国家基础学科公共科学数据中心

National Basic Science Data Center

中国科学院计算机网络信息中心
Computer Network Information Center,CAS

邮箱 E-mail nbsdc@cnic.cn

联系 Tel 010-58812526

来访 Add 北京市海淀区中关村南四街4号中国科学院软件园
4,4th South Street,Zhongguancun,Haidian-District, Beijing, China

邮编 P.C 100190

2020

年度报告

www.nsdata.cn

中心概述

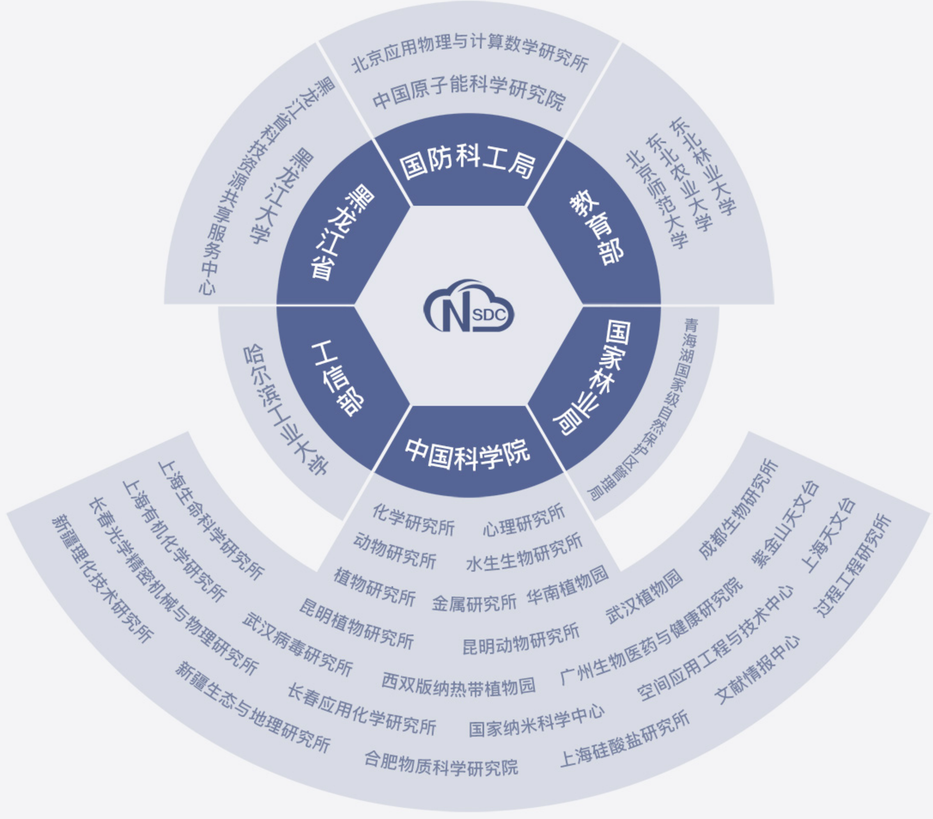


国家基础学科公共科学数据中心（简称“国家基础数据中心”）是中国科学院计算机网络信息中心联合中国科学院、教育部、工信部、国防科工局、国家林业局、黑龙江省等下属研究所共37个单位共同建设的国家科技资源共享服务平台，旨在汇集管理物理、化学、材料、动物、植物、病毒和信息科学等基础学科领域，青海湖、黑龙江、新疆等典型区域长期科研活动积累的科学数据，以及上述领域政府预算资金支持项目生产的基础科学数据。经过多年的持续发展，国家基础数据中心已经形成了完善的基础学科数据资源体系，以及分布式科学数据资源统一管理、集成融合、分析挖掘和应用服务的技术体系、标准体系和服务体系，有力支持了数据驱动的基础学科原始创新、多学科交叉融合创新、国民经济发展、人民生命健康和国际交流与合作。

组织架构



参建单位



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

数据资源

17类
一级学科覆盖

19个
主题库

290个
数据集

1.37PB
数据总量

8.9亿条+
领域数据记录

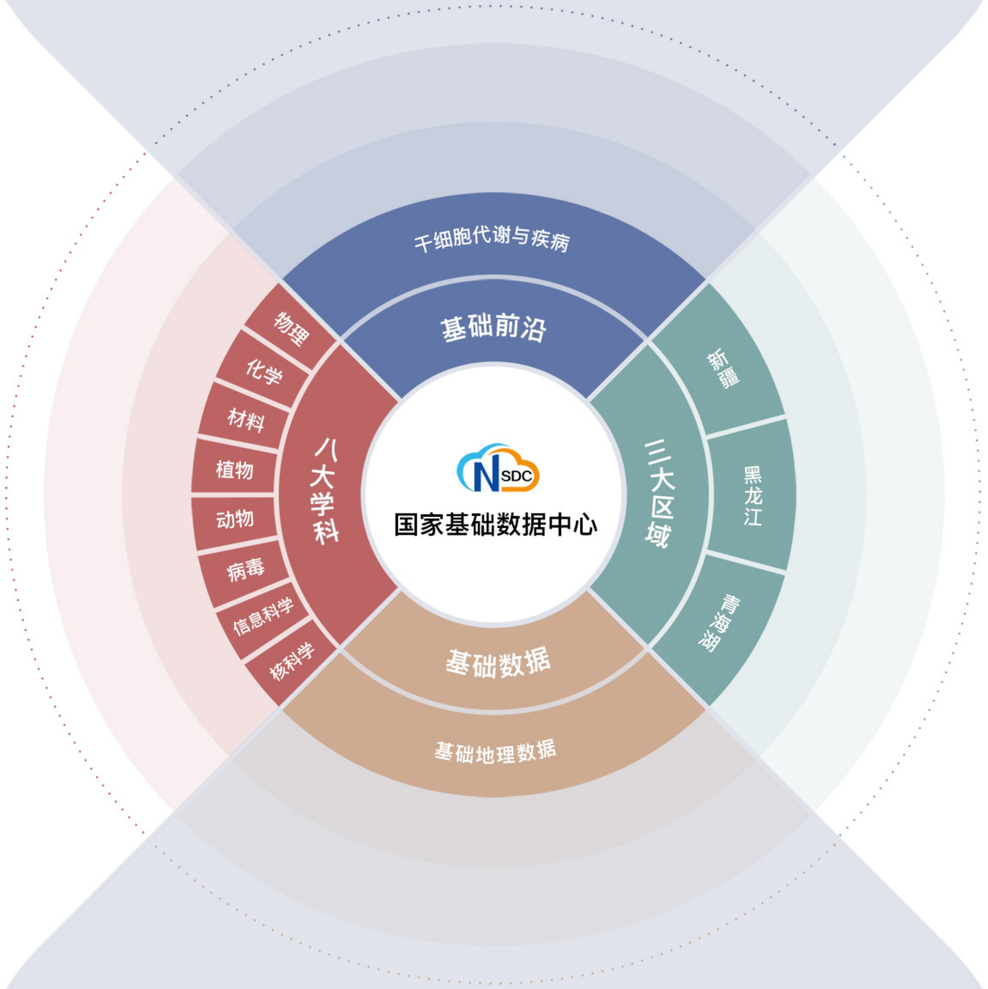
国家基础数据中心系统整合了中国科学院、教育部、工信部、农业部、国防科工局、国家林业局等研究机构在物理、化学、材料、动物、植物和病毒等学科领域，以及青海湖、黑龙江和中亚地区等典型区域长期科研活动中积累的基础科学数据。目前，中心数据资源覆盖17类一级学科，包括19个主题库，290个数据集。数据总量1.37PB，自建数据资源总量712.63TB，国家重点研发计划项目汇交科学数据量671.47TB，数据记录8.9亿余条，并在完善的技术体系、标准体系支撑下构建了基于云门户的“数据服务–信息服务–知识服务”多元化服务体系，有效促进了基础科学数据的资源汇聚、开放共享、多学科交叉融合分析和应用。

科学数据来源

- 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

领域覆盖

动物、植物、化学等领域，数据总量覆盖领域80%以上
金属材料、核物理等领域，数据总量覆盖领域60%以上



基于上述数据资源建设运行的国家基础学科数据云已经成为国内规模最大的基础学科科学数据综合服务平台

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

重大项目科学数据汇交

汇交机制及流程

根据国家《科学数据管理办法》（国办发〔2018〕17号）、《国家重点研发计划项目综合绩效评价工作规范（试行）》（国科办资〔2018〕107号）等文件精神，“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，应由项目牵头单位汇交到相关科学数据中心，各级计划（专项、基金等）项目管理部门应建立先汇交科学数据、再验收科技计划（专项、基金等）项目的机制”。国家基础数据中心积极开展国家重点研发计划项目科学数据汇交工作，编写了《国家重点研发计划项目科学数据汇交工作指南》，组织和参加了22次科学数据汇交交流会。



项目数据汇交情况

截止2020年12月，国家基础数据中心主要服务支撑“科学技术部高技术研究发展中心”、“工业和信息化部产业发展促进中心”及“中国21世纪议程管理中心”三个项目管理机构的国家重点研发项目科学数据汇交工作，共计18个专项、98个项目（完成汇交工作的项目28个），汇聚科学数据总量达671.47TB。

中国21世纪议程管理中心（5）

绿色建筑及建筑工业化	2 项
大气污染成因与控制技术研究	1 项
公共安全风险防控与应急技术装备	2 项

科学技术部高技术研究发展中心（65）

纳米科技	1 项
高性能计算	8 项
新能源汽车	6 项
先进轨道交通	8 项
云计算和大数据	5 项
量子调控与量子信息	1 项
战略性先进电子材料	12 项
增材制造与激光制造	2 项
重大科学仪器设备开发	1 项
重点基础材料技术提升与产业化	11 项
煤炭清洁高效利用和新型节能技术	10 项

工业和信息化部产业发展促进中心（28）

网络空间安全	9 项
智能电网技术与装备	17 项
大科学装置前沿研究	1 项
制造基础技术与关键部件	1 项

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

科学数据存储库
Science Data Bank

2300位+
服务科研人员

2TB+
数据体量

53万次+
访问量

21万次+
下载量

提升科研数据价值
助力学术成果传播

科学数据存储库(ScienceDB)面向学术期刊、科研人员提供数据在线存储、汇交管理、长期保存与获取、共享、出版和引用服务。

2020年9月，ScienceDB成为国内首个被Scientific Data和Springer Nature推荐的通用型数据存储库，是继Dryad、figshare、Harvard Dataverse、OSF、Zenodo、Mendeley Data之后的第7家被收录的通用型存储库。同时，ScienceDB已获得FAIRsharing、re3data、OpenDOAR等认证收录。

2020年12月，《中国科学院科学数据工作要点》（科办〔2020〕7号）正式发布，明确将ScienceDB确定为院论文关联数据首选汇交平台，推动落实中国科学院科技期刊论文关联数据汇交。

www.scidb.cn
sciencedb@cnic.cn

SPRINGER NATURE
scientific data
AGU ADVANCING EARTH AND SPACE SCIENCE
CSData 中国科学数据
FAIRsharing.org standards, databases, policies
re3data.org REGISTRY OF RESEARCH DATA REPOSITORIES
OPENDOAR

ScienceDB 科学数据存储库

集中汇聚·友好访问

免费提供通用型的数据出版服务
支持各种文件类型数据的共享和获取
基于中国科技云，以云托管的方式提供服务

保障权益·提升发现

提供CC0和CC BY 4.0等数据使用许可协议
提供DOI、CSTR、PID的自动注册和管理
提供数据集更新版本、引用格式推荐等服务

超强存储·安全可靠

所有数据文件配套防篡改机制
具备强大底层数据存储、数据处理与数据灾备能力
保障稳定数据在线出版服务、7×24小时无单点故障

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

科学数据库

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、病毒、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19个主题数据库290个数据集，数据资源总量712.63TB，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。

中国动物主题数据库

中国病毒资源基础数据库

中国植物主题数据库

植物园主题数据库

中国植物DNA条形码数据库

化学主题数据库

原子分子物理主题数据库

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

射电天文与深空探测数据库

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

地理空间数据云

青海湖区域综合研究专题数据库

干旱区药用资源数据库

黑龙江省特色基础科技数据库

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学科领域数据为主体，依据国内外相关动物学研究成果进行建设，是目前国内动物学领域类群覆盖全面、内容科学的权威数据库。该数据库涵盖中国鸟类、兽类、两栖爬行动物、鱼类、昆虫及其它无脊椎动物。截至2020年，总计在线子库31个，记录数约800万条，线上及线下年均下载量超过1TB，累计支撑超过20项科学研究项目及公众科普活动，对动物学研究、生物保护决策和科学普及具有重要意义。

林聪田
linct@ioz.ac.cn
中国科学院动物研究所



31个
在线子库

800万条
总记录数

1TB+
年均下载量

20项+
支撑科研项目

主要数据资源

序号	数据库/集名称	数据量
1	中国动物志数据库	36.4GB
2	中国经济动物数据库	1.32GB
3	中国鸟类数据库	29.06GB
4	中国兽类数据库	0.86GB
5	中国蝴蝶数据库	2.76GB
6	中国内陆水体鱼类数据库	1.05GB
7	中国两栖爬行动物数据库	0.774GB
8	中国直翅目与革翅目昆虫数据库	0.398GB
9	中国灵长类物种文献及使用数据库	3.325GB
10	动物图像标注数据库	2GB

服务成效

序号	成果名称	类型	发布时间
1	Huang X , Lin C , Ji L . The persistent multi-dimensional biases of biodiversity digital accessible knowledge of birds in China[J]. Biodiversity and Conservation, 2020, 29(11):3287-3311.	论文	2020
2	Wang H Q , Liu X W . A new genus and a new species of Anabropsinae (Orthoptera: Anostostomatidae) from China[J]. Zootaxa, 2020, 4809(3):509-520.	论文	2020
3	Zhu-Qing H E , Wang H Q , Liu Y F , et al. New and little known Agraeciini species from Hainan, China (Orthoptera: Ensifera: Tettigoniidae: Conocephalinae: Agraeciini)[J]. Zootaxa, 2020, 4748(1):153-162.	论文	2020
4	Jin X , Liu X , Wang H . New taxa of the tribe Meconematini from South-Pacific and Indo-Malayan Regions (Orthoptera, Tettigoniidae, Meconematinae)[J]. Zootaxa, 2020, 4772(1):1-53.	论文	2020
5	《中国生物物种名录》第二卷-动物-脊椎动物(IV) 两栖纲（江建平等）	专著	2020
6	中国动物分类代码 第1部分:脊椎动物	国家标准	2020
7	蔡波,姚明灿,郭宪光.河南发现刘氏石龙子及该物种系统位置分析[J].动物学杂志,2019,54(05):668-676.	论文	2019
8	Chung H V , Tao N T , Vinh L Q. A new species of Leptobrachella Smith 1925 (Anura: Megophryidae) from Thanh Hoa Province, Vietnam[J] The Raffles Bulletin of Zoology, 2019, 67:536-556.	论文	2019
9	贾佳，张美华，高克勤，江建平 中国四川省峨眉山特有小鲵类（两栖纲，有尾目） 龙洞山溪鲵（Batrachuperus londongensis Liu and Tian, 1978）正模标本骨骼μCT 三维数据集. 中国科学数据, 2018, 3(2). DOI: 10.11922/csdata.2018.0005.zh	数据论文	2019

典型案例

为“腾讯企鹅爱地球”项目提供动物物种名录

腾讯企鹅爱地球项目是腾讯联合权威环保公益组织WWF等发起的公益项目，致力于探索数字时代下的网络生态安全和环境保护。2020年11月中国动物主题数据库为该项目提供覆盖鸟类、兽类、两栖类、爬行类等的中国动物物种名录数据，数据内容包括物种学名、中文名及别名。以权威科学的物种名录作为索引，保证物种图像分类准确性、物种知识的科学性，为用户提供科学规范的生物学知识，在不同来源的物种数据整合和生物保护方面发挥积极作用。



持续为《中国生物物种名录》提供数据支撑

中国生物物种名录是国内广泛使用的标准名录，被全球物种2000收录，具有重要影响力。《中国生物物种名录2020》已经于2020年5月正式出版，该名录共新增了15971个物种，其中动物物种增加了9350个，中国动物主题数据库持续为《中国生物物种名录》的维护更新提供蝴蝶、蜘蛛、广翅目昆虫、尺蛾科昆虫、鞘翅目昆虫、直翅目昆虫、两栖动物等多个类群名录的新增与更新数据支持。



中国病毒资源基础数据库

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是基于我国主要病毒资源的背景信息、遗传信息、流行特征以及交流等重要科学信息建立的综合性病毒资源信息共享数据库。该库采集和整合了重要病毒相关宏基因组测序原始数据，拼接数据，序列基因预测结果基因注释，物种的丰度等数据，已经覆盖8个目、84个科、278个属、1703个种(分别占全球已发现病毒物种的15%、50%、20%、26%)。

主要数据资源

序号	数据库/集名称	数据量
01	病毒资源数据库	85.3GB
02	病毒编目数据库	0.124GB
03	病毒敏感细胞库	0.02GB
04	病毒遗传资源数据库	23,865GB
05	野生动物病原宿主组织数据库	0.11GB
06	动物疫病病原综合应用数据库	1.39GB
07	自然疫原性病毒综合应用数据库	3.32GB
08	昆虫病毒综合应用数据库	0.1GB
09	流感综合应用数据库	5,621GB
10	肝炎综合应用数据库	0.02GB
11	中国病毒流行病学调查专业数据库	870GB

范兆军
roy@wh.iov.cn
中国科学院武汉病毒研究所

服务成效

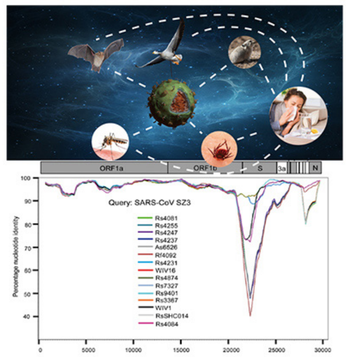
在2020年度新冠疫情期间，本数据库高质高效地配合国家和武汉市政府完成了大量疑似样本接收、转运，检测、筛查和资源标准化入库及分发工作。

302次	115个	50,000个+
在线数据服务申请	在线数据预订	在线样本

典型案例

病毒病原调查数据管理与服务支持

中国病毒基础数据库为“我国重要自然宿主及媒介昆虫的病毒病原调查”提供宏基因组检测数据。通过大数据分析，从大量的检测样本中抓取了病原的本底数据，利用参比数据设计了PCR引物并探测出相关宿主中的病毒病原背景，分离出160余种病毒株。该工作为新发传染病早期诊断和疫苗、药物研制提供支持，为政府决策提供重要科学依据。



为抗疫新冠的应急及常态化疫情防控工作提供支撑

在新冠疫情期间，中国病毒基础数据库中的病毒敏感细胞数据库为新冠病毒专业鉴定、增殖细胞模型建立和药物筛选提供了敏感细胞数据和病毒培养及质控信息，为新冠药物筛选和实验提供了重要支持。同时，病毒遗传资源数据库为新冠前期检测提供了重要的参比数据信息，相关重要源头资源信息协助完成了新冠病毒的前期检测。



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

中国植物主题数据库是以Species 2000中国节点和中国植物志名录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、化石植物名录和标本以及药用植物数据，建成25个数据子库及5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。

肖翠
xiaocui@ibcas.ac.cn
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

主要数据资源

序号	数据库/集名称	数据量
01	植物彩色照片库	120GB
02	植物标本库	260MB
03	文献志书库	35MB
04	京津冀常见野生植物数据库	330KB
05	芳香植物资源数据库	86KB
06	食用植物资源数据库	1.56TB
07	药用植物资源数据库	161KB
08	木材植物资源数据库	150KB
09	油脂植物资源数据库	30KB
10	栲胶植物资源数据库	15KB
11	有毒植物数据库	400KB
12	外来入侵植物数据库	64KB



25个
数据子库

5个
特色软件工具

服务成效

经过多年的发展，目前中国植物主题数据库已拥有包括京津冀常见野生植物数据库、中国芳香植物数据库、植物与昆虫相互关系数据库等在内的25个特色权威数据子库，通过自建和共建的方式，构建了植物彩色照片数据库、植物文献志数据库、古植物化石名录数据库。该数据库为用户提供基础数据，支持众多科研成果产出，服务于国家重大科学工程等项目，对植物学领域的人工智能发展起到积极作用。

典型案例

“花伴侣”植物图像识别应用

中国植物主题数据库的彩色照片库是“花伴侣”植物图像识别应用的重要识别训练素材库。目前，“花伴侣”经过五年的发展，截止2020年底，其安卓手机App用户已超过1000万，较2019年增加了400万；已实现了对2万余种植物的识别覆盖，对常见植物的识别率达90%以上，在国内外同类应用中处于领先地位。花伴侣App可以支持用户获取小范围内物种识别记录的实时动态，还可以辅助新物种（包括物种的新分布地）的发现和记录，极大方便了非植物分类学专业人士，推动了公众科普。



植物名录批量自动校对系统

中国植物主题数据库基于全球植物分子系统学研究的最新成果开发了植物名录批量自动校对系统。其中，石松类和蕨类以PPGI系统（2016）为标准；裸子植物以Christenhusz系统（2011）为标准；被子植物以APGIV系统（2016）为标准。植物名录批量自动校对系统操作简单，界面友好，数据权威。用户无需注册，可批量粘贴植物中文名或拉丁名，一键实现名称校对和系统排序等功能，并根据用户粘贴的植物名称自动生成标准的植物名录。该系统可实现校对植物规范名称、实现不同类植物根据名称排序等，目前被林业院校师生、保护区工作人员、林业科研工作者等广泛使用。



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

植物园主题数据库

植物园主题数据库是国内统一的植物园数据资源综合展示和共享平台。该库根据公众对植物园基本信息和基础数据的需求，将中国科学院核心植物园的数据按照标准进行数据重整、补全，并通过网络整合集成为引种、保育、图谱、生境视频、GIS、基本属性、其它等7大类分类明确又相互关联的虚拟数据集。数据库通过统一门户7X24小时稳定地向公众提供权威、全面的一站式植物园基础信息关联检索和共享服务。

服务成效

703,338个
累计独立访问IP数

2,016,866次
累计访问人次

23,302,948次
累计页面访问数

56,405,846个
累计文件数

1163.39GB
累计下载量

主要数据资源

序号	数据库/集名称	数据量
01	植物引种与保育数据库	63GB
02	热带亚热带植物基础数据库	329.13GB
03	中国植物图谱数据库	172.09GB

杨中清
yang@wbgcas.cn
中国科学院武汉植物园

参建单位
中国科学院西双版纳热带植物园
中国科学院华南植物园

典型案例

服务植物园迁地栽培植物志编撰

科技基础性工作专项“植物园迁地栽培植物志编撰”通过引用植物园主题数据库子库“迁地保护植物数据库”的引种和保育植物名录，减轻工作量的同时提高了数据的准确性。研究人员仅需在经植物分类学专家审核校验的植物名录数据基础上补充植物物候、病虫害、生长记录、凭证标本等数据，便可以形成完整的植物园迁地栽培植物志。相关数据为项目后续植物多样性保护和可持续利用分析奠定了坚实的基础。



服务西南–滇南地区本土植物清查与保护

植物园主题数据库为“中国植物园联盟建设本土植物全覆盖保护计划”提供西双版纳热带植物园历年植物引种、清查、物候、生长量、标本等数据，支撑了西南–滇南地区本土植物清查与保护项目计划制定、项目执行。截至2020年年底，等级2–4的西双版纳受威胁植物共244种已就地保护、近地保护和迁地保护，占总数的69%。普洱等259种受威胁植物目前已就地保护、近地保护和迁地保护，占总数的66%。



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

王亚楠
wangyanan@mail.kib.ac.cn
中国科学院昆明植物研究所

中国植物DNA条形码数据库是我国植物DNA条形码科学数据的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17.3万条，覆盖中国植物2578属11842种。平台提供植物DNA条形码检索和在线比对服务，基于DNA条形码推动新一代智慧植物志(iFlora)建设，为政府和相关产业的植物物种精准鉴定提供序列数据服务。

17.3万条
植物DNA条码数据

11842种
覆盖中国植物

iFlora
支撑建设智慧植物志

主要数据资源

序号	数据库/集名称	数据量
01	植物DNA条形码数据库	17.3万条

服务成效

138篇
支撑论文发表

93篇
SCI收录

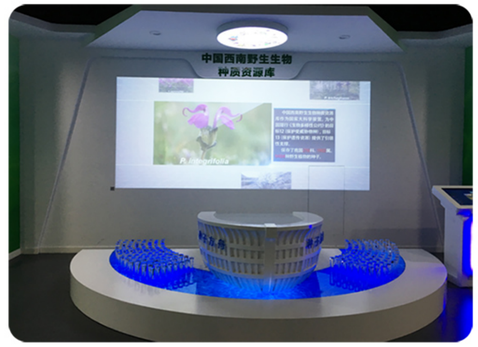
30人+
支撑培养专业技术人员

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，成果丰硕。共支撑发表论文138篇（总影响因子210.5），其中SCI收录93篇，其中8篇代表性SCI论文被引409次，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

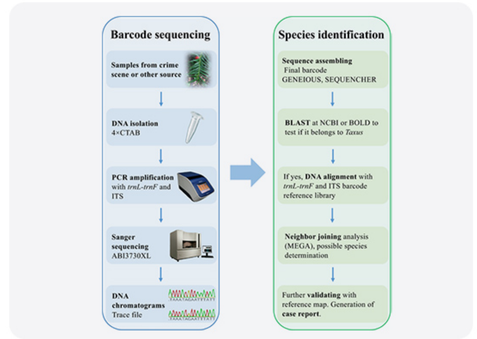
参加“率先行动砥砺奋进—‘十八大’以来中国科学院创新成果展”

中国植物DNA条形码数据库支撑的“种子方舟——中国西南野生生物种质资源库”成果参与了“率先行动砥砺奋进—‘十八大’以来中国科学院创新成果展”，该成果以基于DNA条形码的植物快速鉴定装置作为重要主题，系统展示了遗传信息高效获取技术装备体系和智能化“百科全书”式植物志书的前景，凸显了科学数据工作对科技创新的重要支撑作用。



支持“全球红豆杉属植物制定分子和地理身份证研究”

物种的准确鉴定是生物多样性保护的基石。“全球红豆杉属植物制定分子和地理身份证研究”综合运用DNA条形码、物种分布区模拟和谱系地理学等多学科交叉方法，提出了全球红豆杉属植物DNA条形码标准数据全库(comprehensive reference library)，并将其成功的运用到法医鉴定中，构建了标准的物种鉴定体系。中国植物DNA条形码数据库支撑的该研究为珍稀濒危物种找到了一个快速、准确的鉴定方法，对系列珍稀濒危生物多样性保护策略的制定、非法贸易的遏制起到了示范作用。



化学主题数据库

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

服务成效

2项
支撑项目获奖

39篇
支撑论文发表

149个
支撑科研项目

184,547个
注册用户

839,608人次
访问人次

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，当前登录化合物860万余种，并以此为基础实现了基于化合物唯一标识的异地异构化学数据资源整合与共享；面向全球提供化合物全景式的化学科学数据检索展示、化学数据加工及面向项目的专题数据服务。

赵月红
yhzhao@ipe.ac.cn
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

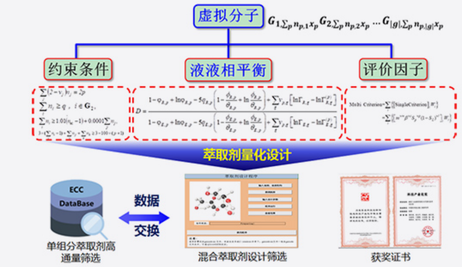
主要数据资源

序号	数据库/集名称	数据量
01	化合物基本信息库	7,290 MB
02	物性及热化学数据库	168 MB
03	纯化合物相变数据库	18 MB
04	化合物结构数据库	4,166 MB
05	天然气水合物数据库	5 MB
06	化学危险品安全与处置数据库	154 MB
07	铁铜液体相互作用参数数据库	31 MB
08	非电解质体系汽液相平衡数据库	92 MB
09	共混聚合物相容性数据库	6 MB
10	聚合物溶液相平衡数据库	40 MB
11	参考文献数据库	36MB
12	离子簇热化学数据库	44.5 MB
13	化合物估算性质数据	50MB
14	化学核心期刊数据库	6,170 MB
15	中国化学文献数据库	1,960 MB
16	中国化学专利数据库	6,390 MB
17	质谱谱图数据库	160.2 MB
18	化合物毒性数据库	213.9 MB
19	MSDS数据库	6 MB
20	药物天然产物数据库	623.1 MB
21	化学试剂数据库	1,610 MB
22	精细化工产品数据库	16.6 MB
23	化合物活性数据库（HIV）	1,060 MB
24	C-13核磁数据库	2,330 MB
25	化合物基础（登录）数据库	8,175MB
26	化学反应数据库	1,970MB
27	化合物生物活性数据库（SIOC）	962.6MB
28	红外谱图数据库	3,780MB
29	化工产品数据库	28.6 MB
30	中药与化学成分数据库	647.5 MB
31	专业情报数据库	3.4 MB
32	量化计算结果数据库	4.6GB
33	稀土物性数据库	140MB
34	稀土萃取数据库	140MB
35	化学物质分析方法	10.15MB
36	药品名称数据库	4.67MB
37	化学配方数据库	46.85MB
38	工程塑料数据库	6.39MB
39	辛烷值数据库	4MB
40	化学物质毒性中文文献数据库	5.6MB
41	化学物质毒性数据库	300MB
42	化学品安全特性数据库	5MB
43	离子簇反应热化学数据库	0.98MB
44	植物化学成分数据库	5,835MB
45	中草药单一化学成分质谱数据库	440MB
46	化合物生物活性数据库	20M
47	中药方剂药物代谢组学数据库	50MB
48	化合物基本信息库	7,290 MB
49	含能材料数据库	10MB



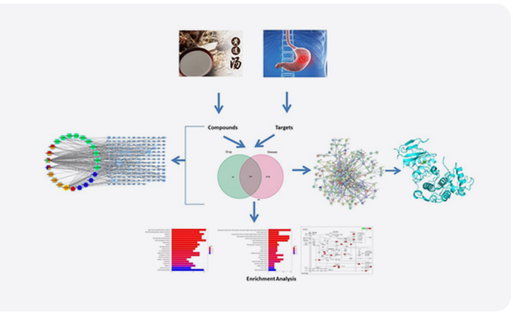
典型案例

煤化工废水新型萃取剂筛选与设计



2020年以化学主题数据库数据资源为基础，结合长期形成的数据管理、分析及应用经验和技术，开展钢铁工业园区水网络全局优化研究。在国内典型钢铁园区供水种类、关键用水点用排水操作参数以及水处理单元操作参数等调研数据的基础上，形成专题数据集，分析园区用排水特征和水网络全局优化方案，为园区绿色升级改造提供科技支撑。研究成果为2020年“中国科技产业化促进会科学技术奖（特等奖，2020-CY-001-T）”获奖项目提供了数据和理论支撑。

为中药现代化研究提供支撑服务



顺应生命科学和生物医药产业的发展趋势，助力中药这一具有中国特色的优势领域和产业走向现代化和国际化，化学主题数据库建设整合了植物-化合物-基因-疾病数据，并实现了与中药、毒性和谱学数据的互连互通。基于此，科研人员采用网络药理学和基因组学等方法研究了多种中药的有效成分与治病机理，2020年，该数据库共支撑了全国各地16家医院和医学院所的中药临床药理学研究工作，包括1个国家重点研发计划、13项国家自然科学基金项目和10项省级科研项目，相关工作在国内外药学期刊共发表论文13篇。

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建立，是目前国内原子分子领域唯一的数据库。原子分子数据在天体物理、武器物理、聚变能源（ICF、Tokomak）和工农业生产等领域都有重要应用。该数据库现有原子光谱、能级、电离势、电子与原子分子碰撞截面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。

主要数据资源

序号	数据库/集名称	数据量
01	原子光谱数据库	182万条
02	原子能级数据库	9万条
03	原子电离势数据库	0.2万条
04	电子原子碰撞激发数据库	36万条
05	电子原子碰撞电离数据库	1.4万条
06	离子分子碰撞数据库	8.2万条
07	电子分子碰撞数据	1.4万条

刘春雷
liu_chunlei@iapcm.ac.cn
北京应用物理与计算数学研究所

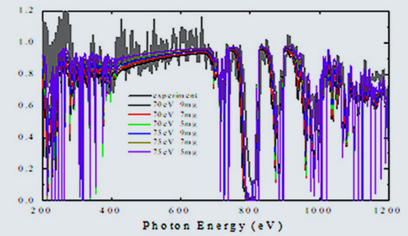
服务成效

原子分子数据在各领域应用广泛，尤其是天体物理等基础科学研究领域、聚变科学等高科技研发领域、照明科技和EUV光源等相关国民经济技术研发领域、以及战略武器研发等相关国防建设领域。原子分子数据库近年支撑发表了10余篇高质量学术论文，发表于The Astrophysical Journal Supplement Series (IF:12.203)、Atomic Data and Nuclear Data Tables (IF:2.165)等期刊。支撑了国家重大专项“X基本参数研究”、科技部大科学装置前沿研究(2017YFA0403200）等重大科研项目的申请和研究，为项目提供了所需的原子光谱和能级等数据。

典型案例

铁等离子体不透明度实验谱的理论模拟分析与诊断

铁（Fe）等离子体的不透明度在武器物理、天体物理等领域都是关键的问题，其不透明度的精度会对一些模拟结果产生重要影响。原子分子数据库采用相对论多组态相互作用方法，为该研究提供了不同离化度铁离子的能级（约5万条）和辐射跃迁数据（约80万条），应用于铁等离子体不透明度实验谱的理论模拟之中。为服务课题进行更高精度的理论模拟和分析，得到更可靠的等离子体状态诊断结果做出了贡献。



为钼等离子体光谱分析、诊断提供原子数据服务

惯性约束聚变是利用热核燃料的惯性来约束热核等离子体,从而实现核聚变反应的一种方法。在国家重大专项“X基本参数研究”中,对复杂的等离子体状态进行模拟需要大量的光谱数据作为支撑，而原子数据的精度对诊断结果有较大影响。原子分子数据库采用相对论多组态相互作用方法，提供了不同高精度的高Z离子能级结构和辐射跃迁数据和光谱展宽数据，应用于高温钼等离子体不透明度谱的模拟与光谱诊断，进行更高精度的理论模拟和分析，得到了更可靠的等离子体状态诊断结果，在评估等离子体温度密度状态、及相关不透明度理论模型的可靠性研究发挥了重要作用。



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类数据库，集成了光学领域内涵盖基础光学、工程光学等的公开数据。该库始于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，包括了结构数据以及各种像质评价结果数据和图形。所有数据通过站点查询下载，与光学设计软件集成两种方式提供数据共享服务，已为国内外上千家行业单位、数十万用户提供数据支持。

娄洪伟
louhw@ciomp.ac.cn
中国科学院长春光学精密机械与物理研究所

主要数据资源

序号	数据库/集名称	数据量
1	光学玻璃数据库	1.01GB
2	光学晶体数据库	6.20GB
3	光学塑料数据库	17.90GB
4	光源数据库	11.05GB
5	光电探测器数据库	0.98GB
6	光学传感器数据库	1.09GB
7	光学镜头数据库	2.40GB
8	光学棱镜	1.99GB
9	光学标准库	2.00GB
10	镀膜材料及膜系数据库	2.03GB
11	光学成像数据库	430.20GB
12	光学检测数据库	4.02GB
13	特色光谱信息数据库	88.10GB
14	激光器件数据库	6.10GB
15	激光传输特性数据库	15.20GB
16	激光光学系统数据	5.00GB
17	光纤光学数据库	6.01GB
18	微光与红外器件数据库	4.70GB
19	红外传输特性数据库	4.00GB
20	微光与红外光学系统数据库	7.01GB
21	短波光学数据库	1.98GB

服务成效

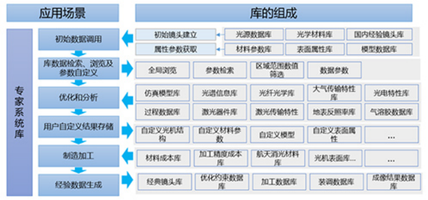
光学技术数据库2010年以来承担国家相关光学系统设计项目20余项，支持国家重大项目超过5项。

支撑论文	19篇	累计独立IP数	7万个
支撑项目	44项	累计访问人数	21万次
FTP批量数据申请	214次	累计页面访问数	76万次
在线计算任务	875个	累计下载量	50TB
在线数据预订	185个		

典型案例

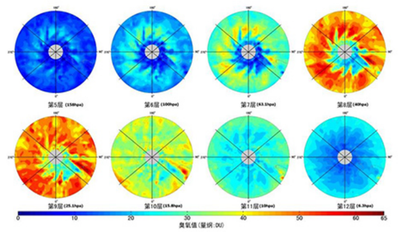
为国产化光学设计分析软件的研制提供光学基础数据

2020年，国家在国产化软件“卡脖子”工程项目中立项支持国产化光学设计分析软件建设，光学技术数据库为该软件研制提供了各类光学基础数据，如光学镜头数据、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，同时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，支持了软件的研制、推广，对我国核心软件自主化生产和生态形成做出了重要贡献。



支持大变倍比高清电动长焦监控用镜头产业化研究

目前国内监控企业生产的镜头以低端短焦产品为主，高清长焦镜头长期依赖进口。该项目目标是将大变倍比高清电动长焦监控镜头研制成果产业化，解决高清电动长焦镜头国产化问题。光学技术数据库支持研究的关键技术包括高清、大变倍比、宽光谱、体积小、重量轻、低成本、适合批量生产的镜头设计技术；长焦镜头加工、装调与检测等批产工艺技术。目前，“大变倍比高清电动长焦监控用镜头产业化研究”项目已经完成原理样机研制，后续将重点解决适合批产的设计和工艺等问题。



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

核物理主题数据库是核基础数据合集，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、裂变产额和核结构、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的重要数据，是核基础研究、核装置设计与运行、核技术应用的重要数据。该库为我国重大核科学装置的建设提供了数据支撑，同时中国评价核数据子库是我国自主研发、国际公认的五大中子评价核数据库之一，具有较大的国际影响力。

主要数据资源

序号	数据库/集名称	数据情况
01	中子诱发核反应全套数据	2800多个核素，包括综合说明信息、共振参数、中子截面、角分布、能谱、双微分截面、及部分核素的协方差数据等
02	原子核特性数据	3000多个核素，包括核素的质量亏损、稳定核素之丰度、稳定核素或长寿命放射性核素之热中子俘获截面、半衰期、衰变能、原子核基态或原子核同核异能态之自旋与宇称等
03	核反应实验数据	国内和国际23000多个核反应实验的数据，共约20万个EXFOR格式的文件
04	常用衰变数据和纲图	300个核素的原子核衰变与能级信息
05	钍铀循环专用核数据库	405个核素的全套中子数据；约1100个产物核裂变产额数据；54个核素的衰变数据

金永利
jyl@ciae.ac.cn
中国原子能科学研究院

服务成效

在线计算和绘图	7.0万次	累计数据下载量	347.32GB
累计访问文件数	3,506,872个	科研院校申请专用账号	36个

典型案例

示范快堆伪裂变产物数据评价

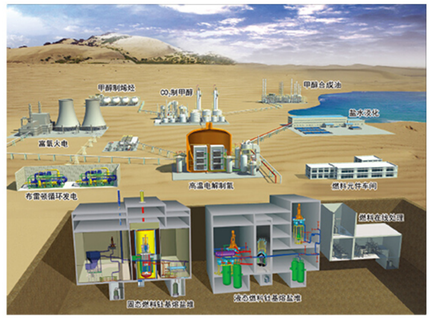
钠冷快堆是第四代核反应堆的主要堆型之一，也是中国核工业发展的标志性工程。为实现快堆的物理和屏蔽设计，科研人员在核物理主题数据库常用核衰变数据库基础上研发了一套包含伪裂变产物在内的专用中子–光子耦合多群常数库，根据裂变产物的裂变产额、衰变数据和中子活化截面等物理量进行主要裂变产物的筛选和伪裂变产物数据的评价。最终筛选出约150种常用裂变产物核，完成了伪裂变产物的评价，并制作成了AMPX格式中子–光子耦合库，支撑了示范快堆的设计与建设。



福建霞浦示范快堆

专用微观核数据库应用于钍基熔盐堆核能系统

钍基熔盐堆核能系统（TMSR）是第四代先进核能系统的 6 种候选之一。为满足中国科学院战略性先导科技专项钍基熔盐堆核能系统（TMSR）项目设计和分析需求，核物理主题数据库与上海应用物理研究所合作研制了一套核素种类完整的评价核数据库CENDL–TMSR。该库包括3个专用微观核数据库，即评价中子数据库、裂变产额数据库、衰变数据库，核物理主题库的最新国际数据支撑了该库的更新和钍铀核燃料循环中的关键数据评价，目前三个专用微观核数据库已用于TMSR项目实验堆的设计和分析研究。



基于TMSR的核能综合利用前景

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

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星系宇宙学与暗能量研究团组承建，以中国科学院紫金山天文台的“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400TB左右，并以每年约30%的速度增加。截止目前，该库已为国内外70余位天文专业注册用户提供数据服务，平均每年支撑论文数十篇，已申请国家发明专利10项。

400TB+
数据总量

70位+
专业注册用户

10项
发明专利

主要数据资源

序号	数据库/集名称	数据量
01	SDSS,WISE,ROSAT星系,星系团的星表库观测数据库	10GB
02	NIHAO流体数值模拟数据库	16TB
03	弱引力透镜全天球光线追踪数值模拟数据库	28TB
04	ELUCID宇宙学数值模拟数据库	64TB
05	FAST观测图像	3TB
06	illustris TNG	64TB
07	CSST光路模拟	20GB
08	270亿粒子模拟	60TB

王蕾
leiwang@pmo.ac.cn
中国科学院紫金山天文台

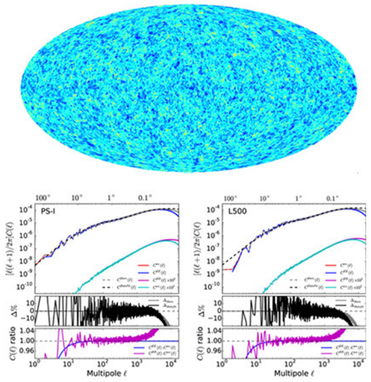
服务成效

支撑论文	10篇	平均年访问量	12万人次
国家发明专利	10项	总计在线计算任务	34211个
支撑项目	11项	总计在线计算时间	2188万CPU小时
平均年数据调用量	430TB	总计注册用户数量	78个

典型案例

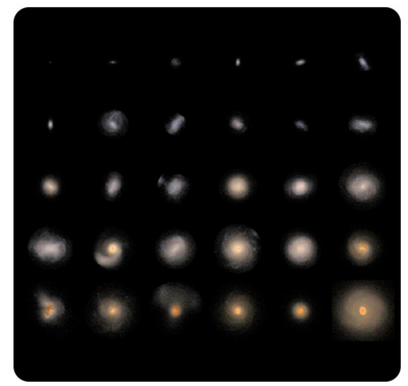
为弱引力透镜研究提供数据支撑

康熙、李国亮等研究人员利用宇宙学数值模拟数据库，结合自主研发的星系形成半解析模型和球面光线跟踪技术，完成了一个全天、高精度的弱引力透镜成图，构建了弱引力透镜全天球光线追踪数值模拟数据库。通过对该数据库提供的信息进行研究发现模拟数据预言的星系形状功率谱与目前国际领先的巡天项目KiDS（千平方度巡天)的观测结果符合度较高，这是首次利用透镜模拟数据再现出实际观测数据。同时该研究对解决弱引力透镜的核心科学问题提供了重要参考。迄今为止基于弱引力透镜全天球光线追踪数值模拟数据库的研究论文已经有十余篇。



为类银河系研究提供数据支持

紫金山天文台与德国马普天文所合作开展“你好”模拟(NIHAO：Numerical Investigation of a Hundred Astro-physical Objects)是近年来国际上比较有影响力的宇宙学的流体数值模拟项目之一。NIHAO流体数值模拟数据库从大尺度环境中挑选了一百个有代表性的盘状星系样本(类似银河系)对其初始状态扩充粒子数后做详细的模拟，从而在更高的分辨率上研究星系的形成与演化细节。这对于大尺度的宇宙学数值模拟是一个有益的补充，也是将来探索星系周介质（CGM）的性质，揭开重子物质失踪之谜（Miss-ing Baryon Problem）的理论武器。基于该数据库已经产生数十篇科研论文，均发表在国际天文学顶级期刊上。



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

射电天文与深空探测数据库



射电天文和深空测量数据库是基于中国VLBI网开展的月球与深空探测项目、天体物理、天体测量和大地测量研究所获实测数据构建的科学数据集。其中，深空测量数据包括探月工程一期嫦娥一号、探月工程二期嫦娥二号、嫦娥三号、嫦娥四号、探月工程三期嫦娥五号再入返回试验和我国首次火星探测天问一号等系列任务的VLBI测量数据。该数据库直接服务于月球及火星探测器的工程测控，并为科学探测研究提供高精度的轨道和位置等重要信息，为我国月球与深空探测工程实现与科学研究提供了重要的数据支撑。其中，嫦娥系列月球探测器VLBI测量数据是国内唯一、也是国际上少有的连续服务于月球与深空探测的VLBI数据集。该库经历了多次工程任务考验，在项目论证、关键技术攻关、系统测试、任务演练等方面发挥了极其重要的作用，为我国月球与深空探测任务中VLBI测量精度的提高和测量能力的进步提供了不可或缺的数据支撑。

陈中
chenzhong@shao.ac.cn
中国科学院上海天文台

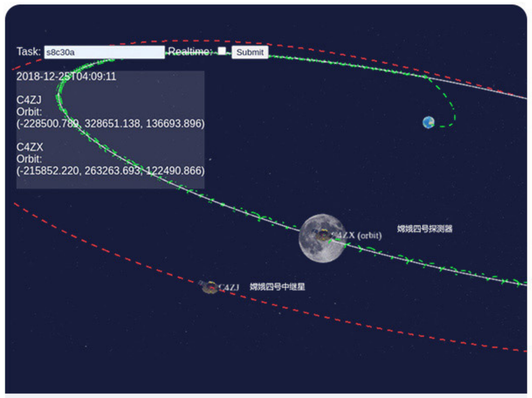
主要数据资源

序号	数据库/集名称	数据量
01	探月工程嫦娥一号VLBI测量数据	1.8TB
02	探月工程嫦娥二号VLBI测量数据	4.2TB
03	探月工程嫦娥三号VLBI测量数据	6.8TB
04	探月工程嫦娥五号再入返回飞行器VLBI测量数据	1.8TB
05	探月工程嫦娥四号VLBI测量数据	9.3TB

服务成效

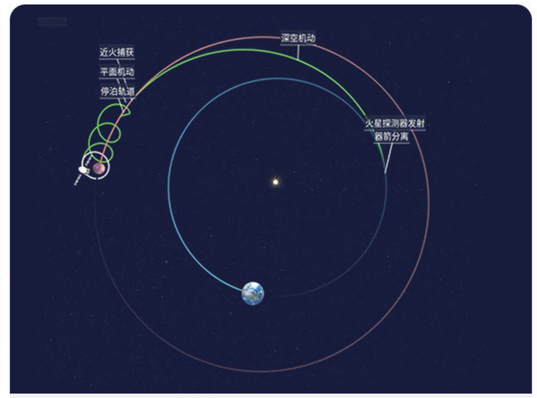
支撑论文	35篇	服务国家自然科学基金项目的数量	26个
探月工程内网在线计算任务	304个	服务省部级项目的数量	10个
服务国家科技重大专项项目的数量	2个	服务其他科技计划项目的数量	4个
支持国家重点研发项目	2个		

典型案例



探月工程嫦娥四号任务测定轨数据服务

探月工程嫦娥四号任务包括中继星和探测器两次任务，中国VLBI网主要开展探测器地月转移段和环月段的ΔDOR测量工作，并承担探测器的轨道确定和预报。射电天文和深空测量数据库保存的嫦娥四号任务VLBI观测数据，被用于嫦娥四号任务验证及演练过程。根据数据需求，该库组织整理出符合特定要求的数据集，提供数据资源路径、数据类别、数据大小、数据报告等在实际任务中所需的数据、元数据和辅助数据等。大大节约了项目开发时间，加快了技术方案验证，是项目实施方案编制的重要依据，为探测器高精度测控提供有力保障，支撑探月工程嫦娥四号任务圆满完成。目前嫦娥四号已经完成世界上首次月背着陆，并将开展长期探测工作。



中国首次火星探测天问一号数据资源服务

天问一号火星探测的主要任务是研制并发射火星环绕器和巡视器、实现火星环绕及软着陆、开展火星环绕及就位探测。中国VLBI网主要承担火星转移段、火星捕获段和进入段、环火探测段等飞行段的VLBI测量。射电天文和深空测量数据库的使用大大节约了项目论证时间，加快了技术方案验证，是项目实施方案编制的重要依据及实际VLBI高精度测量的有力保障。目前天问一号探测器正在奔向火星，已完成三次中途轨道修正和一次深空机动。射电天文和深空测量数据库实时跟踪和持续汇聚高质量的测量数据。根据近4个月的观测数据质量统计，时延测呈平均误差<0.5ns,时延率测量平均误差<0.5ps/s。相关数据将用于后续高精度VLBI测定轨研究，将进一步提升我国VLBI测定轨能力，达到国际前沿水平。

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是由中国科学院空间应用工程与技术中心负责建设、运行、管理和维护的数据库，包括天宫一号、天宫二号、天舟一号应用任务的各类别、各级别空间科学与应用数据产品，覆盖了空间生命科学与生物技术、微重力流体物理及燃烧科学、空间材料科学、微重力基础物理、空间环境与空间物理、空间天文、空间地球科学、空间新技术试验等8个研究领域。数据库服务网站正式注册的国内外用户达300余家单位，其中包含国家部委、科研院所、高等院校和企事业单位等，累计提供的数据产品量达60TB，累计数据产品下载量近10TB，累计下载次数近30万次，数据条目39万余条，累计访问量达120余万次，为推动载人航天空间科学与应用数据的共享和应用发挥了不可替代的作用。

60TB
数据总量

39万条
数据条目

300家+
服务用户

30万次
累计下载

120万次+
平台访问量

8个
研究领域

李轩
msadc@csu.ac.cn
中国科学院空间应用工程与技术中心

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

主要数据资源

序号	平台	载荷/实(试)验名称	数据量
01	天宫一号	高光谱成像仪 电离层扰动探测仪 空间环境探测仪	10TB
02	天宫二号	宽波段成像光谱仪 三维成像微波高度计 多波段紫外临边成像仪 高等植物培养实验 空间冷原子钟 空间环境探测器 伽玛暴偏振探测仪 液桥热毛细对流实验 空地量子密钥 综合材料实验	48TB
03	天舟一号	科学实验检测单元 主动隔振实验 快速水接头 蒸发冷凝实验 非牛顿引力实验 空间生物反应器	733GB

服务成效

37篇
支撑论文

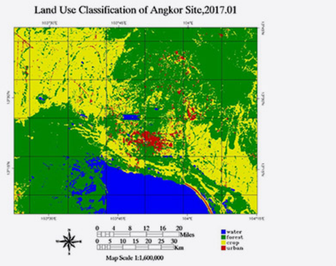
4项
专利

6项
标准

典型案例

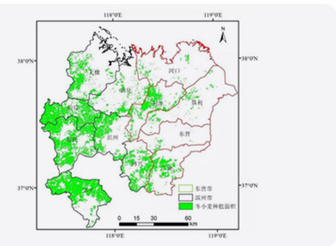
基于天宫二号数据–柬埔寨吴哥窟古迹环境动态监测

在世界文化遗产柬埔寨吴哥窟古迹环境动态监测工作中，利用载人航天空间科学与应用数据库提供的天宫二号宽波段成像光谱仪和三维成像微波高度计数据产品，结合长时间序列其他遥感数据，监测柬埔寨吴哥窟世界文化遗产地本体及周边地区的环境要素动态变化（包括森林、水体、建筑等土地利用类型），并研究其变化的驱动因素，该项工作挖掘了天宫二号宽波段成像仪数据和三维成像微波高度计数据在文化遗产环境监测中的应用潜力，推动世界文化遗产的科学管理与可持续发展。



基于天宫二号的冬小麦与棉花种植面积监测研究

在冬小麦与棉花种植面积监测研究中，科研人员利用载人航天空间科学与应用数据库提供的天宫二号宽波段成像光谱仪数据，结合地面野外调查和统计抽样数据，针对我国冬小麦和棉花主产区之一的山东省滨州市开展了冬小麦、棉花种植区面积和产量估算工作。研究结果表明宽波段成像仪可见近红外第1、第2和第14波段对于识别冬小麦十分有效，可以用来进行区域尺度冬小麦种植面积的计算；同时融合短波红外和热红外图像后，可综合利用冬小麦冠层湿度、温度等信息，进一步提高冬小麦种植面积空间分布识别的精度。相关应用工作为农业生产生活的规划管理起到了积极作用。



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，数据来源主要包括实验数据、手册数据以及文献数据。数据采用了普通分类方法，为基础研究人员、研发工程师和企业界之间建立联系。目前总数据量超过20万条，服务网站的累计数据下载量已达143.58GB，访问量达到369万次。

20万条+
数据总量

143.58GB
下载量

369万次
总访问量

崔丽娜
material@imr.ac.cn
中国科学院金属研究所

参建单位
中国科学院化学研究所
国家纳米材料中心

主要数据资源

序号	数据库/集名称	数据量
01	金属材料节点	11.5万条
02	有机高分子材料节点	9.25万条
03	纳米材料节点	1512条，1.18GB

服务成效

20篇
支撑论文

2项
软件著作权

2项
专利

9项
支撑项目

服务国家科技重大专项、重大工程清单

领域名称	支持项目名称	支持项目来源
国家重点研发计划	国家重点研发计划“材料基因工程关键技术与支撑平台”中“材料基因工程专用数据库架构、标准规范与数据库”（2016YFB0700500）项目	科技部
国家重点研发计划	材料环境失效过程的高通量计算模拟与高通量实验技术（2017YFB0702100）项目	科技部
科技部重点专项	为科技部“重点基础材料技术提升与产业化”重点专项“高品质阻燃纤维及制品关键技术”（2017YFB0309300）项目的研制，提供树脂材料性能评估标准及前沿研究进展、发展趋势。为该项目顺利推进提供了支持。	科技部

典型案例

支持港珠澳大桥耐久性防护工程实施

港珠澳大桥是世界最长的跨海大桥，其设计标准打破了国内通常的“百年惯例”，制定了120年设计标准。港珠澳大桥的基础防护工程主要是对钢管复合桩的防腐施工，钢管桩位于海泥环境中，防腐涂层的破坏方式主要来源于打桩过程中的机械损伤、泥砂碎石磨划伤和泥下腐蚀因素的长期侵蚀、性能衰退等。其中，材料科学基础数据库的腐蚀数据子库为跨海大桥耐久性设计方案提供数据依据，支持了中国科学院金属研究所自主研发的“SEBF/SLF高性能涂层加牺牲阳极”联合防护技术的研制和落实，为港珠澳大桥的建造通车做出贡献。



钢管复合桩内外涂装

为科技部立项的国家重点研发计划提供数据支撑

材料科学基础数据库面向支持为科技部立项的国家重点研发计划“材料基因工程关键技术与支撑平台”中项目“材料基因工程专用数据库架构、标准规范与数据库”（2016YFB0700500）、项目“材料环境失效过程的高通量计算模拟与高通量实验技术”（2017YFB0702100）提供金属材料和有机高分子材料数据，推动了项目的实施和关键技术突破。



项目页面

干细胞与代谢疾病数据库

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

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，基于云计算和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。

该库以中国科学院广州生物医药与健康研究院、广州再生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基础来源，同时汇聚高校、科研单位、医药企业多源异构数据，包括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实验室及研究中心、科技计划项目和应用平台，为人类健康与“健康中国”战略实施提供强有力支撑。

主要数据资源

序号	数据库/集名称	数据量
01	代谢疾病数据资源库	1MB
02	干细胞组学资源库	500GB
03	临床医学知识系统	1TB
04	健康科学数据库	20TB
05	人类致癌基因数据库	400GB
06	人类致癌转移数据库	10GB
07	自身抗体生物标志物数据库	30MB
08	单细胞肿瘤数据资源库	4.8GB
09	小鼠细胞图谱	60TB
10	人类细胞图谱	50TB
11	细胞多组学数据库	1TB
12	国际开放数据库	50GB

赵娟娟
zhao_juanjuan@gibh.ac.cn
中国科学院广州生物医药与健康研究院

服务成效

数据汇聚共享

促进数据融合治理和交叉研究，实现PB级数据汇聚融合、分析处理、可视化工具与知识发现服务

数据挖掘分析

对细胞组学数据及临床代谢疾病多源异构数据的典型处理算法进行优化，搭建高通量数据处理的可视化交互工作流，提升对数据的流程化、标准化统合分析

数据创新利用

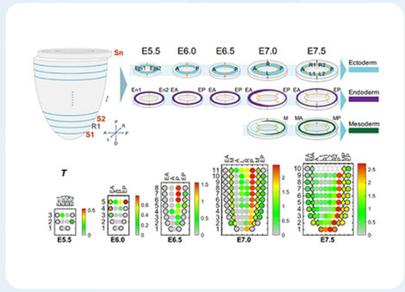
建立干细胞与代谢疾病领域数据清洗、转换、标准化与标注的数据治理体系

SCI论文23篇	广东省自然科学奖 1项	专利 1项

典型案例

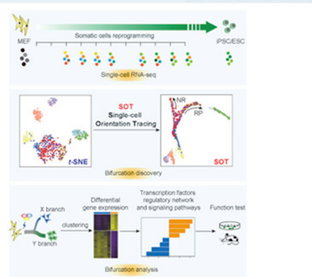
小鼠早期胚胎发育过程中全胚层时空分子谱系

干细胞与代谢疾病数据库汇集的小鼠细胞图谱数据资源在基于激光显微切割的低起始量空间转录组分析方法(Geo-seq)建立、和小鼠早期胚胎着床后高分辨率时空转录组图谱的构建起了重要的支撑作用。科学家首先完成了小鼠早期胚胎的三维空间分子图谱，建立起百科全书式全基因组的时空表达数据库，并构建了小鼠早期胚胎发育过程的系统发生树，从分子层面重构了胚层谱系的发生过程。在对信号通路富集分析的基础上，发现胚层谱系发生过程中关键调节机制。



单细胞测序解析体细胞重编程路径

干细胞与代谢疾病数据库搭建的流程化数据挖掘、可视化交互式平台，促进了Single-cell Orientation Tracing (SOT)新算法的顺利研发。汇集的人类细胞图谱和细胞多组学数据库，在OCT4/SOX-2/KLF4 (OSK) 重编程单细胞测序数据挖掘分析，利用SOT算法绘制重编程路线以及非重编程分支过程发挥重要作用。基于数据与计算平台，研究人员提出了广泛适用于细胞命运决定的分支模型，并且提供了利用单细胞测序分析复杂命运变化的研究中“发现分支-预测原因-功能验证”的研究范式。



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是中国科学院心理研究所磁共振成像研究中心主导建立的包括广泛个体差异的人群的脑影像数据库。该数据库依托于中国科学院心理研究所，包括中国彩巢计划、中国重度抑郁症数据库、国际信度与可重复性(CoRR)大数据共享等数据资源，此外包括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，以及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。数据库自正式上线以来，已有数万名国内外脑科学领域的科研人员下载并使用该库共享数据。

1万+
多模态人脑活体成像

数万人次
访问量

CCS+DPABI
独立研发人脑影像数据处理软件

主要数据资源

序号	数据库/集名称	数据量
01	中国彩巢计划	521套
02	中国重度抑郁症数据库	2428套
03	国际信度与可重复性(CoRR)大数据共享	1629套

左西年
zuoxn@psych.ac.cn
中国科学院心理研究所

服务成效

3项
支撑项目

4578套
数据量

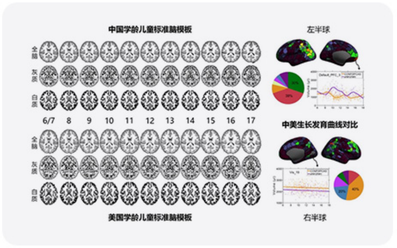
40篇
支撑论文

2项
获奖

典型案例

中国彩巢计划数据支持中国学龄儿童标准脑模板研发

“学龄儿童青少年脑与心智发育研究”项目完成了521人次6–18岁学龄期儿童青少年脑影像及行为数据的采集，构成了我国首个学龄儿童脑与心智发育队列“彩巢计划—成长在中国”。基于该中国活体人脑成像共享数据子库，中国科学院心理研究所主导研发了中国学龄儿童标准脑模板，绘制并比较了中、美学龄期儿童的脑发育曲线，推进了脑发育测量的信效度方法学，填补了脑智发育模板跨文化比较的空白，提高了国内脑成像研究在国际领域内的影响力。



中国重度抑郁症功能磁共振数据集

中国重度抑郁症功能磁共振数据集包含国内共17家医院、25位研究者参与贡献的数据。目前已汇集2428名被试（1300个抑郁症患者和1128个健康对照样本），形成了国际最大的抑郁症功能磁共振数据集。该数据集通过对已有的抑郁症静息态功能磁共振成像数据进行重新分析和再挖掘，发现了中国人的抑郁症全脑静息态活动的异常，特别是抑郁症患者默认网络内部功能连接的异常，将静息态功能磁共振成像在抑郁症的临床应用方面推进一大步。





服务成效

地理空间数据云已建立成完善的数据资源储备库，实体数据汇聚量157TB，专业用户数超40万人，广泛服务于地学、遥感、生态、水利、交通等众多领域，为科学研究提供重要数据和技术保障。“大规模地理空间数据云服务关键技术与应用”作为地理空间数据库核心组成部分，获2016年度北京市科技奖二等奖。

注册用户数	416,157人	CNIC论文支持数	7,474篇	累计独立IP数	188.56万个
FTP批量数据申请	15,457次	SCI论文支持数	914篇	累计访问人次	406.42万次
在线计算任务	61,948个			累计页面访问数	2.97亿次
在线数据预订	17,967个	Googlescholar论文支持数	2,970篇	累计下载量	826.66TB

地理空间数据云是先进的“一站式”地学数据综合服务平台，自主研发了可靠的大规模地理空间数据实时分析引擎，实现了亿级元数据高效管理和PB级实体数据分析服务。数据库汇聚地理空间数据资源，生产优质数据产品，创新数据服务模式，开展数据搜索、数据下载、数据预定、批量申请、数据众包、模型计算、在线制图等多种数据服务。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。

林青慧
data@cnic.cn
中科院计算机网络信息中心

地理空间数据云

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

主要数据资源

序号	数据库/集名称	数据量
01	LANDSAT系列数据	50.01TB
02	MODIS陆地标准产品	36.76TB
03	MODIS中国合成产品	4.54TB
04	MODIS-L1B标准产品	46.29TB
05	DEM数字高程数据	0.92TB
06	EO-1系列遥感数据	3.24TB
07	NOAA AVHRR数据产品	1.56TB
08	大气污染插值数据	1.37TB
09	SENTINEL系列遥感数据	12.8TB
10	用户“数据汇”分享数据	4GB

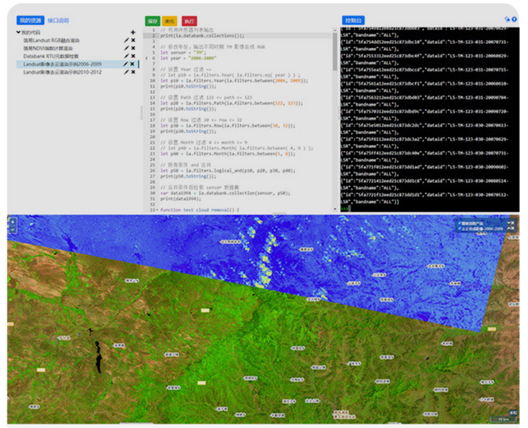
157TB
数据总量

750万条
数据条目

40万+
服务用户

1万余篇
支持论文

典型案例



为“地球科学大数据”提供全面支撑

围绕科研人员在地理空间数据搜索、获取、存储、在线分析和可视化的迫切需求，地理空间数据云基于已汇聚的数十颗国际遥感卫星数据和国内陆地观测卫星数据，建立完善的数据储备库，并对大规模遥感对象存储、高并发遥感对象IO驱动、数据粒度分布式任务调度、免预切大规模地图引擎、按需交互渲染计算、交互分析应用和算法数据集成等多项关键技术进行突破，形成一套面向遥感大数据管理、分析和应用的软件技术体系，实现海量数据实时在线分析服务，支撑地球大数据领域的技术创新与地球系统科学的重大突破和科学发现。



支撑中国历史地理数据库建设

中国历史地理数据库基于地理空间数据云核心技术和数据建设，是为社会科学类科研工作者、历史博物馆和社会大众等提供丰富数据资源和历史文化展示的平台。该库由地理空间数据云与中国历史研究院专家密切合作，系统整理和集成中国自史前起至中华民国共23个历史断代的历史地理、文化数据资源，并通过三维地图等先进动态可视化方式进行数据生动的展示，是信息全面、内容权威、技术先进的历史文化科普系统和学术研究工具，为落实“坚定文化自信、建设文化强国”提供了有力支撑。

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

青海湖区域综合研究专题数据库

青海湖区域综合研究专题数据库是集青海湖区域植被、兽类、水鸟、生态水文、遥感微生物等在内的基础数据资源收集、整合、关联性分析、集成、共享，融合数据、模型为一体的数据处理和应用环境，是基本完整的数据资源体系和一站式数据服务系统。该库共建设五大类44个子库，资源总量达20TB。服务成效方面，目前年平均下载量100GB，年平均访问量20万人次，为生物多样性保护、疫病防治、水环境监测、水华防治和保护区管理等提供了准确、及时的特色服务。

- 5类44个 子库
- 20TB 资源总量
- 100GB 年均下载量
- 20万 年平均访问人次

主要数据资源

序号	数据库/集名称	数据量
01	青海湖区域植被调查数据库	15GB
02	青海湖区域动物物种调查数据库	20GB
03	青海湖区域本底数据库	20TB
04	青海湖区域生物信息学数据库	52MB
05	青海湖区域文献资料数据库	40MB

罗泽
luoze@cnic.cn
中科院计算机网络信息中心

参建单位
中国科学院动物所
中国科学院武汉病毒研究所
中国科学院水生生物研究所
青海湖国家级自然保护区管理局

典型案例

为上海科技馆公众教育工作提供数据服务

青海湖野外鸟类监控视频自2015年7月正式接入上海科技馆至今已连续6年对公众展出，主要通过“文明的史诗”主题板块的常设主题展区“未来之路”进行展示。该展区以中国地图为背景，在不同地理位置放置了10余块用来展示保护区实时监控视频的显示屏，展示包括青海湖鸟类保护在内等工作情况，使得社会大众可以在科技馆中实时远程看到青海湖的鸟类真实生活状态，具有很好的生态保护宣传教育意义。



支持毗邻月牙湖变红事件调查与成因分析

2017年4月12日青海湖毗邻月牙湖突然变成粉红色，当地群众拍照并在网络上传播后，引起了社会各界广泛关注。青海湖数据库为本次调查事件报告的编写和成因分析、以及后续持续保护政策制定提供了重要数据支持，主要包括遥感数据、水体调查数据（水体调查植物群落数据、底栖动物数据、浮游动物数据、浮游植物数据、理化数据）、气象数据等，为该事件不同层面（管理局、地方政府、中科院）报告撰写提供了重要的依据，为青海湖流域2018–2020年持续生态保护工作决策提供了重要的支持。



服务成效

SCI论文	20余篇	专利	3项
其他论文	70余篇	国家重大项目	10余项
软件著作权	30余项	科技计划项目	5项

特色软件

野外巡护移动终端、野生动物识别等软件已在车八岭、卧龙、呼伦湖、长白山、无量山等多个保护区进行了安装和试运行。

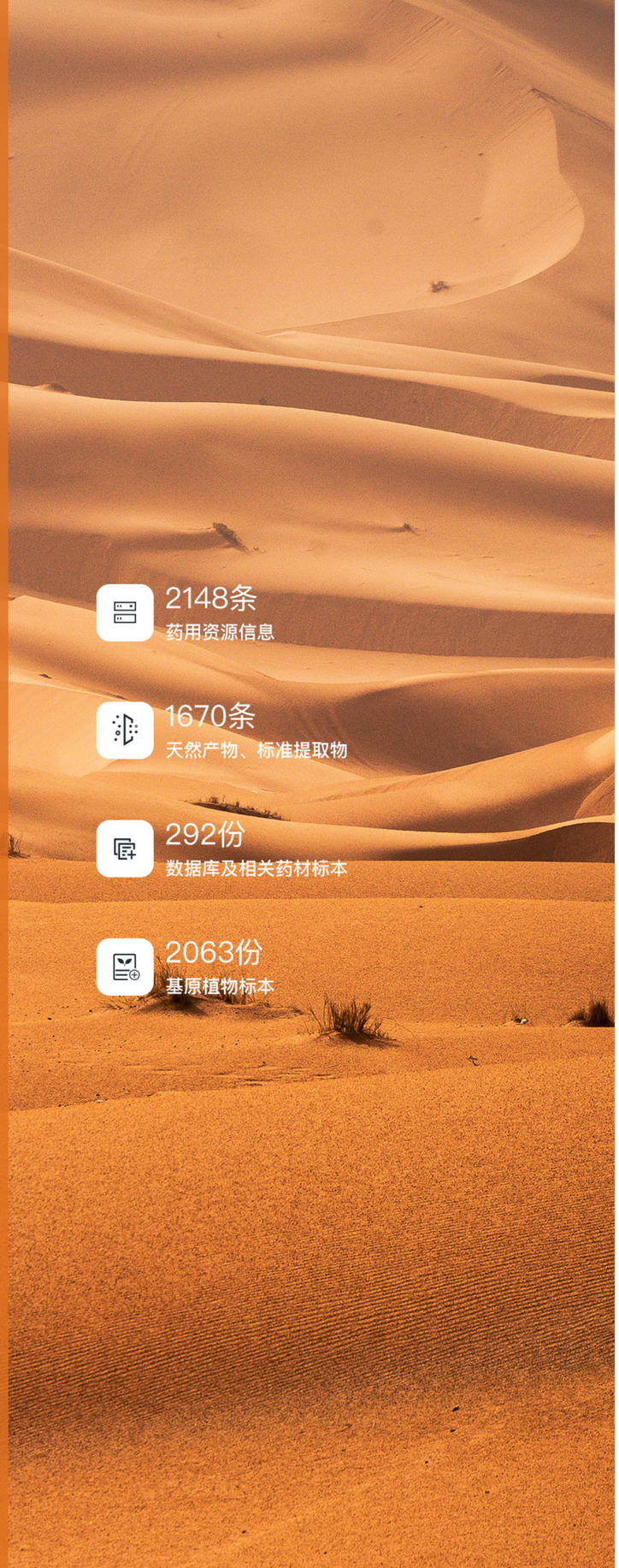
特色数据集

MODIS产品集、刚毛藻水华空间分布数据集等有效支持地方生态保护。

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是以我国新疆、中亚等范围内，涵盖干旱、半干旱地区药用资源的综合数据库，数据资源包含药用资源信息（2148条）、天然产物、标准提取物信息（1670条）、药材标本（292份）及其数据库、基原植物标本（2063份）、标准品等。该库实现了干旱区药用数据资源规范化线上发布，可面向用户提供药材基原，有效成分化合物和标准提取物等数据查询和下载服务，2020年度该库支持发表学术论文51篇、学位论文22篇、授权专利20个。



2148条
药用资源信息

1670条
天然产物、标准提取物

292份
数据库及相关药材标本

2063份
基原植物标本

刘戈宇
liugy@ms.xjb.ac.cn
中国科学院新疆理化技术研究所

参建单位
中国科学院新疆生态与地理研究所

服务成效

干旱区药用资源数据库可提供的数据服务包括：（1）药材基原信息、基原植物生境信息、标准提取物、化合物等数据查询；（2）为活性筛选提供有效成分化学结构、有效部位等数据查询；（3）为新药开发提供古籍名方基础数据。

2020年度，该数据库的典型服务案例包括：（1）支持了中药6.1类新药“乃孜来颗粒”临床前研究；（2）药用植物资源、古籍名方等基础数据为地方药企和科研机构提供道地药材质量标准信息咨询和药材鉴定服务，并协助新疆特色医药企业、民族医院完成药品、院内制剂质量标准提升；（3）通过对该数据库资源的有效利用对南疆贫困地区社会大众开展药用资源科普，促进当地药材种植业发展，为全面脱贫攻坚提供科学指导；（4）依托数据库支撑的科研项目、成果转化以及科技扶贫等方面工作成效，本数据库负责人阿吉艾克拜尔·艾萨在2020年初的《科技盛典——CCTV2019年度十大科技创新人物》评选中被评为科技创新人物。

典型案例



服务一带一路中亚联合实验室的药用植物研究

干旱区药用资源数据库依托单位与乌兹别克斯坦科学院植物化学研究所、生物有机化学研究所、塔吉克斯坦科学院等科研单位，分别建立天然药物联合实验室、生物药联合实验室和中塔天然产物创新联合实验室。干旱区药用资源数据库中的药用植物资源数据及化合物资源支持了联合实验室开展翠雀属、乌头属、阿魏属、蒿属植物等研究，同时在目标药用植物的筛选、药材基原鉴定、特征有效成分的比对分析等方面发挥了关键作用。



服务新疆医疗制剂标准提升和南疆脱贫攻坚

干旱区药用资源数据库为新疆民族医院院内制剂标准提升提供药材基原信息、提取物标准和基础科研数据查询服务，完成了98个院内制剂的药材质量标准建立和制剂成分标准提升。同时，依托库中数据资源开展“访惠聚”村民药用植物资源科普培训，参与总数达500人次。通过开展新疆特色药用植物资源和道地药材等培训，更多农村地区百姓了解药用植物种类、辨别基本药材，并掌握了药用资源和药材种植栽培技术，相关科普培训为南疆精准扶贫工作的顺利开展提供了科学依据和指导。

黑龙江省特色基础科技数据库

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

黑龙江省特色基础科技数据库是汇集黑龙江省寒地特色物种、论文、专利及大型科学仪器的数据库。近五年来，通过数据库平台进行资源共享和服务，累计支持中小微企业及单位等使用上万次。其中2017–2019年服务企业714家，涉及服务项目623项、从业人员总数近万人，支持企业及企业类服务机构新增就业人数234人，获得经费补贴6300万元。

主要数据资源

序号	数据库/集名称	数据量
01	论文数据库	1G
02	专利库	500M
03	仪器设置库	2G

荀妍妍
191540343@qq.com
黑龙省科技资源共享服务中心

参建单位

哈尔滨工业大学　黑龙江大学
东北林业大学　黑龙江省农科院土肥所
哈尔滨工程大学　黑龙江省科信科技信息中心

服务成效

153篇
支撑论文

8,835,478时
仪器库对外服务机时数

12,453个
服务单位/个人

4,948台/套
仪器库对外服务仪器

714家
政策性经费支持企业

6,300万
经费补贴

典型案例

为项目“哈长城市群综合科技服务平台研发与应用示范”平台建设提供服务

国家重点研发计划“哈长城市群综合科技服务平台研发与应用示范”的项目目标是建设哈长城市群综合科技服务平台，实现科技资源有效精准对接科技企业进行服务。黑龙江省特色基础科技数据库的数据资源支持了哈长城市群综合科技服务平台数据资源富集。目前黑龙江省特色基础科技数据库已经成功应用于哈长城市群综合科技服务平台上，截至2020年已利用相关数据资源发表论文13篇、申请软著7项。



为东北农业大学科研项目组提供科学数据服务

东北农业大学科研项目组在进行科学研究过程中，通过黑龙江省特色基础科技数据库平台查询并获取黑土长期施肥下的土壤肥力数据，该数据的分析利用对该课题的进展和成果产生了极大的促进作用，支持发表论文2篇。



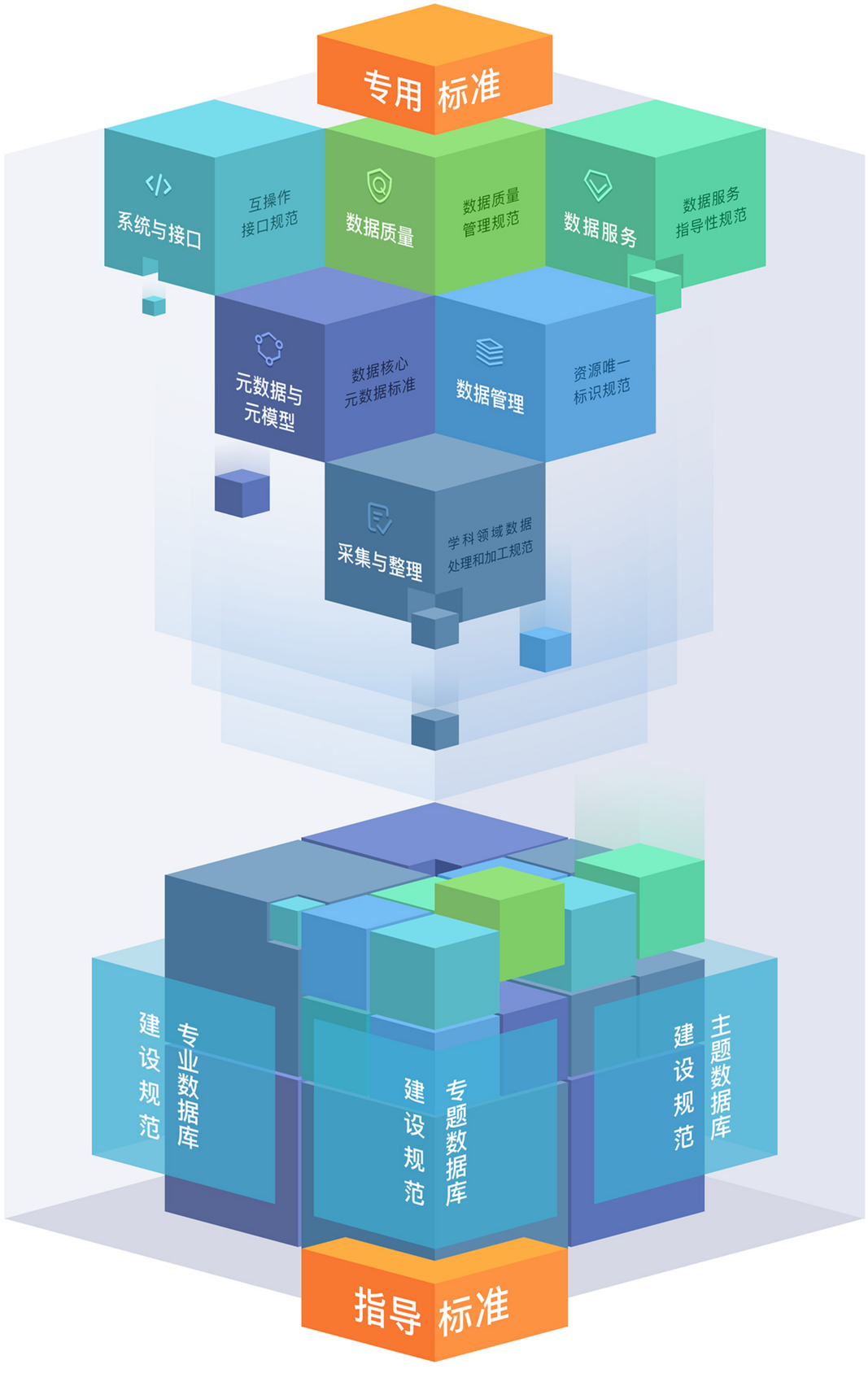
标准体系

标准体系全流程覆盖

国家基础数据中心的科学数据标准体系覆盖了从数据资源建设到应用及效果评估全生命周期流程，分为专用标准和指导标准。专用标准包括采集与整理、元数据与元模型、系统与接口、数据管理、数据质量、数据服务等多种类型。形成的40余项标准规范，在科学数据资源建设和应用服务中发挥了积极作用。中心积极推动国家标准的研制工作，共牵头完成6项，参与11项。2020年9月《科学数据标识编码规范》、《数据论文出版元数据》和《科学数据收割规范》三项标准通过国家标准立项公示。

牵头或参与完成的国家标准列表

序号	国家标准名称	国家标准号
01	生态科学数据元数据	GB/T 20533–2006
02	检测资源信息共享体系建设指南	GB/Z 27414–2012
03	科技平台 元数据注册与管理	GB/T 30524–2014
04	科技平台 资源核心元数据	GB/T 30523–2014
05	科技平台 一致性测试的原则与方法	GB/T 31071–2014
06	科技平台 服务核心元数据	GB/T 31073–2014
07	土壤科学数据元数据	GB/T 32739–2016
08	信息技术 科学数据引用	GB/T 35294–2017
09	信息技术 数据溯源描述模型	GB/T 34945–2017
10	信息技术 大数据 术语	GB/T 35295–2017
11	信息技术 数据交易服务平台 交易数据描述	GB/T 36343–2018
12	信息技术 数据质量评价指标	GB/T 36344–2018
13	信息技术 数据交易服务平台 通用功能要求	GB/T 37728–2019
14	科学数据标识编码规范	研制中
15	数据论文出版元数据	研制中
16	科学数据收割规范	研制中



工具与技术体系

大数据处理与分析平台

管理工具集

- PiFlow — 大数据流水线
- PackONE — 大数据环境快速部署

分析工具集

- DVIZ — 在线可视分析
- iAnalysis — 在线交互分析

数据管理与支撑服务

公共服务

- MSIS — 监控统计
- DRS — 咨询服务
- RESSTAT — 资源量统计
- DataPid — 科学数据唯一标识服务平台
- ScienceDB — 科学数据存储库

汇聚发布·自由组配

- **共享与服务** — 发现访问 · 推荐过滤 · 社交评价
- **数据管理与发布** — 中心发布 · 本地发布 · 融合管理
- **数据汇聚传输** — 逻辑汇聚 · 物理汇聚 · 混合模式

面向科学数据中心体系化、系统化建设及管理的需求，国家基础数据中心研发重点包括分布式数据资源自主管理与服务，数据资源集成整合管理与服务，数据资源与服务监控、统计分析及评估管理，数据资源统一发布、共享、服务及应用等，同时面向不同类型数据提供统一数据资源管理和外延服务，形成了系列工具。



01 体系化的解决方案

采用体系化的设计理念，提供自顶向下的总体设计，协同策略、系统、规则、设施等各方面，保障数据资源高效、安全、可靠、敏捷的管理与服务

02 完备的服务平台和软件体系

提供包括集成化数据服务、数据资源监控与管理、数据资源整合、分布式数据资源自主管理工具在内的软件体系，并为用户个性化需求提供定制和支持

03 统一的管理和外延服务

基于标准化的数据资源管理机制提供一整套有效的管理办法，包括数据库建设、系统运维、服务咨询在内的多项外延服务

04 建设经验丰富多样

十余年来，技术体系为国家基础数据中心数据资源建设与服务有效助力，在动物、植物、化学、物理和核科学等领域深入探索，服务于科学领域信息化建设，提供了体系架构、系统设计、开发、服务一体化的建设方案，支撑服务科研单位60余家

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

DataPID
数据标识系统

数据标识系统DataPID为各类数据分发全球唯一永久标识(PID)，支持数据开放共享，是构建数据的管理注册和应用生态的基础支撑系统。基于全球Handle标识高效解析机制和灵活开放的应用管理策略，研发标识分级注册技术和异构多址解析技术，实现了海量数据实时注册校验、高并发解析查询及元数据策略开放定制支持数据标识的统计、引用关联等融合应用。

郭志斌
gzb@cnic.cn

平台服务

- Handle、CSTR双标识注册解析查询
- 标识元数据修改、发布机制
- 标识子节点自动部署和管理
- 支持自定义元数据注册、查询与检索
- 支持OpenAPI、XML、JSON多方式数据对接
- 标识编码规范
- 标识关联、标识图谱、标识追溯、标识区块链

目前，数据标识系统安全稳定运行，与国际数据标识系统互通。其已为中国科学院A类战略性先导科技专项“地球大数据科学工程”、高能物理所散裂中子源数据平台、国家空间科学数据中心、地理空间数据云、《中国科学数据》期刊等提供数据标识服务，累计标识注册量49万余条。

典型应用

支持“地球大数据科学工程”数据标识

支持空间科学数据标识

支持论文标识

系统特性

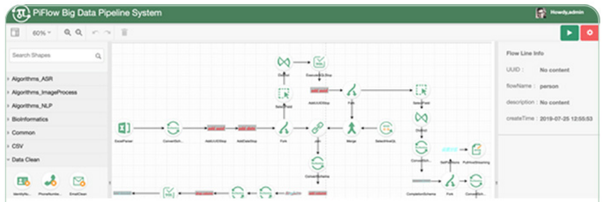
简单易用 提供可视化界面配置流水线，实时监控流水线运行状态，查看日志，提供模板功能

功能强大 提供100+的数据处理组件， 包括Hadoop Spark、MLlib、Hive、Solr、Redis、MemCache、ElasticSearch、JDBC、MongoDB、HTTP、FTP、XML、CSV、JSON等，同时集成了AI、NLP等领域的相关算法

按需扩展 支持自定义开发数据处理组件、满足项目实施过程中的特定需求

性能优越 基于分布式计算引擎Spark开发，性能表现优越

典型应用



PiFlow在基金委项目中的具体应用

PiFlow已应用在多个项目中，以国家自然科学基金委大数据知识管理服务平台为例。将基金委内部各个子系统在数据层面进行整合，自动化实现人员、项目、机构、成果的数据抽取和汇聚，有利于隔离业务生产环境和数据分析环境，高效满足基金委科研项目人员关联分析、回避关系挖掘、历史数据多维统计等分析需要，促进科研项目的发布、申报、立项、结题、成果检索等高效管理。其中涉及100多项流水线配置开发，完成数据的采集、清洗和汇聚，数据体量达亿级，性能较Apache开源项目Nifi提升至少3倍。PiFlow获得“码云最有价值开源项目”。

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

PiFlow
大数据流水线系统

PiFlow是一个基于分布式计算框架Spark开发的大数据处理流水线系统，将数据的采集、清洗、计算、存储等各个环节封装成组件，以所见即所得的方式进行流水线配置。

朱小杰
xjzhu@cnic.cn

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

VDB4.0
分布式数据汇管服平台

国家基础数据中心应用云计算和大数据技术，建成了通用可定制数据管理与发布平台VDB4.0。该平台由数据汇聚与传输系统、数据管理与发布系统、数据共享与服务门户系统构成，定位于实现分布式多源异构数据物理汇聚、集中管理发布、统一共享与服务，支撑科学数据中心服务体系构建。

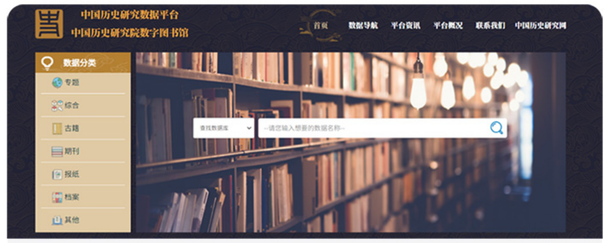
韩芳
hanfang@cnic.cn

系统特性

VDB4.0平台底层由全文搜索引擎(Elastic-Search)集群和分布式文件存储库(MongoDB)集群支撑数据存储、管理和检索，具有通用性，高可定制性特点，支持数据的汇聚、管理、服务全流程定制化管理，已在多个数据库建设中应用服务。

通用化　可定制　全流程

典型应用



中国历史研究数据平台



水生生物数据平台

应用成果

科技领域大数据知识图谱已经汇聚超10亿科研实体（论文、机构、专家等）、上百亿关系的大数据知识网络，并已在国家自然科学基金大数据知识管理服务平台、中国科协计算机与人工智能大数据知识管理平台、智慧中科院、空间中心大数据管理与应用服务平台、烟草科技知识图谱构建等项目中应用，取得了较好效果。

十亿+　科研实体
百亿+　大数据知识网络

典型应用



国家自然科学基金大数据知识管理服务平台



烟草科技知识图谱服务平台

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

Knowledge Graph
知识图谱技术

针对科技领域知识图谱需求特征，国家基础数据中心提出相应的知识图谱技术框架。在大数据基础环境支撑下，基于自主研发的PiFlow系统，利用分布式大数据采集与融合技术，设计一整套领域知识图谱构建流水线。以科技领域为切入点，突破大规模网络数据存储、实体发现与链接预测、行业趋势与交叉学科预测等关键技术，构建领域大数据知识图谱。

杜一
duyi@cnic.cn

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

iAnalysis
交互分析云服务系统

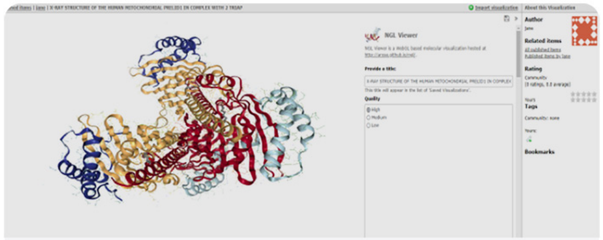
iAnalysis(IA)针对科学大数据分析中数据迁移代价大、数据分析依赖多的特点提供一种科学数据云分析服务。IA通过有效扩展的计算和存储资源、可选的算法资源库、高效的数据存取接口、安全的用户访问方式、便捷的用户交互工具辅助领域科学家进行数据分析，有效避免大规模数据迁移和编程语言处理、算法环境搭建、版本适配、资源调用等问题，应用于多个学科领域重大项目，有效提升了科学家的数据分析能力。此外，提供的中英双语版本服务于国内外用户，扩大了国产自主科学软件的国际影响力。

孟珍
zhenm99@cnic.cn

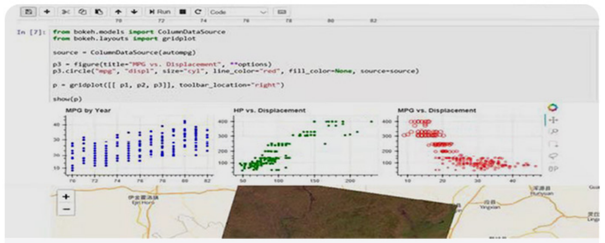
系统特性

- 资源扩展
- 版本可选
- 健康维护
- 安全访问
- 资源隔离
- 简单易用
- 高效便捷

典型应用



Darwintree分子应用环境



面向地理空间数据云应用

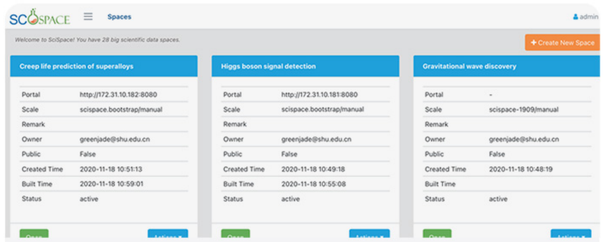
国家基础学科公共科学数据中心
National Basic Science Data Center

PackOne
大数据集群管理工具

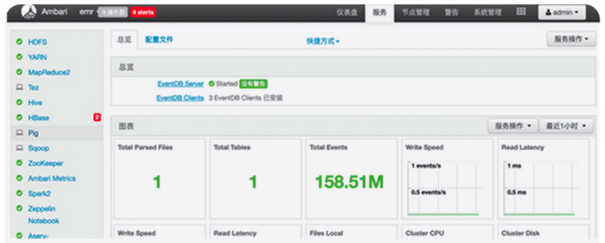
系统特性

- 支持在空白虚拟机上完成大数据处理集群的全自动部署
- 通过Apache Ambari对已部署的大数据软件进行状态监控、配置管理
- 通过将模版集群物化为系统镜像，实现新集群的分钟级快速部署
- 通过集群节点的全自动增删，实现各类大数据软件处理能力的分钟级弹性伸缩
- 在同一个界面上对来自不同云的虚拟机、存储卷、镜像、模版等进行CURD操作

典型应用



SciSpace–科研人员大数据云端分析环境



科学大数据管理系统

PackOne致力于实现主流大数据软件在云端的快速弹性部署。通过对云API和Apache Ambari API的联合调用，完成Hadoop、Spark、NiFi、PiFlow、Kylin、MangoDB、Neo4J等流行的大数据管理/处理软件在云端的一键部署和一键伸缩。

王华进
wanghj@cnic.cn

2019年
国内科学数据工作
十大进展

自《科学数据管理办法》发布以来，我国科学数据工作硕果累累。受国家科技基础条件平台中心委托，国家基础数据中心近年连续承担撰写《国家科学数据资源发展报告》任务。为全面、客观呈现我国科学数据的年度发展态势，国家基础数据中心在对国内科学数据工作前沿动态收集整理的基础上，经包括20个国家科学数据中心主任、领域专家、部分行业和部门主管领导等推荐，于2020年1–2月间开展了首次国内科学数据工作十大进展评选工作。其中，“数据作为生产要素按贡献参与分配”理论创新、首批国家科学数据中心名单公布、《地球大数据支撑可持续发展报告》、《科研数据北京宣言》和《中国科学院科学数据管理与开放共享办法(试行)》等入选“2019年国内科学数据工作十大进展”。

数据“生产要素”理论提出

中国共产党十九届四中全会提出“数据作为生产要素按贡献参与分配”为新时期科学数据工作明确了理论基础

中国科学院科学数据中心体系初步建成

中国科学院印发《中国科学院科学数据管理与开放共享办法(试行)》，并初步建成中科院科学数据中心体系，包括1个总中心、11个学科数据中心和8个优秀所级数据中心

国家科学数据中心正式成立并明确五年发展目标

科技部、财政部联合公布20个国家科学数据中心名单，并指导国家科学数据中心完成五年发展规划及其专家论证

科学数据标准国家重点研发计划项目启动

国家重点研发计划首次设立以科学数据标准为重点方向和研究内容的“科学数据安全技术及基础技术标准研究”项目，并进行立项公示

人类首张黑洞照片发布我国科学家和科学数据发挥积极作用

人类拍到的首张黑洞照片正式发布，中国科学院上海天文台等国内机构参与观测和数据处理，数据规模总量超过10PB

10

《地球大数据支撑可持续发展报告》提交联合国第74届大会

《地球大数据支撑可持续发展报告》正式发布，并提交联合国第74届大会，展示了中国利用地球大数据技术支持2030年议程落实和政策决策的探索和实践

“中国生态系统长期观测研究数据”发布

国家生态科学数据中心、国家生态系统观测研究网络、中国生态系统研究网络、中国通量观测研究联盟共同发布“中国生态系统长期观测研究数据”

CODATA发布《科研数据北京宣言》

CODATA“开放科学数据政策与实践”国际研讨会在北京召开，并发布《科研数据北京宣言》

第六届中国科学数据大会在贵阳召开

以“科学数据与重大科研基础设施”为主题的第六届中国科学数据大会顺利召开，社科数据和行业数据首次设专题进行研讨

2019世界数据系统会议在京召开

2019世界数据系统（WDS）亚洲–大洋洲会议、亚太区域会议先后在北京召开，进一步推动了我国科学数据的国际交流和合作，提升了国际影响力

运行保障

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。国家基础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。

- 315PFlops 计算资源
- 150PB 存储资源
- 52个 综合平台
- 409个 科研软件
- 9类 典型云服务
- 840万+ 免费CPU计算机时
- 160万+ GPU计算机时
- 1000台 云服务器
- 10PB 云备份资源

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 22239–2008信息安全技术 信息系统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