟	64043GB
208	宇宙学数值模拟数据库	FAST观测图像	3079GB
209	宇宙学数值模拟数据库	illustris TNG	64078GB
210	宇宙学数值模拟数据库	CSST光路模拟	21GB
211	宇宙学数值模拟数据库	中国空间站工程巡天望远镜CSST高精度宇宙学数值模拟	977033GB
212	宇宙学数值模拟数据库	中国空间站工程巡天望远镜CSST的观测数据仿真	193016GB
213	宇宙学数值模拟数据库	银河系星流数值模拟数据集	20GB
214	宇宙学数值模拟数据库	孪生宇宙模拟	6300GB
215	宇宙学数值模拟数据库	中国空间站工程巡天望远镜CSST的观测数据仿真7平方度	103010GB
216	VLBI月球与深空探测数据库	嫦娥一号VLBI测量数据集	1.8TB
217	VLBI月球与深空探测数据库	嫦娥二号VLBI测量数据集	2.5TB
218	VLBI月球与深空探测数据库	嫦娥三号VLBI测量数据集	6.8TB
219	VLBI月球与深空探测数据库	嫦娥五号再入返回试验VLBI测量数据集	5.6TB
220	VLBI月球与深空探测数据库	嫦娥四号VLBI测量数据集	1.96TB
221	VLBI月球与深空探测数据库	嫦娥五号VLBI测量数据集	2.58TB
222	VLBI月球与深空探测数据库	天问一号VLBI测量数据集	3.3TB
223	载人航天空间科学与应用数据库	天宫一号对地观测领域数据库	35TB
224	载人航天空间科学与应用数据库	天宫一号空间科学领域数据库	20GB
225	载人航天空间科学与应用数据库	天宫二号对地观测领域数据库	60TB
226	载人航天空间科学与应用数据库	天宫二号宽波段成像仪青藏高原专题数据集	12.2GB
227	载人航天空间科学与应用数据库	“一带一路”沿线港口城市天宫二号多光谱影像数据集	28GB
228	载人航天空间科学与应用数据库	天宫二号遥感图像自然景物分类数据集	8.7GB
229	载人航天空间科学与应用数据库	天宫二号多模态遥感场景分类数据集（MRSSC1.0）	214MB
230	载人航天空间科学与应用数据库	天宫二号多模态遥感场景分类数据集（MRSSC2.0）	211MB
231	载人航天空间科学与应用数据库	天宫二号空间科学领域数据库	30GB
232	载人航天空间科学与应用数据库	空间站空间科学与应用领域数据库	30TB
233	材料科学基础数据库	金属材料节点	11.7万条
234	材料科学基础数据库	有机高分子材料节点	9.40万条
235	材料科学基础数据库	纳米材料数据资源节点	1.3GB
236	材料科学基础数据库	高分子材料形态结构	2.2GB
237	干细胞与代谢疾病数据库	健康科学数据库	20TB
238	干细胞与代谢疾病数据库	人类致癌基因数据库	400GB
239	干细胞与代谢疾病数据库	人类致癌转移数据库	10GB
240	干细胞与代谢疾病数据库	自身抗体生物标志物数据库	30MB

序号	主题库名称	数据库/集名称	数据总量
241	干细胞与代谢疾病数据库	单细胞肿瘤数据资源库	4.8GB
242	干细胞与代谢疾病数据库	小鼠细胞图谱	60TB
243	干细胞与代谢疾病数据库	人类细胞图谱	50TB
244	干细胞与代谢疾病数据库	细胞多组学数据库	1TB
245	干细胞与代谢疾病数据库	国际开放数据库	50GB
246	干细胞与代谢疾病数据库	allen_cell数据库	1876GB
247	干细胞与代谢疾病数据库	人类蛋白质数据库	2GB
248	干细胞与代谢疾病数据库	人类细胞蓝图（HCL）数据库	702968条
249	干细胞与代谢疾病数据库	SC2disease	946481条
250	干细胞与代谢疾病数据库	细胞标志物（Marker）数据库	13605条
251	干细胞与代谢疾病数据库	Tabula Muris	529823条
252	干细胞与代谢疾病数据库	GIBH论文关联数据	21148.85GB
253	干细胞与代谢疾病数据库	hECA	156GB
254	干细胞与代谢疾病数据库	Spatial_Organ_Atlas	952GB
255	干细胞与代谢疾病数据库	SCXA	1000GB
256	干细胞与代谢疾病数据库	singlecellportal	500GB
257	干细胞与代谢疾病数据库	scRNAseqDB	11480GB
258	干细胞与代谢疾病数据库	SpatialDB	6GB
259	干细胞与代谢疾病数据库	VascularSingleCell	2GB
260	干细胞与代谢疾病数据库	iSyTE	5GB
261	干细胞与代谢疾病数据库	CeDR Atlas	1GB
262	干细胞与代谢疾病数据库	Gut Cell Atlas	30GB
263	干细胞与代谢疾病数据库	stemformatics	1GB
264	中国活体人脑成像共享数据库	中国彩巢计划	900GB
265	中国活体人脑成像共享数据库	I See your brains (ISYB)	61.98G
266	中国活体人脑成像共享数据库	国际信度与可重复性(CORR)大数据共享	1.5T
267	地理空间数据库	LANDSAT系列数据	60.91TB
268	地理空间数据库	MODIS陆地标准产品	39.51TB
269	地理空间数据库	MODIS中国合成产品	4.86TB
270	地理空间数据库	MODIS-L1B标准产品	48.12TB
271	地理空间数据库	DEM数字高程数据	1.40TB
272	地理空间数据库	EO-1系列数据	3.24TB
273	地理空间数据库	NOAA AVHRR数据产品	1.56TB
274	地理空间数据库	大气污染插值数据	1.37TB
275	地理空间数据库	Sentinel数据	13.7TB
276	地理空间数据库	高分一号WFV数据产品	2.8TB
277	干旱区药用资源数据库	药用资源信息平台及标本库	2GB
278	干旱区药用资源数据库	药材标准信息平台及标本库	2GB
279	干旱区药用资源数据库	标准提取物数据库	2GB
280	干旱区药用资源数据库	天然产物资源库	2GB

序号	主题库名称	数据库/集名称	数据总量
281	中国病毒资源基础数据库	中国病毒性病原调查专业库	870GB
282	中国病毒资源基础数据库	病毒性病原调查基础数据库	0.1GB
283	中国病毒资源基础数据库	病毒性病原历史疫情数据库	0.1GB
284	中国病毒资源基础数据库	高通量病原检测数据库	0.1GB
285	中国病毒资源基础数据库	病毒性病原遗传资源数据库	0.1GB
286	中国病毒资源基础数据库	病毒性病原本底数据库	0.1GB
287	中国病毒资源基础数据库	病毒资源数据库	85.3GB
288	中国病毒资源基础数据库	病毒编目数据库	0.124GB
289	中国病毒资源基础数据库	病毒敏感细胞数据库	0.02GB
290	中国病毒资源基础数据库	野生动物病原宿主组织数据库	0.11GB
291	中国病毒资源基础数据库	动物疫病病原综合应用数据库	1.39GB
292	中国病毒资源基础数据库	病毒遗传资源数据库	23865GB
293	中国病毒资源基础数据库	自然疫原性病毒综合应用数据库	3.32GB
294	中国病毒资源基础数据库	昆虫病毒综合应用数据库	0.1GB
295	中国病毒资源基础数据库	流感综合应用数据库	5621GB
296	中国病毒资源基础数据库	肝炎综合应用数据库	0.02GB
297	中国病毒资源基础数据库	2020年武汉地区新冠疑似病例样本信息基础数据集	0.026GB
298	中国病毒资源基础数据库	2020年鄂州新冠样本采集数据集	0.001GB
299	中国病毒资源基础数据库	2020年鄂州新冠样本检测数据集	0.001GB
300	中国病毒资源基础数据库	新冠防控操作标准规范数据集	0.001GB
301	中国病毒资源基础数据库	2020年内蒙地区重要病毒媒介昆虫宏基因组数据集	4100GB
302	中国病毒资源基础数据库	2021年新疆地区蜱传病毒宏基因组数据集	100GB
303	中国病毒资源基础数据库	2021年内蒙地区蜱传病毒宏基因组数据集	100GB
304	中国病毒资源基础数据库	高原冻土宏基因组数据集	100GB
305	中国病毒资源基础数据库	高原冰川宏基因组数据集	50GB
306	中国病毒资源基础数据库	猴粪便样本数据集	20条
307	中国病毒资源基础数据库	极端环境宏基因组数据集	40.1GB
308	中国病毒资源基础数据库	2022年湖北地区蜱传病毒宏基因组数据集	100GB
309	中国病毒资源基础数据库	2022年海南地区蜱传病毒宏基因组数据集	100GB

配套设施

国家基础数据中心依托中国科技云开展基础设施建设和数据服务。中国科技云是基于云计算的国家科研信息化基础设施，持续开放汇聚国内外优秀信息化基础设施和服务。基础科学数据中心通过与其在技术创新和集成创新等方面的合作，推动国家基础学科数据云建设，实现基础学科各类资源融合、智能运营和动态调度，提供安全、按需、智能化的云服务。

- 计算资源 315PFlops
- 存储资源 150PB
- 综合平台 81个
- 科研软件 1000个+
- 典型云服务 9类
- 免费CPU计算机时 840万+
- GPU计算机时 160万+
- 云服务器 3000台+
- 云备份资源 10PB

服务

网络传输
云计算
云存储
通用性大数据处理环境
高性能计算网格
人工智能计算
数据服务

资源

关联搜索
共享文献
知识产权
领域云数据库

超融合通讯服务

电子邮件
即时通讯
在线会议

地域分布



资源安全保障

国家基础数据中心参照相关国家标准对平台的基础设施与网络环境、技术实现、数据保护、管理制度体系等方面进行整体设计。

执行国家标准：

《GB/T 35274–2017 信息安全技术 大数据服务安全能力要求》

《GB/T 22239–2019信息安全技术 网络安全等级保护基本要求》



建设数据全程可控的可信数据存储库

建立科学数据应急管理灾备和长期保存机制

积极开展科学数据安全框架和区块链技术研究