



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

中国科学院 科学数据总中心服务

陈昕

中国科学院计算机网络信息中心/中国科学院科学数据总中心

2023年11月29日

汇报提纲

一、背景与挑战

二、服务与实践

中国科学院科学数据工作的牵头和运行单位



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

40年科学数据工作基础

牵头中国科学院科学数据项目，随着需求与技术发展，不断演进

在资源、标准、技术、服务等方面具有扎实的基础



国家计划委员会文件

计科(资)(1986)956号

关于中国科学院科学数据库及其信息
系统工程计划任务书的批复

中国科学院：

你院(85)科发计字1014号文收悉。
现就中国科学院科学数据库及其信息系统工程计
划任务书批复如下：

一、为了适应当前信息技术发展的需要并为
四化建设提供信息服务，同意在北京市中关村建

1986-2000
年，科学数
据资源积累，
研究所/课题
组自主自治

2001-2005，
院“十五”信
息专项支持，
规范化和规模
化资源积累

2006-2010，院
“十一五”信
息化专项和国家科
技基础条件平台
等支持，领域数
据整合与系统性
开放共享(科学
数据网格)

2010-2015，
院“十二五”
信息专项和国家
科技基础条件
平台等支持，
资源深度整合
与应用服务
(科学数据云)

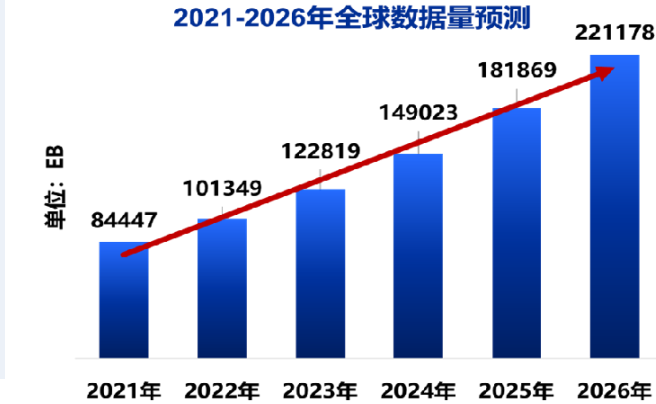
2016-2020，
院“十三五”
信息化专项和
国家科技基础
条件平台等支
持，数据资
产管理(科学
大数据)

2019-2020，
院“十三五”
信息化专项，
院科学数据中
心体系建设
(科学数据中
心体系)

2021-，
院“十四五”
网信专项，科
学大数据工程
三期建设(科
学数据网络)

近年来，科学数据的重要性日趋凸显

- 科学数据是支撑国家重大需求的战略性资源
- 科学数据是急需释放价值的新型生产要素
- 科学数据是科技基础能力的重要组成
- 科学数据是创新发展的重要驱动力

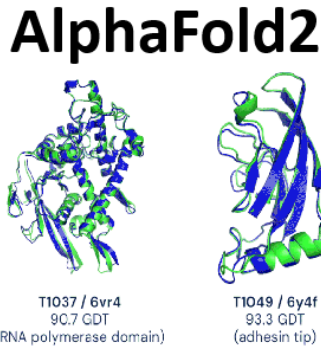


IDC预测：2020年以前，人类产生的数据量每两年翻一倍；
2026年前后，全球数据量将达到惊人的约220ZB
IDC Worldwide Global DataSphere and Global StorageSphere Structured and Unstructured Data Forecast, 2022–2026



20万个
已知结构

30万个
未知结构



● Experimental result
● Computational prediction

7 理论

加强科技基础能力建设

认真学习宣传贯彻党的二十大精神

- 2018-03 **《科学数据管理办法》**
加强全生命周期管理，强化法人主体责任，要求强制性汇交，提出科学数据中心建设
- 2019-06 **《关于进一步弘扬科学家精神加强作风和学风建设的意见》**
“论文等科研成果发表后1个月内，要将所涉及的实验记录、实验数据等原始数据资料交所在单位统一管理、留存备查”
- 2021-06 **《中华人民共和国数据安全法》**
- 2021-11 **《中华人民共和国科学技术进步法(2021年修订) 》**
加强开放共享，将“开放科学”写入
- 2022-12 **《中共中央、国务院关于构建数据基础制度更好发挥数据要素作用的意见》**
数据产权、流通交易、收益分配、安全治理，“活起来、动起来、用起来”

科技项目和论文关联数据的汇交与出版进入全面实践阶段

国家重点研发计划等已启动数据汇交

- 2018年12月,《**国家重点研发计划**项目综合绩效评价工作规范(试行)》
- 2019年7月,《科技部基础研究司 资源配置与管理司关于开展**科技基础性工作专项**项目综合绩效评价工作的通知》
- 2019年2月,《中国科学院科学数据管理与开放共享办法(试行)》明确提出科研项目数据汇交要求

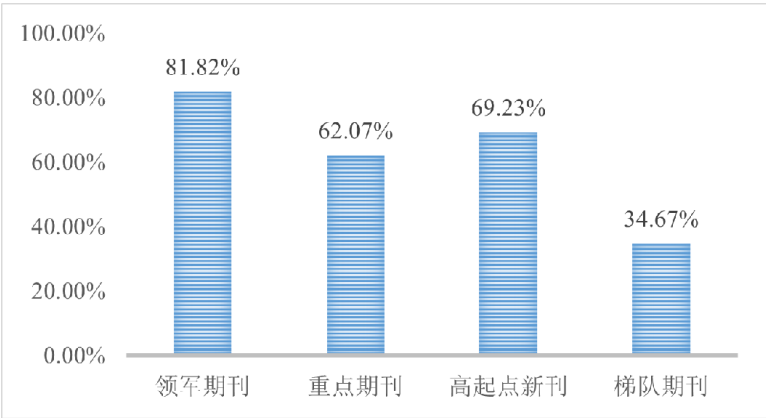
01 明确科研项目数据汇交要求



注 国际合作项目:中科院院属法人单位与国外/境外组织或个人的合作项目,在国外/境外所产生的科学数据,应通过该院属法人单位汇交。国外/境外组织或个人依托院属法人单位开展的合作项目,在中华人民共和国境内所产生的科学数据,应通过该院属法人单位汇交。

期刊出版对论文关联数据的要求

- 2022年4月,国家新闻出版署将期刊的数据汇交与管理情况纳入《2021年度期刊核验表》
- 2022年11月,发布《中国科协办公厅 中国科学院办公厅关于组织开展期刊论文关联数据汇交工作的通知》
- 2021年6月,中国科学院发布《中国科学院传播局关于进一步加强和规范论文关联数据管理的通知》



<http://ncdc.csdata.org/p/891/>

科学数据管理与服务面临新的挑战



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

更规范的数据管理

- 科研活动过程中数据管理活动的支撑
- 以科学数据中心等形式，提供常态化

更广泛的汇聚与开放

- 支持项目与论文数据的汇聚与共享
- 为更广泛的科技基础设施提供服务

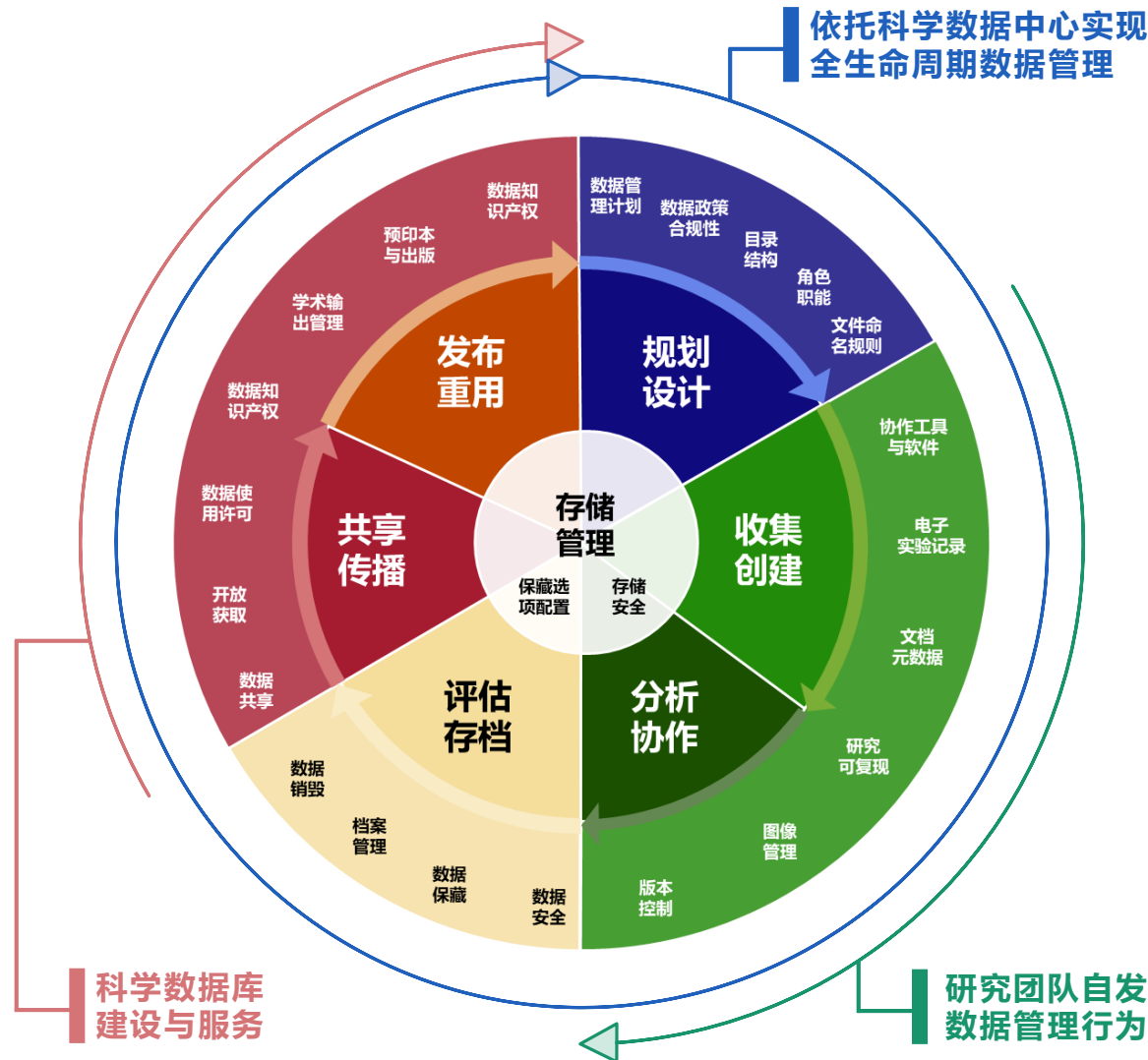
更高效的利用与流通

- 数据互联、跨域协同、智能调度
- 活起来、动起来、用起来

安全可信的治理与服务

- 分类分级管理与安全监测
- 用户权益保护、追溯、认证等要素化治理的机制与实践

依托科学数据中心实现
全生命周期数据管理



汇报提纲

一、背景与挑战

二、服务与实践

- 中国科学院科学数据中心体系于2019年启动建设
- 牵头院科学数据中心体系建设，承担院科学数据总中心
- 为1个总中心 + 18个学科中心 + 13个所级中心的中国科学数据中心体系提供支撑服务，推动高效运转
- 负责统筹协调及评价、通用支撑、公共服务、交叉应用、安全服务





跟踪与咨询

国际科学数据组织

生物

领域科学数据组织

物理

公共基础设施

地学

化学

材料

其他

ESGF

标准规范



技术工具与公共服务

Software Market

FindQta

Convergence Research Network

Repositories

institutes

general

disciplinary

Science Data Bank

Security

AAI

Monitoring

Registries

DataPid

CASDMP

Research Information catalogue

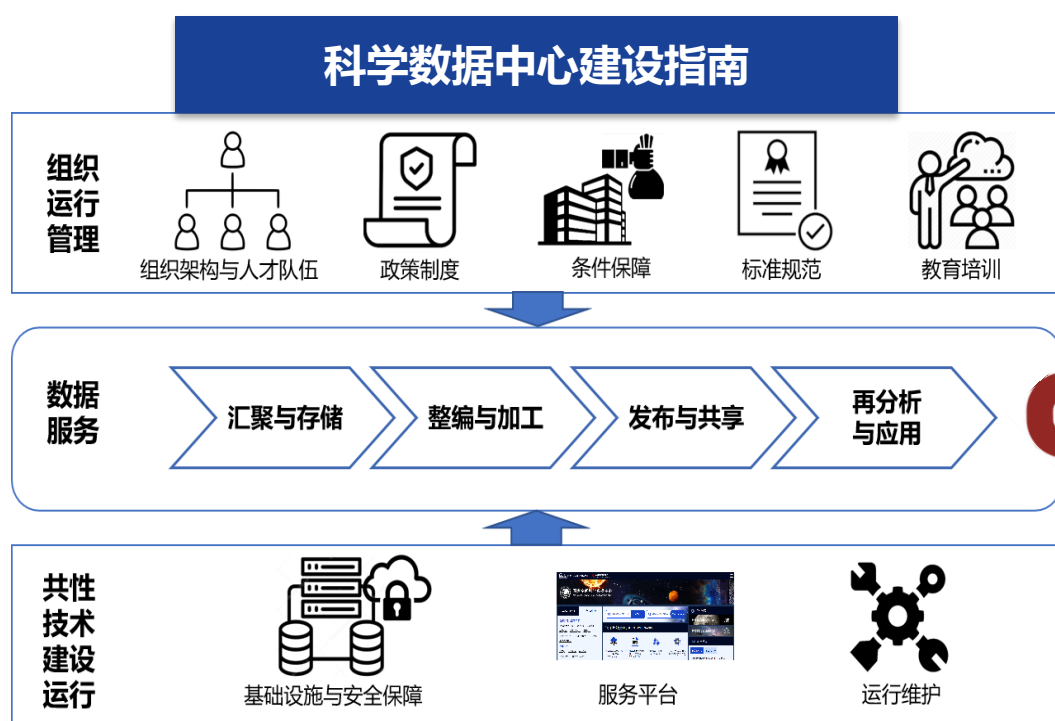
认定评价与运行服务



1.更规范的管理—科学数据中心建设指南及软件栈



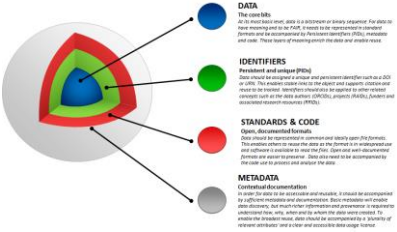
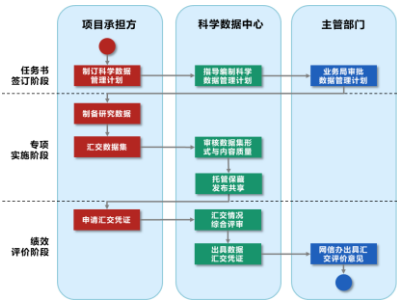
中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences



项目数据汇交

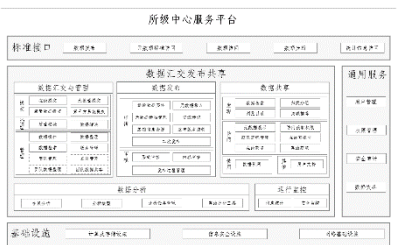
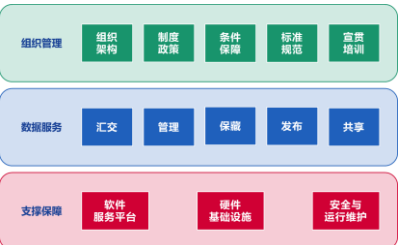
落实汇交管理各方责任，详细流程，主要阶段性成果的格式和填写要求

- 作者本人汇交
- maDMP
- FAIR数据对象
- 数据汇交凭证



数据中心建设

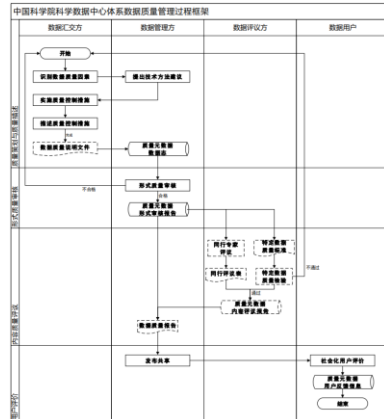
- 院数据中心能力体系3个维度，13个关键过程域，48项实施建议
- 数据中心IT系统参考方案
- 院数据中心体系互操作标准服务接口



数据质量管理

探索科学数据质量描述规范，汇交流程各阶段性成果的质量要求与校验规则

- 数据质量元数据
- 质量说明文件
- 参照论文审核流程设计
- 安全/伦理审核管理



数据共享

面向数据汇交与共享场景，规定过程中的访问权限和授权许可声明方式，数据引用、转载要求等

- 访问权限描述框架
- 开放协议+国产协议
- 明确数据引用支持
- 转载保护首发权益
- ma共享策略描述框架

数据集

推荐CC-BY 4.0，对方使用数据时应适当署名，如引用、致谢

已发表论文

不得改变原文的许可，出版社协议不允许再传播的也不得公开

软件

建议使用开源软件常用协议，如MIT、BSD、Apache、GPL等

专利、软著

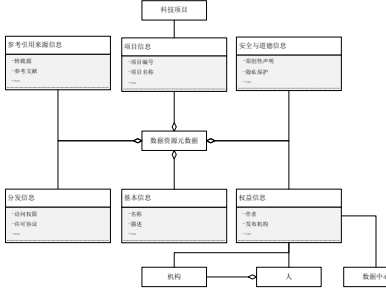
不涉及通过科学数据中心授权许可

图例	说明	格式	语言	是否开放	是否共享	是否引用	是否转载	是否许可	是否其他
CC-BY	CC-BY 4.0	CC-BY	CC-BY	CC-BY	CC-BY	CC-BY	CC-BY	CC-BY	CC-BY
CC-BY-NC	CC-BY-NC 4.0	CC-BY-NC	CC-BY-NC	CC-BY-NC	CC-BY-NC	CC-BY-NC	CC-BY-NC	CC-BY-NC	CC-BY-NC
CC-BY-ND	CC-BY-ND 4.0	CC-BY-ND	CC-BY-ND	CC-BY-ND	CC-BY-ND	CC-BY-ND	CC-BY-ND	CC-BY-ND	CC-BY-ND
CC-BY-NC-ND	CC-BY-NC-ND 4.0	CC-BY-NC-ND	CC-BY-NC-ND	CC-BY-NC-ND	CC-BY-NC-ND	CC-BY-NC-ND	CC-BY-NC-ND	CC-BY-NC-ND	CC-BY-NC-ND

元数据

发展面向汇交业务的元数据格式规范，形成元数据模板，元数据内容校验规则，元数据互操作语法和语义词表

- 已支持8类资源
- InstDB内置通用模板
- 面向汇交业务设计
- 各种ma可互操作



科学数据中心支撑软件-FairStack



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

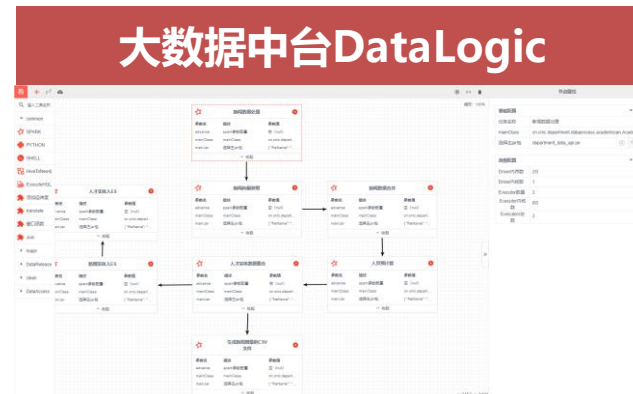
团队数据管理DataSpace



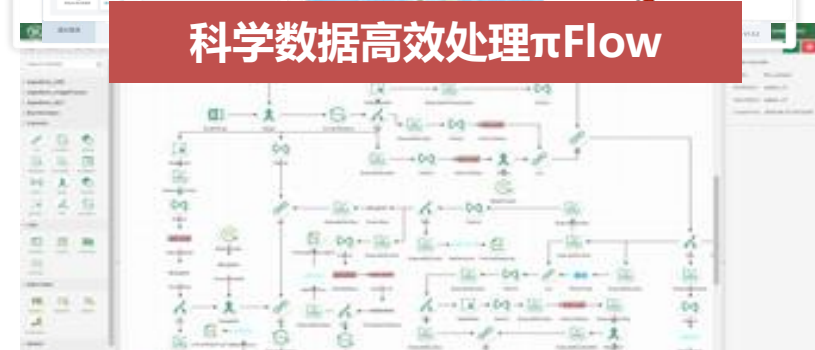
机构数据仓储InstDB



大数据中台DataLogic



科学数据高效处理 π Flow



科学数据交互式分析DataLab



科学数据质量管理软件DataQ



科学数据安全防护DPGuard

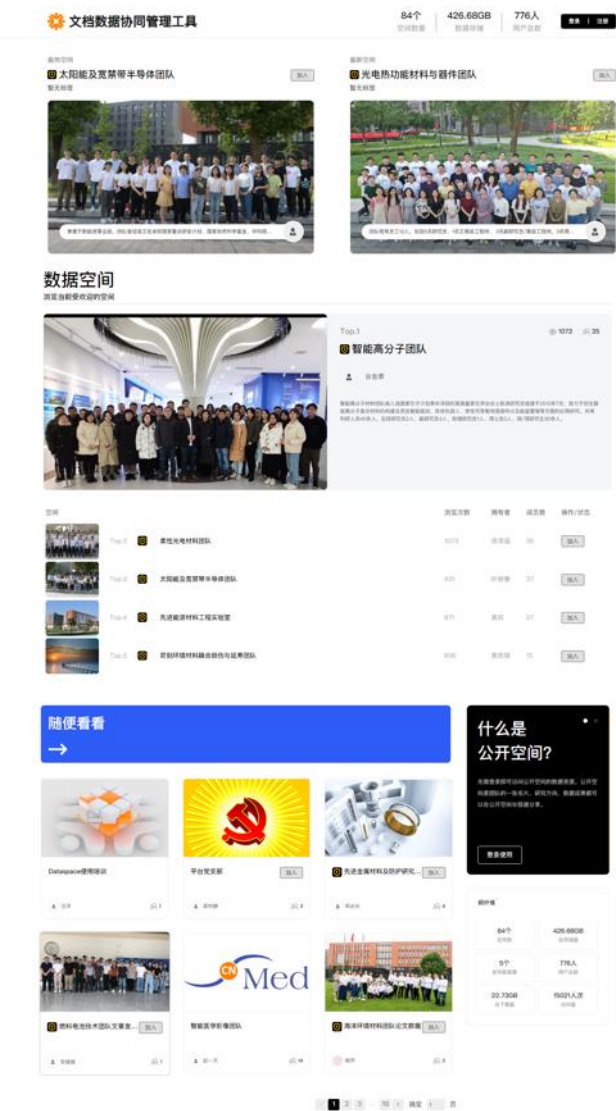


自主研发软件+组件，集成国际
开源、免费软件，形成科学数据
中心软件栈

<https://market.casdc.cn/>



软件栈支撑宁波材料所日常科研数据管理

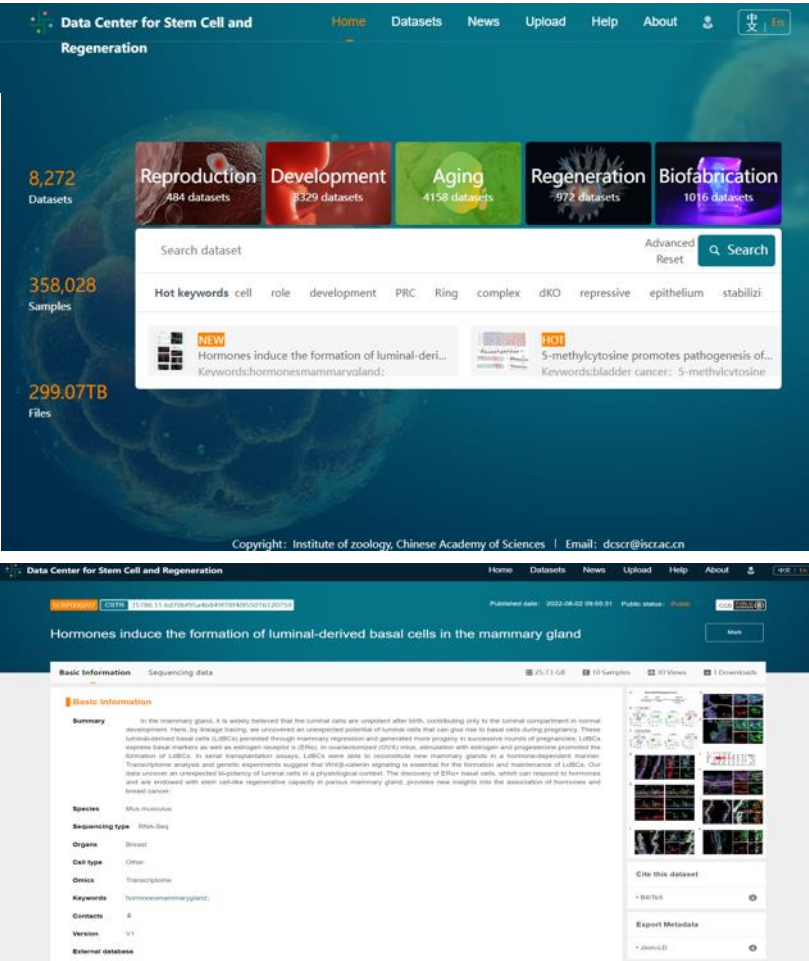


宁波材料所在DataSpace中日常科研数据的管理与共享，目前为全所科研人员、科研团队提供安全、便捷、高效的仓储式科研数据协同管理服务，截止目前已覆盖有机光学诊疗材料与技术团队、纳米生物材料团队、智能高分子材料团队、海洋环境材料团队、先进功能膜等团队**1000+个**科研用户和**800GB+**科研数据，助力宁波材料所领域**科学数据的高效管理**，为科学数据发挥最大的价值提供基础支撑。



软件栈支撑干细胞与再生医学数据中心建设

科学数据中心软件栈软件工具支撑干细胞与再生医学数据中心建设，已实现汇聚/汇交领域高通量数据**700TB**，支撑“器官重建与制造”与“干细胞与再生医学研究”两期先导专项数据的在线汇交与管理，服务来自全球**18个国家19万余次**的访问。



DataLab支撑空间数据中心构建领域分析云平台 iSpace3.0



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

科学数据中心软件栈软件工具DataLab支撑国家空间科学数据中心构建领域分析云平台ispace3.0，实现空间天气事件模型算法的在线集成和分析服务，目前已完成**5类空间天气**算法模型的集成，有效提升空间天气领域科研效率，提供领域科学分析新方法，探索空间天气领域科研范式变革路径。

首页 分析工具 实验 交互式编程 管理配置 个人中心 凌瑞 | 退出

空间天气传播链典型事件知识挖掘
“大数据+人工智能”
赋能空间天气领域科研创新

致力于融合空间天气科学“数据+工具”，基于“大数据+AI”技术提供空间天气典型事件识别、特征提取、关联分析与智能模型计算，探索领域科研范式变革新途径。

开始

版权:
Copyright © 中国科学院计算机网络信息中心
备案号:
京ICP备05064604号-28
京公网安备11010502044069号

Powered by DataLab v1.0.4

NSSC

返回列表 删除 分享 发布

实验步骤 添加

已选组件: 全域紫外极光演化与预测模型

我的数据 实验数据 来自分享 公开数据

新的实验步骤(2)
极区极光识别与特征提取工具

新的实验步骤(3)
等离子体形态分割与特征提取工具

新的实验步骤(4)
地磁扰动 (Pc1-2) 事件自动识别工具

新的实验步骤(5)
地磁扰动 (Pc3-5) 事件自动识别工具

新的实验步骤(6)
地磁扰动期间自动分类工具

新的实验步骤(7)
全域紫外极光演化与预测模型

概述:
Polar/UVI极光图像序列预测。本工具利用ConvLSTM网络对Polar/UVI极光图像序列进行学习并提取特征，生成其在未来时间的图像序列预测图像。输入为前序时间段经过预处理的mpx文件，输出为若干长度的极光预测图像序列

展开详情

src_path
15

dst_path
AuroraPredictionTestOutput_0920

out_len
10

运行成功
任务ID: 6d1b9f25530448e4914af12cd1
详细日志
==== DataLab Start Function ====
==== DataLab Start Function ====
==== DataLab Start Function ====

删除此步骤 保存 执行

名称 输入要搜索的内容

AuroraPredictionData
seg_flip_data
SpectrumCisBicDDData
Pc345Data
Pc1-2Data
SegPlasmaBubbleData
ArcDetectionData
example
data
Test_plasma_bubble

上传本地数据 通过FairLink链接导入数据

16

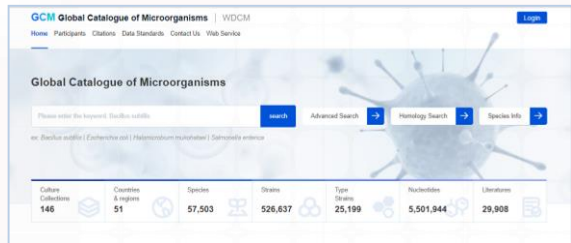
2.更广泛的资源汇聚与开放



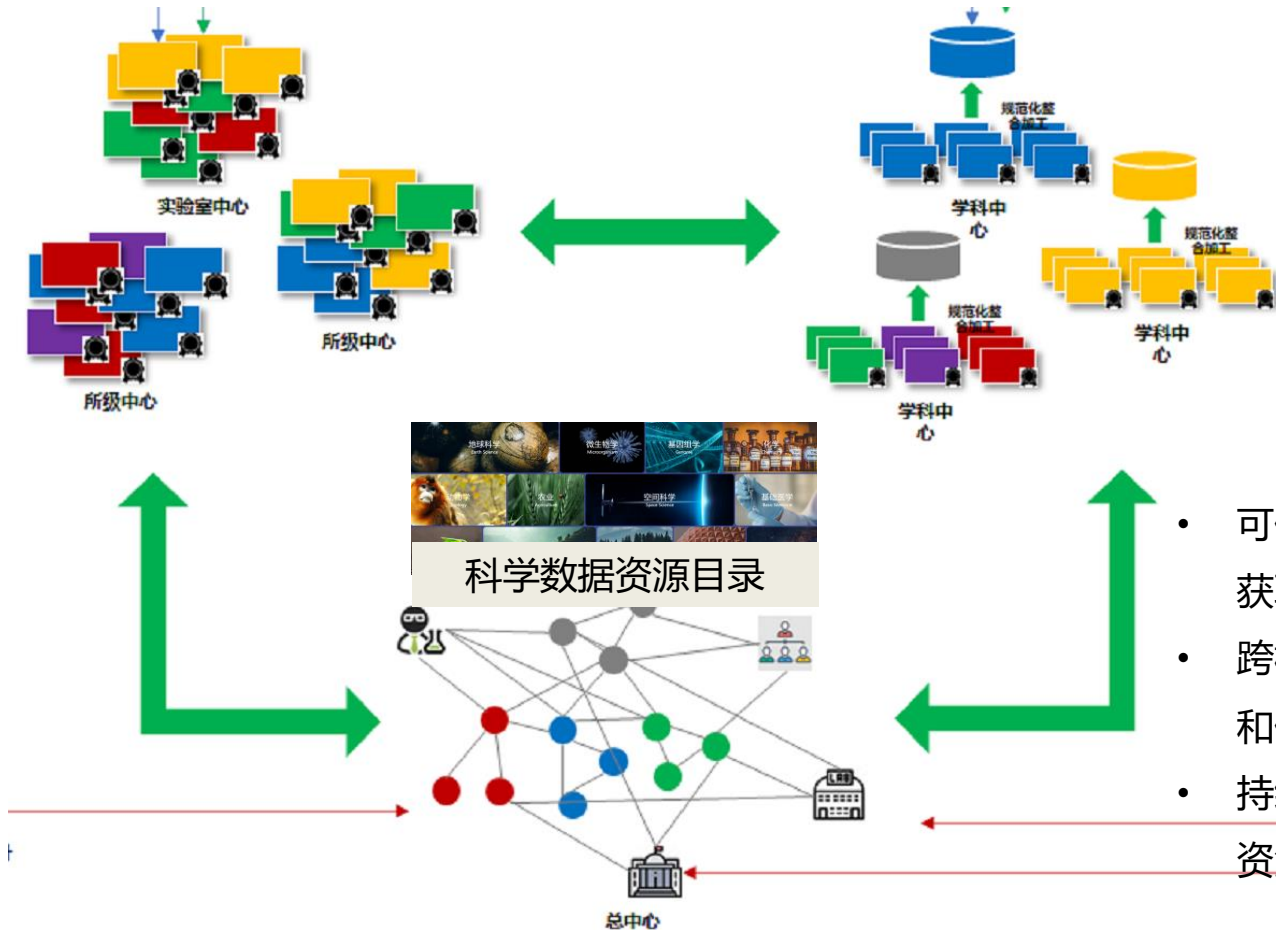
统筹管理的项目数据汇交



具有国际影响力的论文关联数据汇交和出版



权威科学数据建设



DMP平台统筹项目科学数据汇交管理

面向科技项目数据汇交的全面推进，构建科技项目数据汇交管理平台

- 强化统筹管理和全流程服务
- 采用集约化方案减少重复建设

- 项目承担方

统一入口和汇交指导
- 科学数据中心

云化汇交管理平台
- 项目管理机构

全面掌握汇交进度和数据成果
- 平台管理机构

数据整体情况、数据中心评价依据

统一入口、统筹管理、进度跟踪、
全维度统计、DMP可量化



快速构建项目数据汇交支持能力



以某专项为例

ScienceDB助力论文关联数据出版

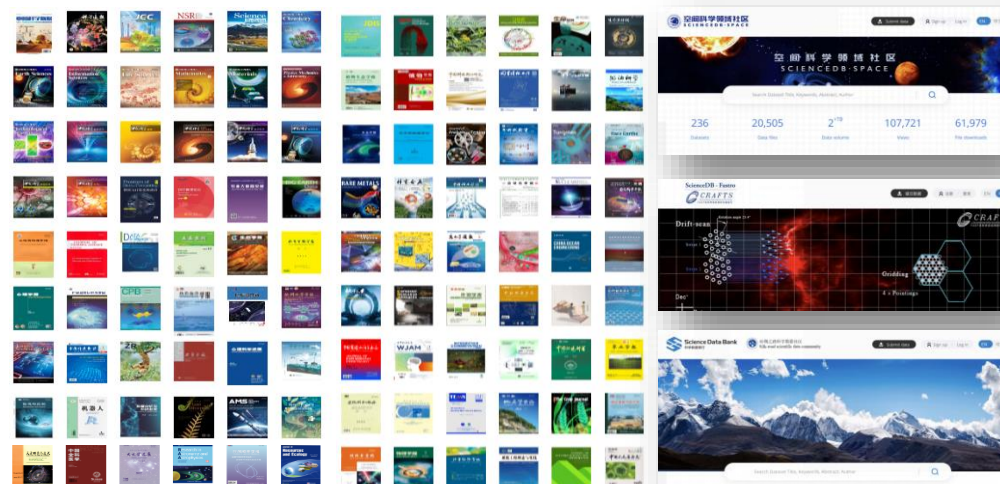


中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

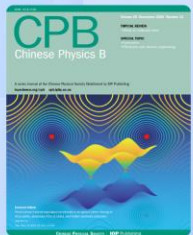
- 科学数据银行ScienceDB获得国际主流期刊认可，被DCI等平台索引
- 发展数据社区模式，打造具有国际影响力的数据出版品牌



- 获SPRING NATURE/WILEY/ELSEVIER等多家国际顶尖出版商认可
- 获Data Citation Index/Google Dataset Search/BASE等多家国际知名数据库收录
- 已为全球190多个国家和地区科研人员服务，访问超过8亿次



- 探索**多元化数据社区**发展模式，**服务数据中心、大装置**等多种需求场景，加强数据共享生态建设
- 已构建论文关联数据社区285个
 - 数据中心：19家国家科学数据中心
 - 大装置、重大科研创新团队：FAST | 寰（&大气物理所）
 - 重点国际区域：“丝路数据社区”（2023重点在乌兹别克斯坦）
 - 百余种科技期刊、出版社



中国物理B Chinese Physics B

Chinese Physics B (formerly Chinese Physics), sponsored by Chinese Physical Society and Institute of Physics, Chinese Academy of Sciences, is one of the most prestigious periodicals published in China, and ranks the top in Chinese physics journals. It publishes "original research papers", "rapid communications", "reviews", and "data papers" on the latest developments and achievements in all branches of physics worldwide except nuclear physics and physics of elementary particles and fields. Subject coverage includes: condensed matter and materials physics; atomic, molecular, and optical physics; statistical, nonlinear, and soft matter physics; plasma physics; interdisciplinary physics. Chinese physics B is collaborating with the IOP Publishing (IOPP, UK) in distributing the journal worldwide. Chinese Physics B is among the Chinese journals first indexed by SCI and currently incorporated in major international indexing systems, such as SCI, EI, Scopus, CA, INSPEC, JICST, AJ, CSTPCD, CSCD, etc.

输入作者，关键字，标题等检索数据集



44 条检索结果

排序
出版时间降序

ScienceDB·Space

Log in

数据提交

中文

空间科学领域社区
SPACE SCIENCE ARTICLE DATA REPOSITORY



空间科学领域社区
SCIENCEDB·SPACE

315

Datasets

23,108

Data files

2+TB

Data volume

34,545,137

Views

71,153

Downloads

Chinese Medical Association Publishing House Data Bank

Log in

数据提交

中文

开放可信 共治共享

Open Medical Data for Better World

中华医学会杂志社科学数据存储库邀您共享医学数据，引领
开放科学新气象，共建医学科研共同体



中华医学会杂志社
科学数据存储库

Co-constructed by CMA Publishing House and Science Data Bank

22

Datasets

38

Data files

550+MB

Data volume

14,300

Views

69

Downloads

Feedback

从内部管理到开放共享，无缝衔接！ScienceDB东北地理所社区实现研究成果快速发布共享

ScienceDB 科学数据银行 2023-08-16 10:12 发表于北京

科研数据是科学研究的重要成果之一，通过良好的数据管理，有助于实现数据共享和再利用，使其成为新的科学研究的基础。[1]

科学数据银行（ScienceDB）通过打通与数据协同管理工具DataSpace的数据管理和分享路径，贯通机构科研数据从内部协同管理到对外开放共享的链路，实现研究成果的快速发布共享，并在ScienceDB中国科学院东北地理与农业生态研究所数据社区（<https://www.scidb.cn/agdc>）应用实践，目前发布数据125个。

中国科学院东北地理与农业生态研究所

Log in

数据提交

EN



Science Data Bank for
中国科学院东北地理与农业生态研究所
Northeast Institute of Geography and Agroecology, CAS

125

发表数据集

107

数据集

7+GB

数据体量

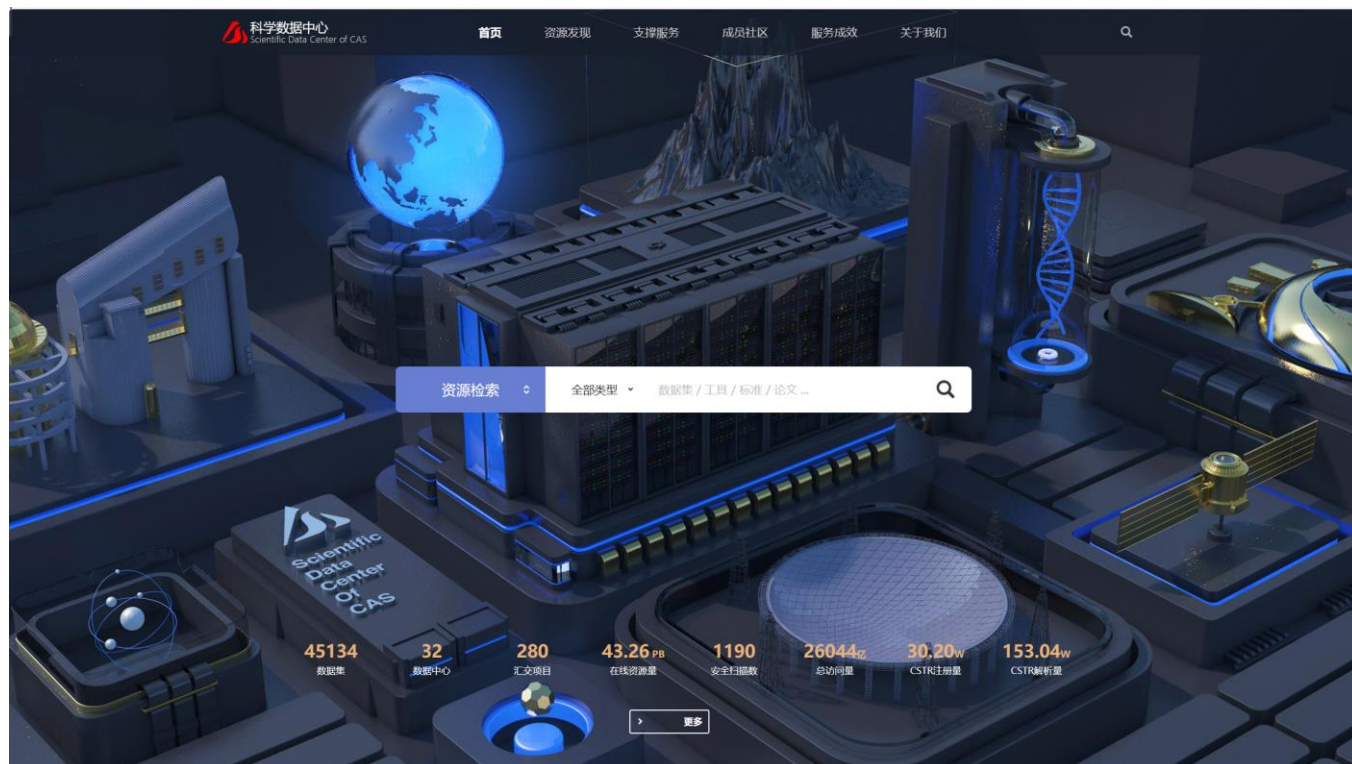
11,275

数据集访问次数

1,843

数据集下载人次

中国科学院科学数据中心体系服务门户，是全院科学数据资源的一体化服务的统一入口，实现科学数据相关各类资源的统一发现与服务。



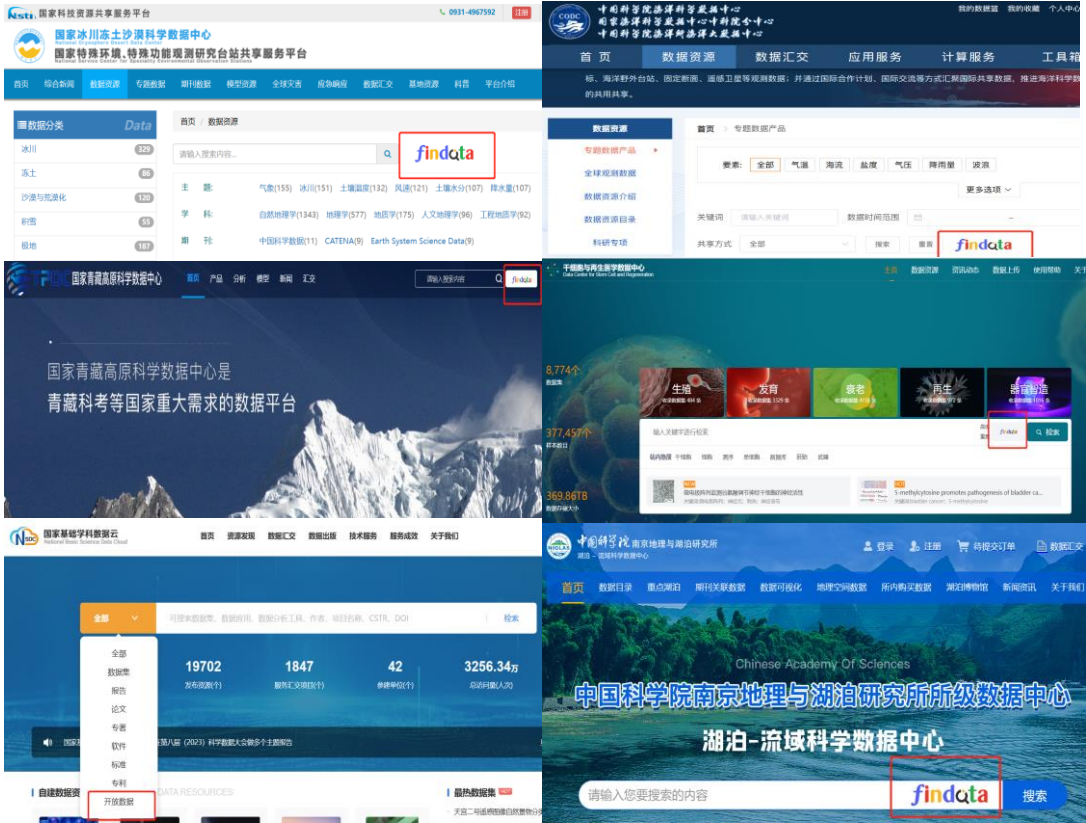
- 集成了数据集，数据工具，标准，论文和著作5类主要科学数据资源，累计集成各类资源元数据45,134条
- 集成各类公共支撑服务，覆盖科学数据管理、资源可信发布与共享全流程
- 提供科学数据中心运行服务情况大屏

Findata提供科学数据发现服务

findata



试试 *Actively Repeating FRB 20190520B* 或 *ASTER GDEM 30M 高程数据*



>220万访问量



>4.6万个独立IP数



106个访问国家数

为数据中心提供科学数据搜索服务
2个国家级 18个院级 2个院内其他平台

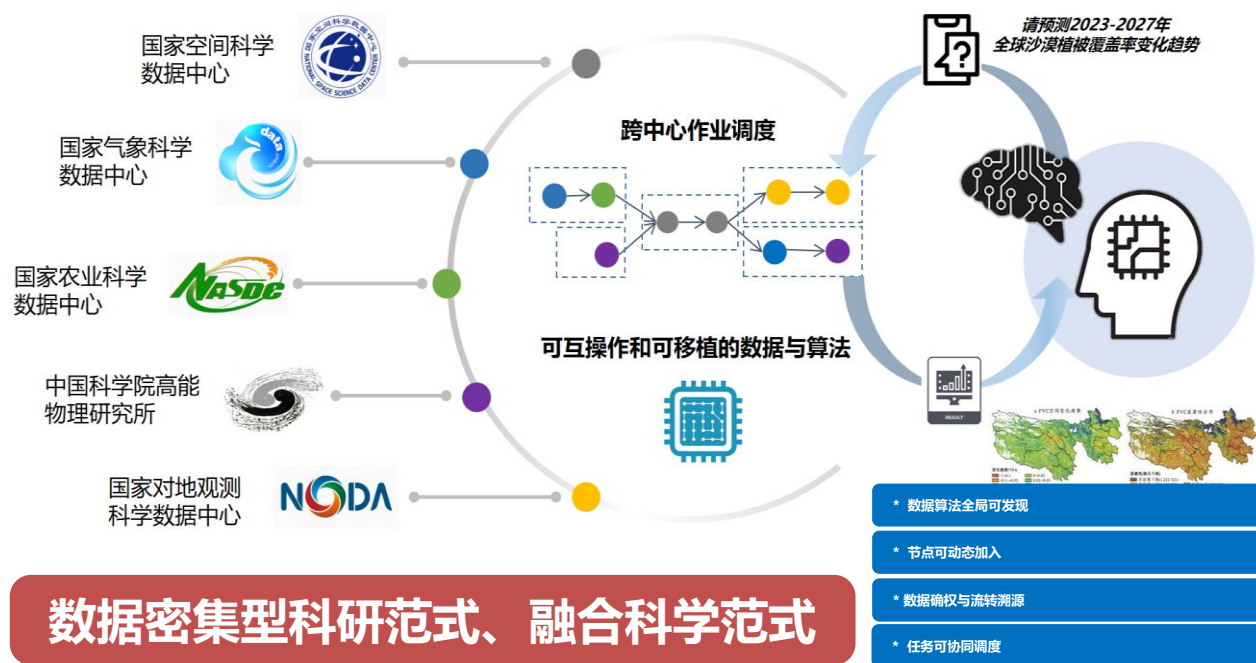
3.更高效的利用与流通



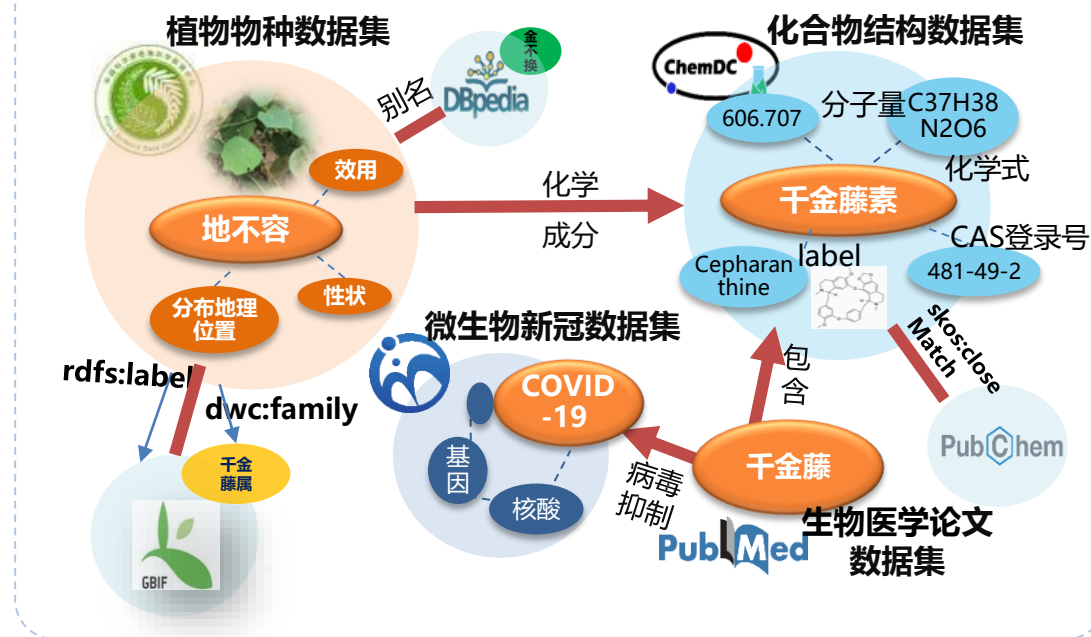
中国科学院计算机网络信息中心
Computer Network Information Center,
Chinese Academy of Sciences

面向新型科研范式需求，建设分布式科学数据资源协同调度网络系统

数据协同：分布式异构数据高效供给



数据语义：科学数据语义集成与互操作



- ◆ 科学数据协同调度技术：整合8个数据中心30个服务，加速生态、科考等数据的分布式采集、质控、治理利用
- ◆ 科学数据语义技术：整合了8个数据机构的29个数据集，形成9.9千万个语义三元组、4.1千万个语义链接
- ◆ 支撑国家重点研发计划项目7项

① 多元异构科学数据融合管理

- 提出复杂结构数据集原位细粒度关联组织机制、**千亿级异构数据**弹性存储技术
- 组合查询较Neo4j快3-4倍**



② 可溯源配置化大数据流水线

- 发布3个版本，性能达到 **1000w/s**数据处理能力
- 较Apache NiFi快3倍以上

③ 低延时高可靠遥感数据高效分析

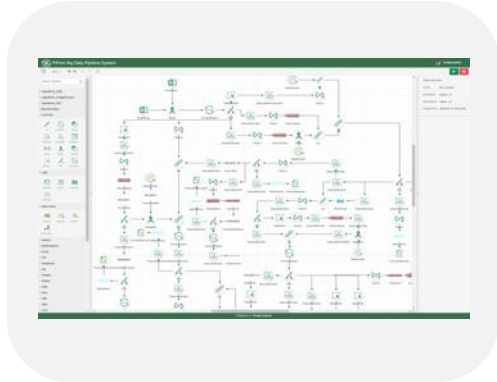
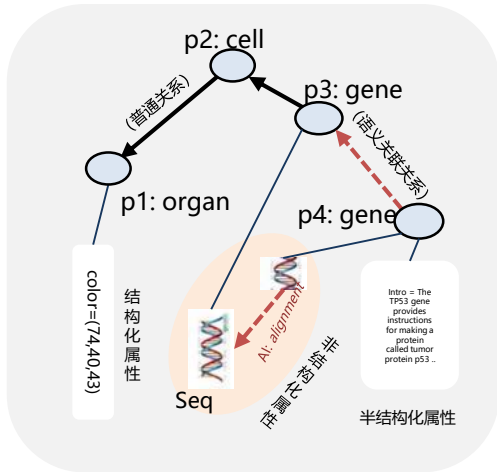
- DATABANK大规模数据解析入库、数据在线分析服务
- 在空天院部署格网数据引擎GeoDataBox

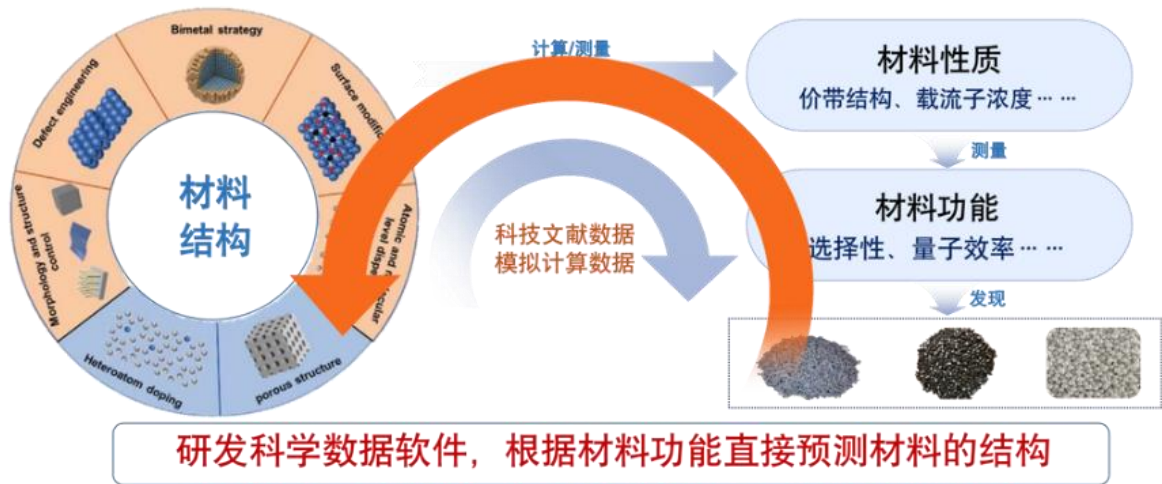
④ 跨域分布式数据协同调度

- 提出了跨节点科学数据流调度框架
- 实现了移动计算、不移动数据的**启发式跨中心调度策略**

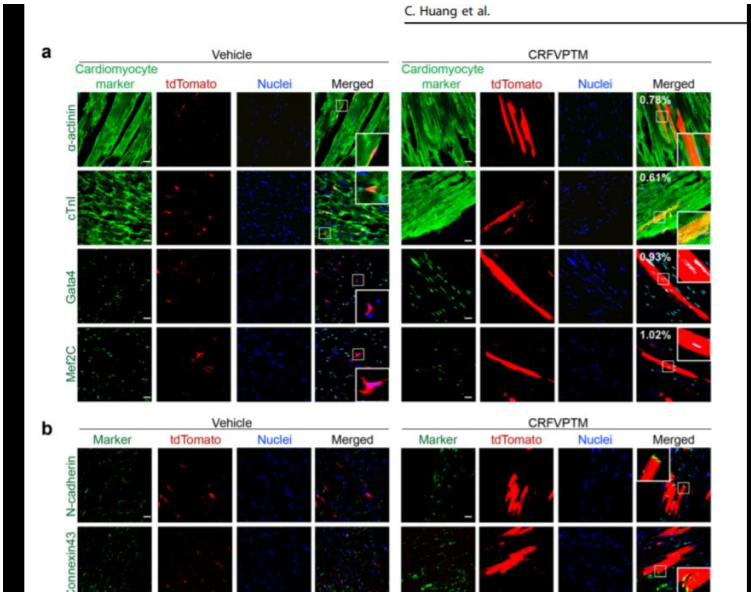
⑤ 跨模态知识挖掘与融合

- 突破基于远程监督的跨模态知识抽取、**基于主动学习**的模型增量训练等技术
- 在科技、军事、医疗健康领域应用





国家重点研发计划青年科学家项目：联合国家纳米中心



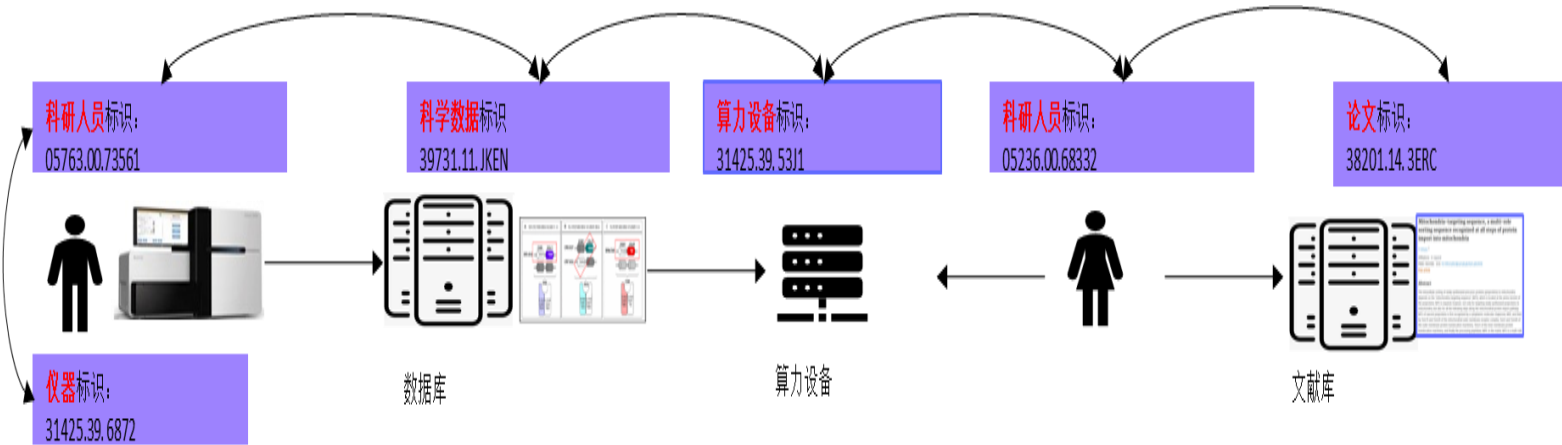
支撑动物所生命模拟创新应用

4. 安全可信的治理与服务

发展科学数据确权、发现、溯源、认证等公共基础服务，促进数据流通

科技资源标识服务

- 建立自主科技资源标识服务平台 CSTR，为科技资源分配“身份**证**”，支持科技资源溯源，推动数**据确权**



科技资源标识服务平台
CSTR Identification Platform

标识服务 国际合作 关于我们 登录 EN 中文

全球科技资源标识，开放共享基础设施

学术资源 关键词: 作者

2.4亿 学术论文

20万 学位论文

819万 科学数据

1.8万 图书

28万 专利

13.8万 物种名录

1.3万 仪器设备

60万 科研项目

17万 科技资源

10万 科研机构

124万 科研人员

中国科学院冰川冻土沙漠科学数据中心
CSTR标识: 11738

注册量: 2,443 次

总排名: 12/233

解析量: 78,608 次

总排名: 1/233

引用量TOP

来源国TOP

中国科学院计算机网络信息中心
CSTR标识: 15406

注册量: 3,770 次

总排名: 10/225

解析量: 77,369 次

总排名: 2/233

引用量TOP

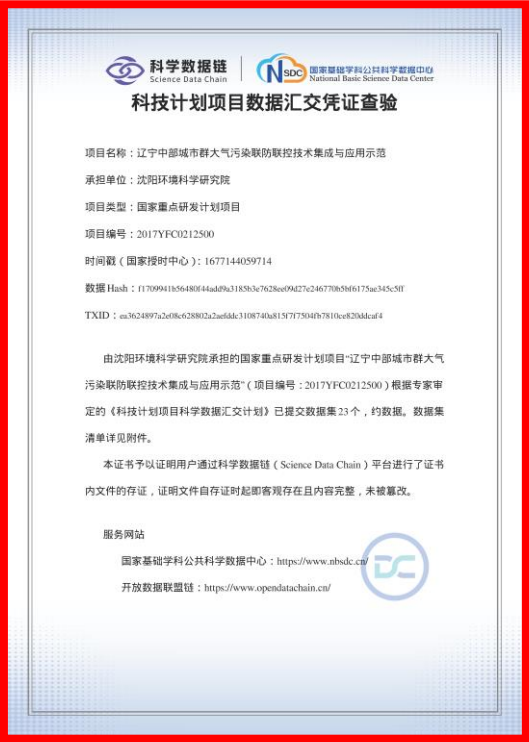
来源国TOP

科链支撑科学数据的存证、确权 and 溯源

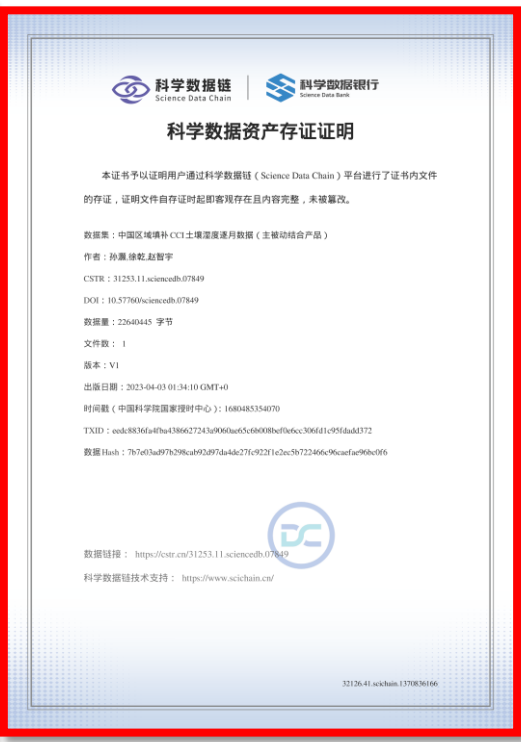
发展科学数据确权、发现、溯源、认证等公共基础服务，促进数据流通

科学数据区块链服务

- 建设科学数据链SciChain，支持科学数据的存证与追溯
- 科学数据领域唯一获得网信办备案的区块链服务



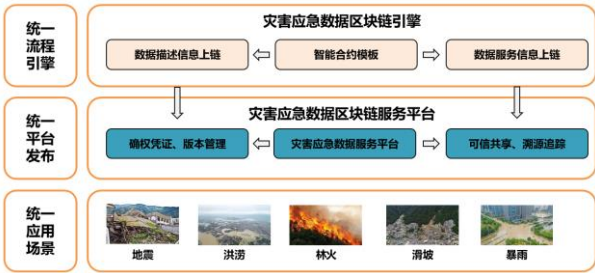
存证服务



数据归属或者首发



微生物所隐私计算证明



发展科学数据确权、发现、溯源、认证等公共基础服务，促进数据流通

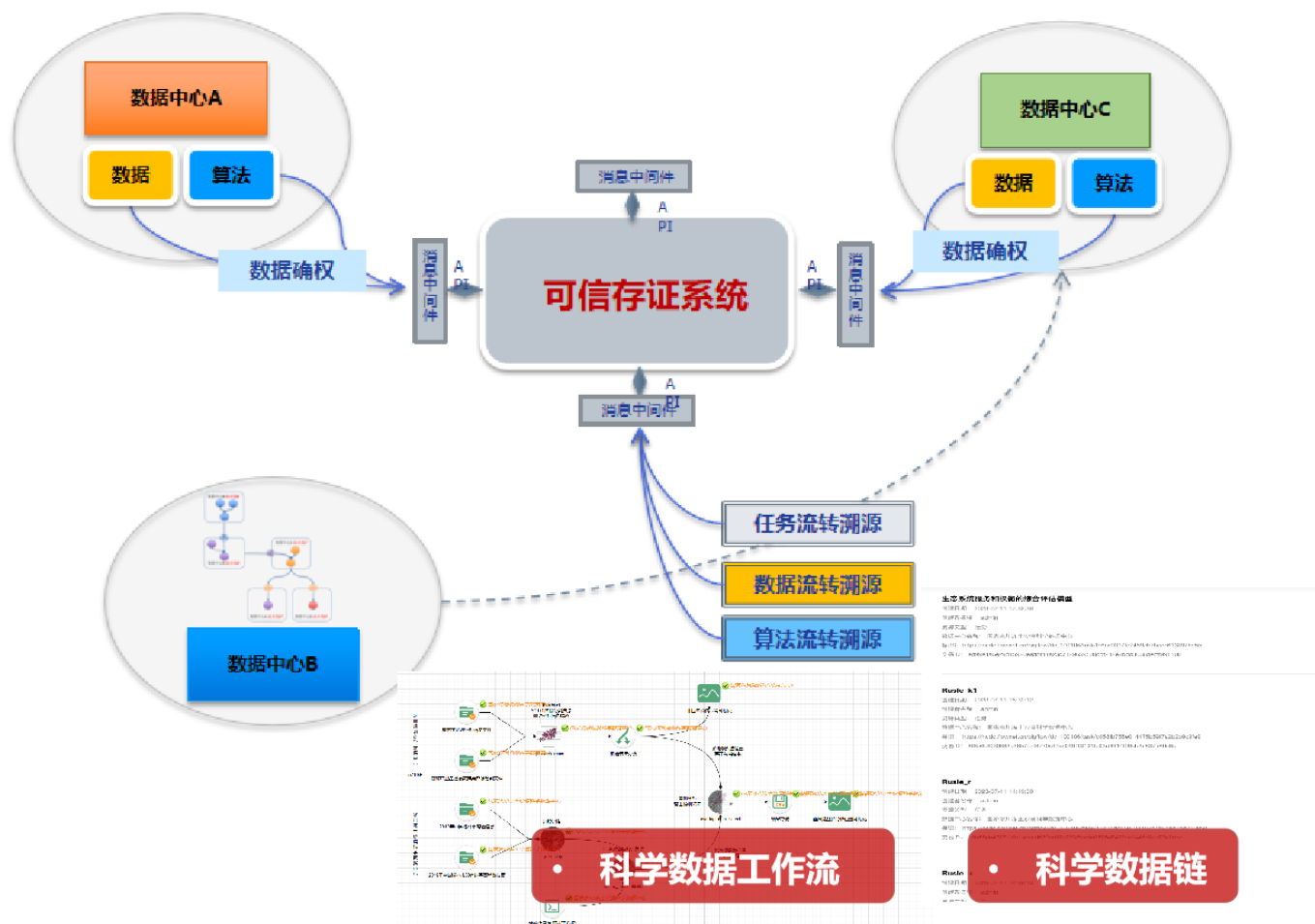


科学数据认证服务

- “注册-指南-认证-监测”机制的实践经验为基础，面向科学数据中心、存储库、知识库等科学数据服务平台，提供注册、认证等多层级服务
- 为科学数据管理平台提供可信评估与改进建议
- 为科研人员快速发现高质量科学数据服务平台提供依据



基于隐私计算、区块链等技术，探索“数据可用不可见”，提升跨机构数据利用能力



- 基于科学数据工作流和区块链技术，探索跨中心的协同分析框架
- 已在生态观测台站数据的采集与处理中得到应用

支撑重大项目的数据管理



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

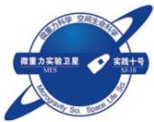
空间科学领域重大工程

运行保障

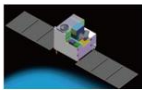
建设中



暗物质粒子探测卫星
2015.12.17至今
(在轨运行时间: 2834天)



实践十号科学实验卫星
2016.4.6~4.26
(在轨运行时间: 21天)



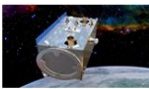
量子科学实验卫星
2016.8.16至今
(在轨运行时间: 2591天)



硬X射线谱仪卫星
2017.6.15至今
(在轨运行时间: 2288天)



天绘二号01组卫星
2019.4.30至今
(在轨运行时间: 1604天)



太极一号卫星
2019.8.31至今
(在轨运行时间: 1481天)



引力波暴高能电磁对应
天体全天监测器
2020.12.10至今
(在轨运行时间: 1014天)



天绘二号02组卫星
2021.8.19至今
(在轨运行时间: 762天)



先进天基太阳天文台
2022.10.9至今
(在轨运行时间: 346天)



中法天文卫星 (SVOM)
预计2023.12发射

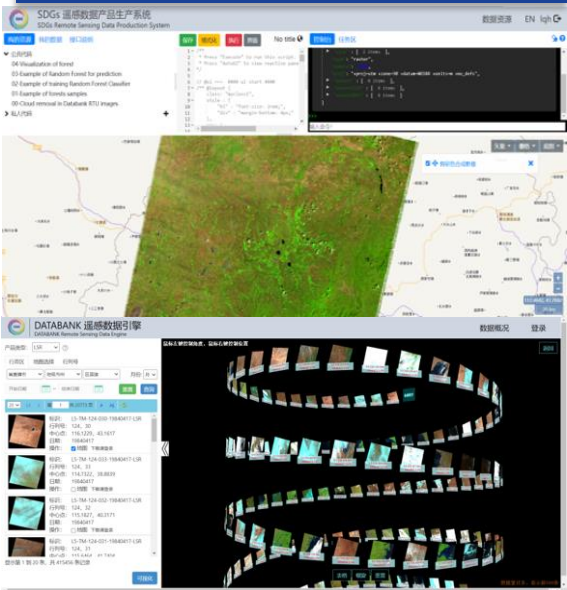


爱因斯坦探针 (EP)
预计2023.11发射

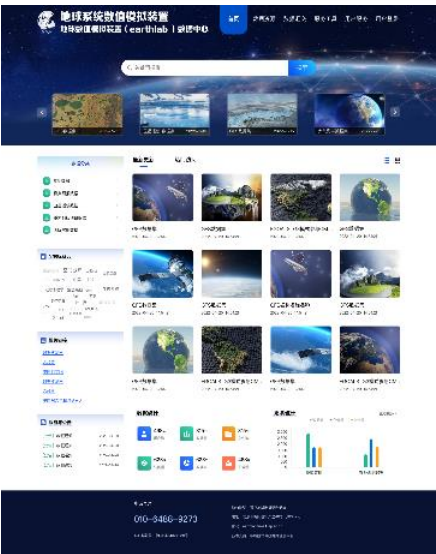


太阳风磁层相互作用全
景成像卫星 (SMILE)
预计2024.12发射

地球大数据科学工程



地球系统数值模拟大装置





科学数据中建设指南与管理规范
科学数据中心建设指南
科学数据质量管理、发布共享指南
元数据内容与汇聚要求
元数据规范技术实现

科学数据管理支撑工具/公共服务平台介绍与实操
科学数据中心建设支撑工具介绍
CSTR科技资源标识服务平台
论文关联数据出版与数据社区构建
科学数据关联与应用
端到端的科学数据协同分析框架

科学数据管理典型业务场景类
先导专项项目科学数据汇交管理要求
先导专项项目科学数据汇交数据治理与元数据汇聚要求
数据论文写作与发表
图数据库技术与应用
大模型技术在科学数据分析中的探索应用



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

感谢聆听，欢迎关注



科苑数据公众号