中国科学院国家天文台
NATIONAL ASTRONOMICAL OBSERVATORIES, CAS

NADC National Astronomical Data Center 国家天文科学数据中心

China-VO

天文科研云平台

李长华

国家天文科学数据中心（NADC)

中国科学院国家天文台

主要内容

- 天文科学平台建设背景
- 天文科学平台发展趋势
- 天文科研云平台进展
- 天文科研云平台应用案例

主要内容

National Astronomical Data Center
国家天文科学数据中心

- **天文科学平台建设背景**
- **天文科学平台发展趋势**
- **天文科研云平台进展**
- **天文科研云平台应用案例**

天文数据





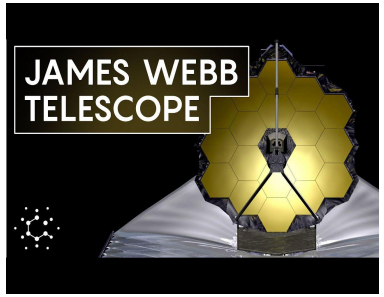


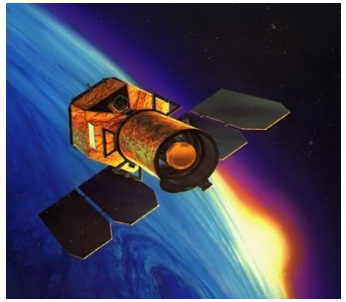


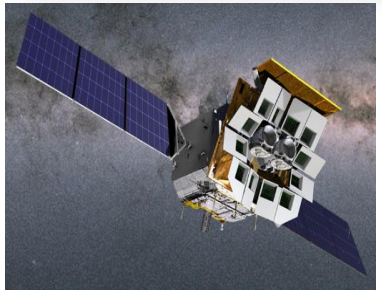






















VO & IVOA

- Virtual Observatory (VO) is a data-intensively online astronomical research and education environment, taking advantages of advanced information technologies to achieve seamless, global access to astronomical information.
- The Virtual Observatory (VO) aims to provide a research environment that will open up new possibilities for scientific research based on data discovery, efficient data access, and interoperability (FAIR).
- International Virtual Observatory Alliance



VO & IVOA

- **VO Applications for Astronomers**
  - IVOA Applications Working Group
- **VO Standards**

Technical Specifications

Group	Title	Most stable
App	SAMP - Simple Application Messaging Protocol	1.3
	VOTable - VOTable Format Definition	1.5
	MOC - HEALPix Multi-Order Coverage Map	2.0
	HiPS - Hierarchical Progressive Survey	1.0
DAL	DALI - Data Access Layer Interface	1.1
	DataLink	1.1
	Simple Cone Search	1.03
	SIA - Simple Image Access	2.0
	SLAP - Simple Line Access	1.0
	SSA - Simple Spectral Access	1.1
	STC-S: Space-Time Coordinate Metadata Linear String Implementation	1.0
	TAP - Table Access Protocol	1.1
	TAPRegExt - A VOResource Schema Extension for Describing TAP Services	1.0
	ADQL - Astronomical Data Query Language	2.1
	SimDAL - Simulation Data Access Layer	1.0
	VOEvent Transport Protocol	2.00
	SODA - Server-side Operations for Data Access	1.0
	Object Observability Simple Access Protocol	
	EPN-TAP: Publishing Solar System Data to the Virtual Observatory	2.0
	LineTAP: IVOA Relational Model for Spectral Lines	1.0
DaM	PHOTDM - Photometry Data Model	1.1
	SimDM - Simulation Data Model	1.0
	STC - Space-Time Coordinate Metadata for the Virtual Observatory	1.33
	Data Model for Astronomical DataSet Characterisation	1.13
	SSLDM - Simple Spectral Lines Data Model	1.0
	SpectralDM - IVOA Spectral Data Model	1.2
	ObsCore - Observation Data Model Core Components and its Implementation in the Table Access Protocol	1.1
	VODML - A Consistent Modeling Language for IVOA Data Models	1.0
	DatasetDM - Dataset Metadata Model	
	CubeDM - N-Dimensional Cube/Image Model	
	ProvenanceDM - Provenance Data Model	1.0
	Astronomical Coordinates and Coordinate Systems	1.0
	WCS Transform Model	
	Meas - Astronomical Measurements Model	1.0
	Observation Locator Table Access Protocol	1.0
	MIVOT - Model Instances in VOTable	1.0
	ObsCoreExtensionForRadioData - IVOA Obscore Extension for Radio data	
	ObsCoreTimeExtension - IVOA ObsCore Metadata Extension for Time Properties	
	CAOM - IVOA Common Archive Observation Model	
	ProposalDM - A Data model for Observation Proposals	
	MANGO: Metadata ANnotation for Generic Objects	
GWS	PDL - Parameter Description Language	1.0
	SSO - Single-Sign-On Profile: Authentication Mechanisms	2.0
	VOSpace service specification	2.1
	Credential Delegation Protocol	1.0
	UWS - Universal Worker Service	1.1
	VOSI - IVOA Support Interfaces	1.1
	GMS - Group Membership Service	1.0
ReR	IVOA Identifiers	2.0
	IVOA Registry Interfaces	1.1
	RM - Resource Metadata for the Virtual Observatory	1.12
	StandardsRegExt: a VOResource Schema Extension for Describing IVOA Standards	1.0
	SimpleDALRegExt - Describing Simple Data Access Services	1.2
	VOResource - an XML Encoding Schema for Resource Metadata	1.2
	VODataService - A VOResource Schema Extension for Describing Collections and Services	1.2
	RegTAP - Registry Relational Schema	1.2
	DocRegExt - Educational Resources in the VO	
Semantics	VOUnits - Units in the VO	1.1
	UCD - An IVOA standard for Unified Content Descriptors	1.10
	UCD1+ Controlled Vocabulary version <= 1.3: recommendation Version >= 1.4: endorsed note	1.6
	Maintenance of the list of UCD words	2.0
	Vocabularies in the Virtual Observatory	2.1
SDP	DocStd - IVOA Document Standards	2.0
VOE	VOEvent - Sky Event Reporting Metadata (VOEvent)	2.0
	VOEventRegExt - An XML Encoding Schema for Resource Metatdata for Collections of Events	



https://ivoa.net

Applications (in alphabetical order)
3D View
Aladin
APERICubes
AppLauncher
CASSIS
CDS Xmatch Service
Data Discovery Tool
Filter Profile Service
Firefly
Iris
Montage
Octet
SkyView
Specview
SPLAT
SVO Discovery Tool
TAPHandle
TAPsh
TESELA
TOPCAT/STILTS
VESPA
VisIVO
VOConvert
VODesktop
VOPlot
VOSA
VOSED
VOServices (Footprint, Spectrum, Filters, ...)
VOSpec

Functionality
Search for Images:
Aladin, Datascope, SkyView, SVO Discovery Tool, VODesktop, Data Discovery Tool, Firefly
Search for Spectra:
Aladin, CASSIS, Datascope, SPLAT, Specview, SVO Discovery Tool, VOServices, VOSpec, Data Discovery Tool
Search for Catalogues:
Aladin, Datascope, SVO Discovery Tool, TOPCAT, VODesktop, Data Discovery Tool, Firefly
Search for Time Series Data:
Time Series Search Tool, Firefly
Image visualisation:
Aladin, SkyView, Firefly
Spectra visualisation:
CASSIS, SPLAT, Specview, VOServices, VOSpec
Catalogues visualisation:
Aladin, TOPCAT, VOPlot, Firefly
Cross-correlation:
Aladin, STILTs,TOPCAT, CDS Xmatch Service, Cross Comparison Tool
Scatter, 3D plots and histograms:
TOPCAT, VOPlot, Firefly
Statistics:
AstroStat
Footprint Service:
Aladin, VOServices
Table format conversion:
TOPCAT, VOConvert, STILTS
Filter curves:
VOServices, Filter Profile Service
SED building:
VOSA, VOSED, VOSpec, Iris
Fixing WCS:
Aladin, WCSFixer
Query Databases
Seleste, TOPCAT, TAPHandle, TAPsh, Firefly

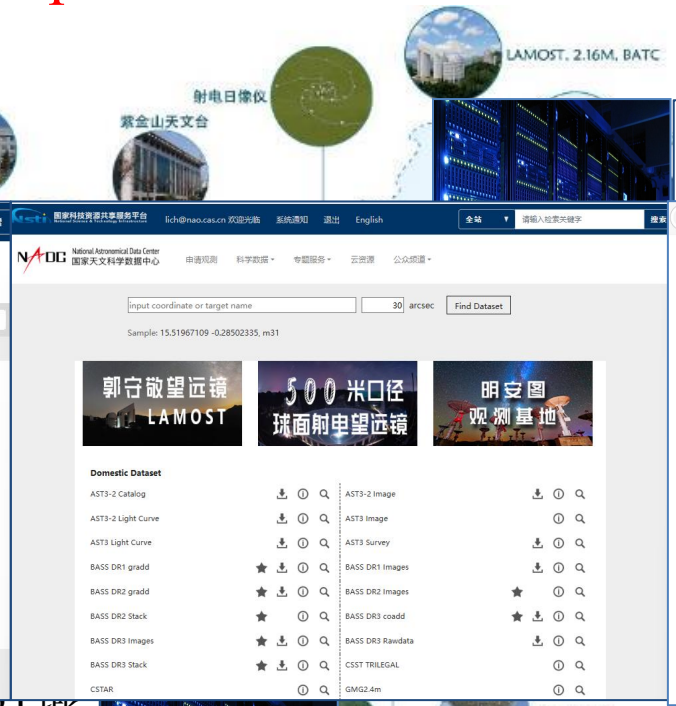
VO-compliant Tools & Services
DS9: Image visualisation
GOSSIP: SED fitting
VirGO: Search for Images and Spectra
IRAF: Image Reduction & Analysis
World Wide Telescope
Gaia - Graphical Astronomy and Image Analysis
SIMBAD
TESELA
VizieR
Browse the Registries
WIRR - Web Interface to the Relational Registry
RSS Feed of New & Updated Services in the VO
AppLauncher
or use VODesktop
Manuals, Tutorials, How-tos
ADQL & TAP Tutorial
Aladin User manual
CASSIS documentation
Datascope how to
Data Discovery Tool Tutorial PDF (video)
Firefly documentation
Iris How-to (video)
Montage help
SkyView documentation
Specview examples
SPLAT documentation
STILTS documentation
TOPCAT documentation
VisIVO how to
VOSpec User manual

天文数据服务

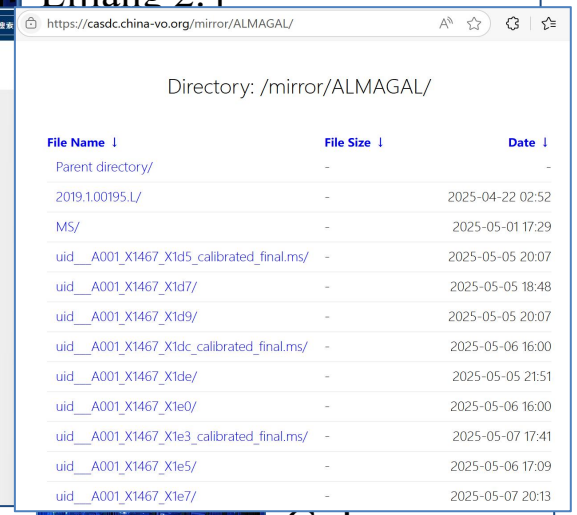
FAIR: Findable、Accessible、Interoperable、Reusable

FAST Pulsar





LAMOST
Lijiang 2.4



FAST 第21批
FAST 第20批
FAST 第19批
FAST 第18批
……





Galaxy
Simulation
……

天文大数据的挑战

National Astronomical Data Center 国家天文科学数据中心

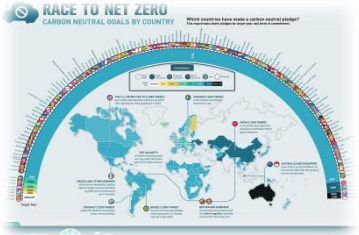
- **TB时代（~2000）**
  - 斯隆数字化巡天（SDSS）
  - 郭守敬望远镜（LAMOST）
- **PB时代（~2015）**
  - 盖亚天文卫星（Gaia）
  - 500米口径球面射电望远镜（FAST）
  - 泛星计划（Pan-STARRS）
- **EB时代（<2030）**
  - 平方千米天线阵（SKA）
  - 巡天空间望远镜（CSST）
  - 大口径全天巡视望远镜（LSST）
  - 欧几里得空间望远镜（Euclid）

Survey Project	Data volume	catalogue
2MASS	10TB	500 million
GALEX	30TB	300 million
ALLWISE	10TB	700 million
SDSS DR16	100TB	1.2 billion
LAMOST	200TB	20 million
UnWISE	**175TB**	**2.2 billion**
DESI Legacy Imaging Surveys	**200TB**	**2 billion**
FAST	**~70PB**	
CSST	~ 5PB	10 billion
LSST	~ 150PB	20 billion

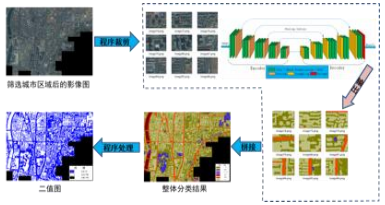
开放科学

National Astronomical Data Center 国家天文科学数据中心

为充分利用全球科技资源应对人类面临的重大挑战，UNESCO发起和推进开放科学运动



碳中和、碳达峰计划进度



异构、复杂环境数据分析



迎接后COVID19时代全球重构



人口与健康的长期工作



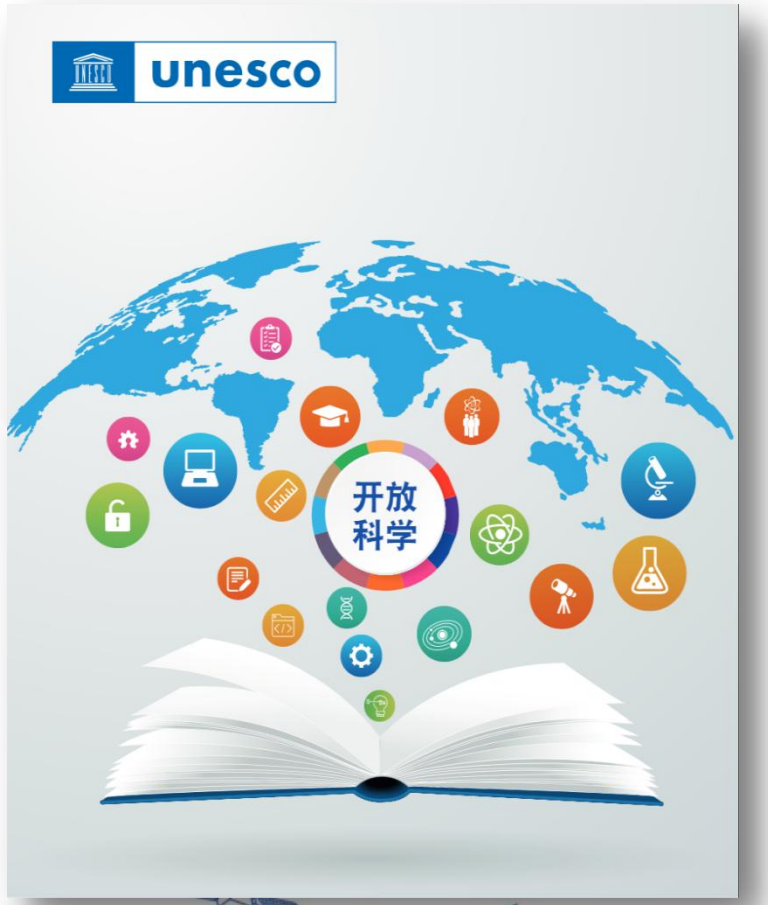
技术革新迭代的迫切需求



联合国可持续发展目标



开放科学



教科文组织开放科学建议书
UNESCO Recommendation on Open Science,2021

开放科学定义：一个集各种运动和实践于一体的包容性架构，旨在实现人人皆可公开使用、获取和重复使用多种语言的科学知识，为了科学和社会的利益增进科学合作和信息共享，并向传统科学界以外的社会行为者开放科学知识的创造、评估和传播进程



四大支柱：开放式科学知识、开放科学基础设施、科学传播、社会行为者的开放式参与以及与其他知识体系的开放式对话



开放科学基础设施

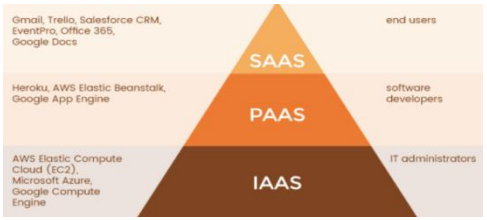
National Astronomical Data Center 国家天文科学数据中心

资源特征与研究焦点

- 计算资源
  - 特征：流转效率高
  - 分类：超算资源/云计算资源
  - 焦点：资源释放、资源使用激励机制
- 存储资源
  - 特征：流转效率低
  - 分类：异地备份服务
  - 焦点：效率、安全、长期可持续性
- 软件资源
  - 开源软件-云化部署
  - 自主研发软件-软件资源池
  - 商业软件-统一购买license+云化服务（选择性）
- 数据资源
  - 特征：生产成本、维护成本高，流通成本较低
  - 焦点：成本补偿机制，数据融合机制，数据质量控制

- 四类科研资源链式、网状推进机制
  - **资源** 数据-存储/计算/软件相互依附
  - **成本** 分摊机制
  - 生产成本、运行维护成本、定制化服务成本（成本计量）
  - 标的资源流通成本-激励机制（交换价值）
  - 标的间融合交叉（复合价值）





FAIR (Findable、Accessible、Interoperable、Reusable) + Computable

主要内容

National Astronomical Data Center
国家天文科学数据中心

- 天文科学平台建设背景
- **天文科学平台发展趋势**
- 天文科研云平台进展
- 天文科研云平台应用案例

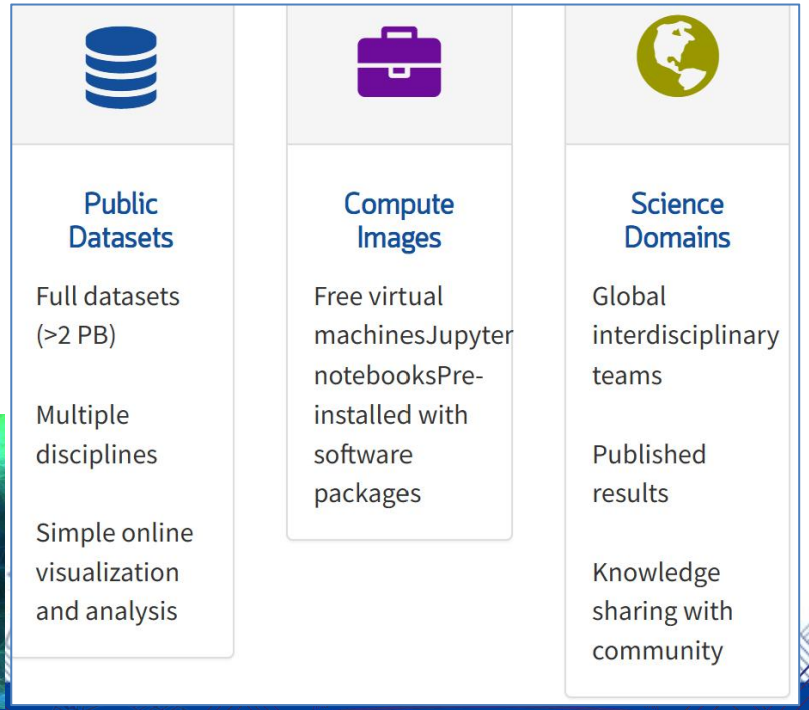
SciServer



National Astronomical Data Center
国家天文科学数据中心

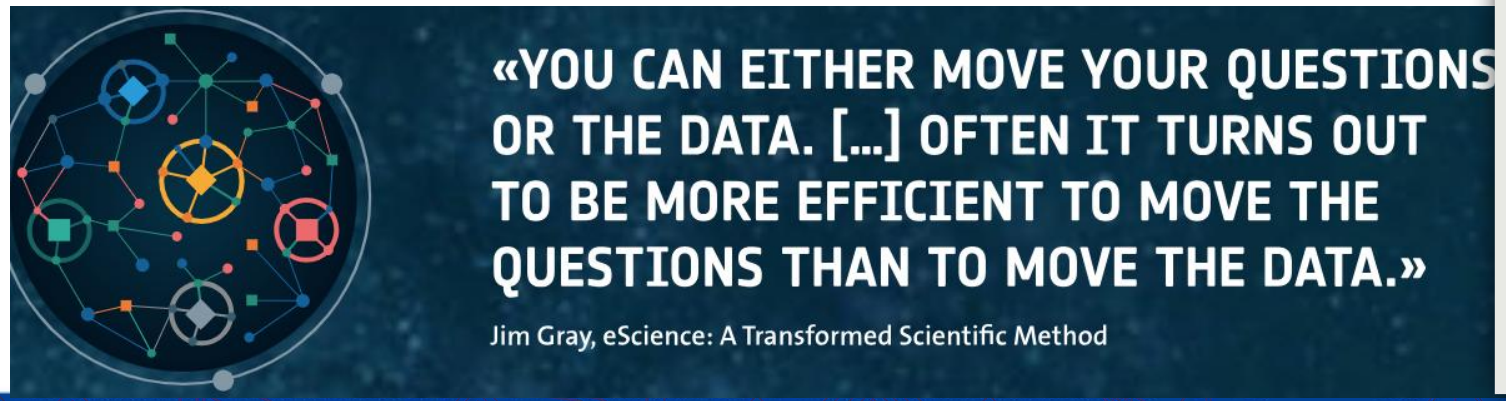
- SciServer is a fully integrated cyberinfrastructure system encompassing related tools and services to enable researchers to cope with scientific big data.
- SciServer enables a new approach that will allow researchers to work with **Terabytes or Petabytes of scientific data**, without needing to download any large datasets.
- This SciServer approach offers tremendous advantages to scientists:
  - By offering Big Data search and analysis capabilities online…
  - By offering worldwide access to large simulation datasets…
  - By providing a cloud-based scientific data storage system…
  - By adapting existing, working tools…
  - By developing new Citizen Science Projects…
  - By offering an extensive set of collaborative features…



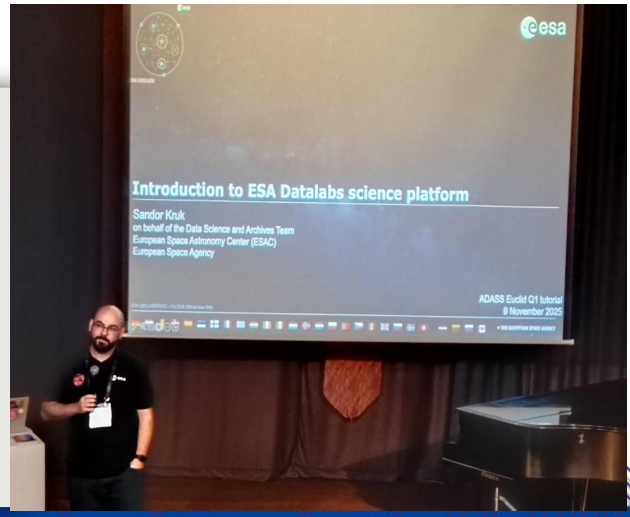


ESA datalabs

- ESA Datalabs is a state-of-the-art e-science platform, crafted to empower users by allowing them to run their code directly on ESA's infrastructure while seamlessly accessing the vast ESA archives.
- The versatility of ESA Datalabs is encapsulated in its domain-agnostic architecture. It advances research and innovation by melding big data access, notebook technologies, and domain-specific software.
- ESA datalabs gives you a privileged position; bring your code directly to ESA’s infrastructure – there is a great set of tools and programming languages are flexible – and execute it with direct access to ESA’s archives.
- ESA datalabs offers a catalog of datalabs you can use. They range from new tools that are quickly become de facto standards to older software that has been repackaged to run inside virtual computers. All are accessible via your web browser.



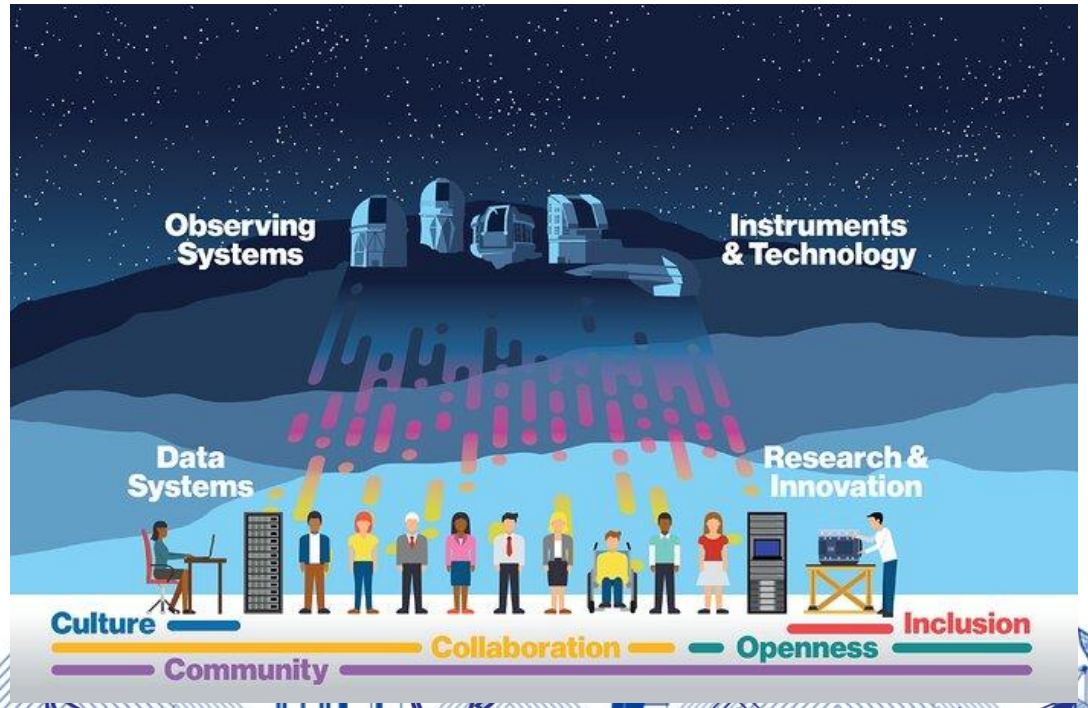


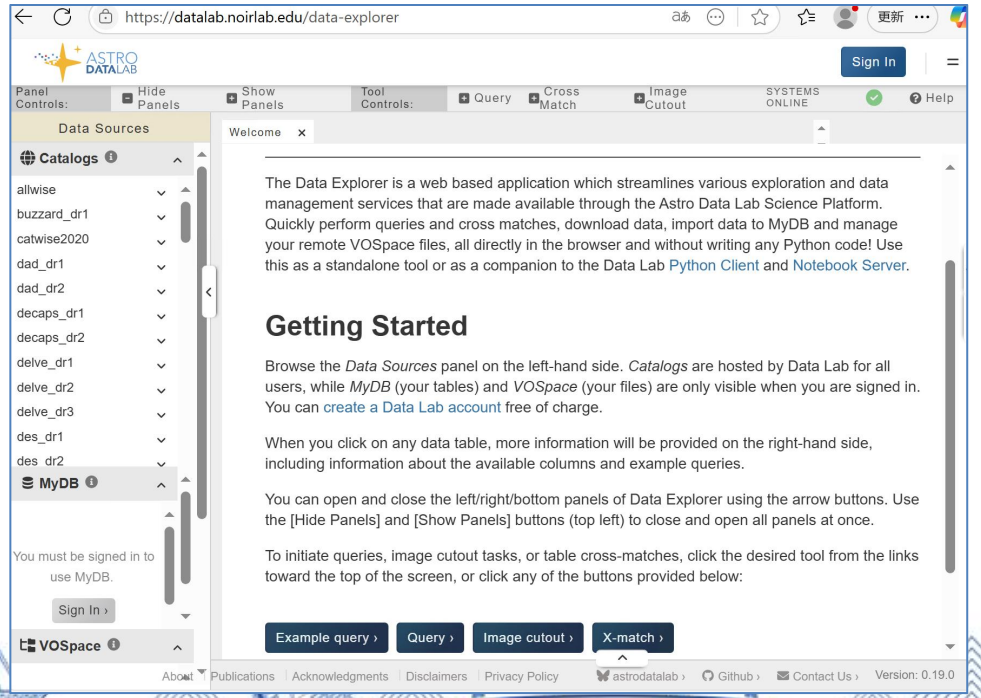


NOIRLab's Astro Data Lab

National Astronomical Data Center 国家天文科学数据中心

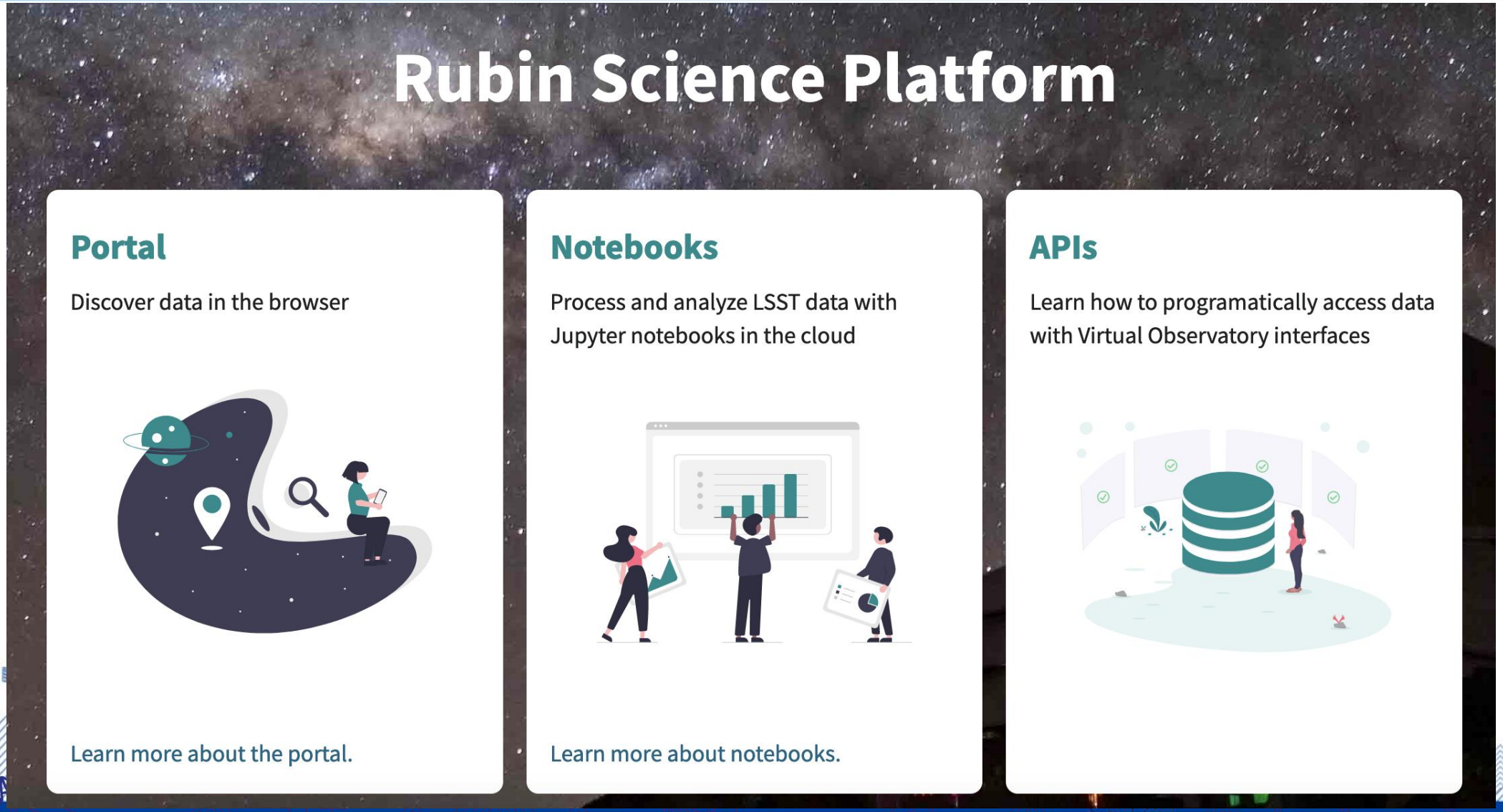
- NOIRLab's Astro Data Lab is a community science platform. Its foremost goal is to enable efficient exploration and analysis of the large datasets now being generated by instruments on NOIRLab's telescopes. Astro Data Lab is being developed at NOIRLab's Community Science & Data Center.
- Enable easy and open access to high-value survey catalogs from NOIRLab and external sources (e.g. Sloan Digital Sky Survey, Gaia, WISE), and to NOIRLab - based images linked to catalog objects.





Rubin Science Platform

National Astronomical Data Center
国家天文科学数据中心



Rubin Science Platform

National Astronomical Data Center 国家天文科学数据中心

- **Portal Aspect**
  - Implemented with IPAC **Firefly**, an open-source astronomical data access and visualization toolkit
  - Almost entirely based on either IVOA query protocols or lower-level services via DataLink
  - Really a general IVOA data explorer - only superficial Rubin-specifics (skinning, help text, defaults…)
- **Notebook Aspect**
  - JupyterLab, JupyterHub, Rubin-created deployment & launch tooling
  - Data services are either accessed from our (and external) IVOA services via PyVO, or for file-like data via the Rubin Butler Python API
- **API Aspect**
  - IVOA protocol standards (TAP, ObsCore, DataLink, SODA, …)
  - Rubin-specific services based on DALI, VOSI, etc. with service parameters discovered via DataLink service descriptors
  - “Heavyweight” services for forced photometry, data product re-creation, bulk cutouts
  - Many microservices supporting links between datasets (time series, coadd inputs, bulk cutouts)

主要内容

National Astronomical Data Center
国家天文科学数据中心

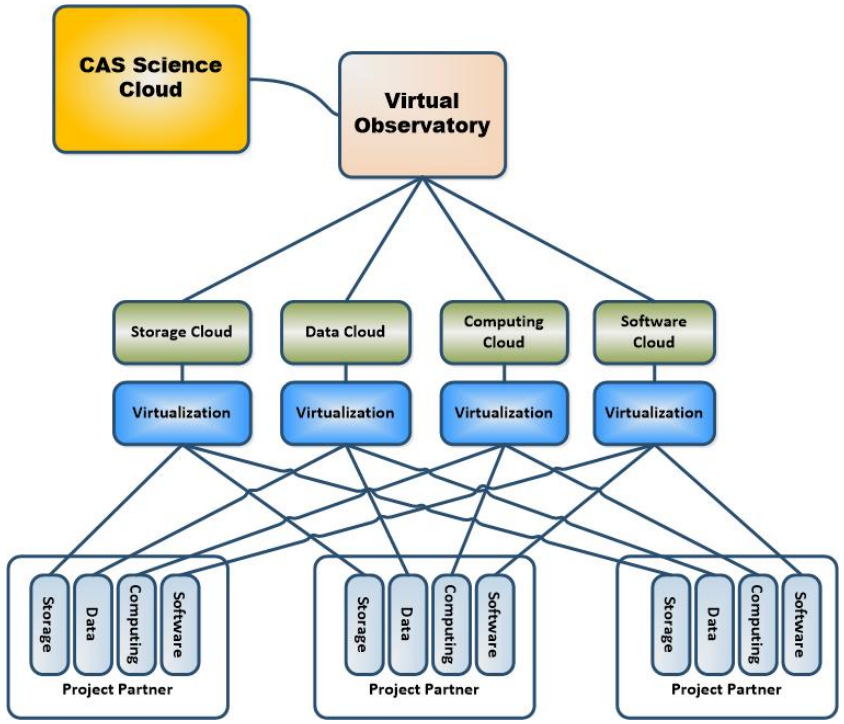
- 天文科学平台建设背景
- 天文科学平台发展趋势
- **天文科研云平台进展**
- 天文科研云平台应用案例

天文科研云平台的发展

National Astronomical Data Center 国家天文科学数据中心

□ 天文科研云平台 V1.0

- 始于2013年，院信息化专项资助





天文科研云平台的发展

National Astronomical Data Center 国家天文科学数据中心

□ 天文科研云平台 V2.0

- NADC 获资助，天文科学数据归档与集成



持续整合天文领域数据资源

- 数据集1072个
- 已发布1029个
- 总量87.43PB（含FAST）

承担科技计划和科研论文数据汇交

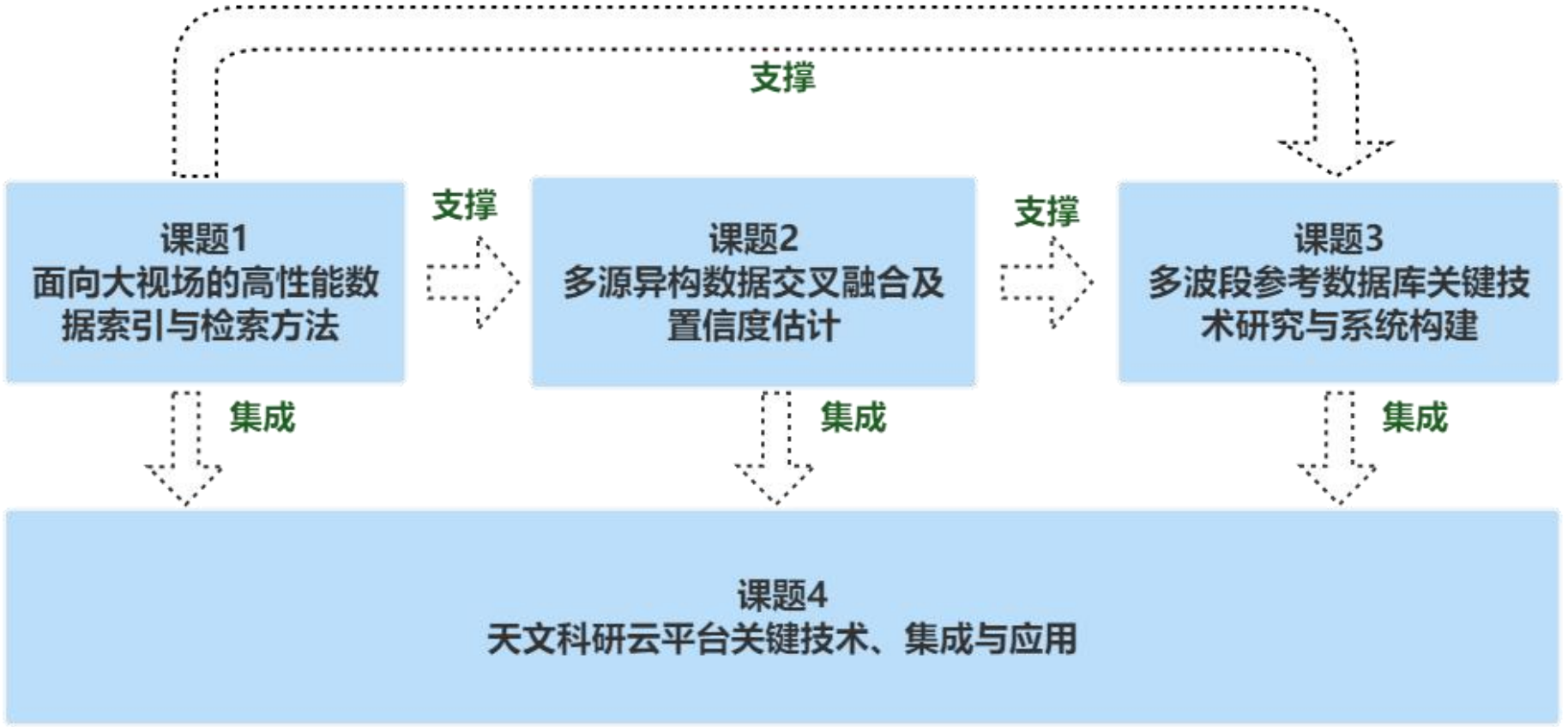
- 23个重点研发专项的数据汇交
- 202篇天文领域相关论文数据汇交

天文科研云平台的发展

National Astronomical Data Center
国家天文科学数据中心

□ 天文科研云平台 V3.0

- 获重点研发项目资助：海量多波段天文数据融合关键技术与科学应用 （2022-2025）



平台架构

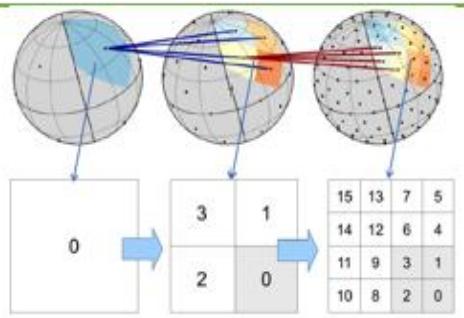
National Astronomical Data Center
国家天文科学数据中心



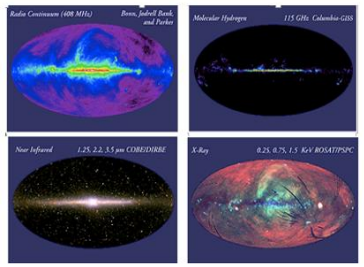
Interactive Tools



Analysis Tools



Search Algorithm



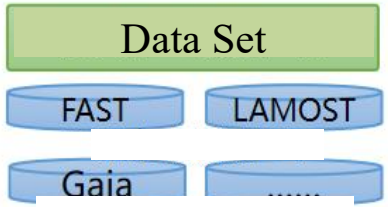
Science Research Environment



Archiving & Fusion



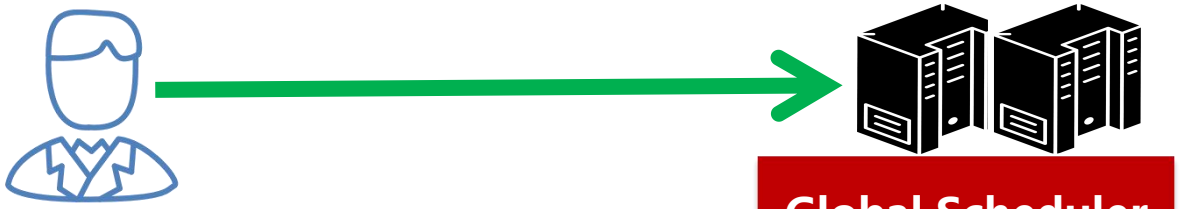
Computing



Storage

资源节点

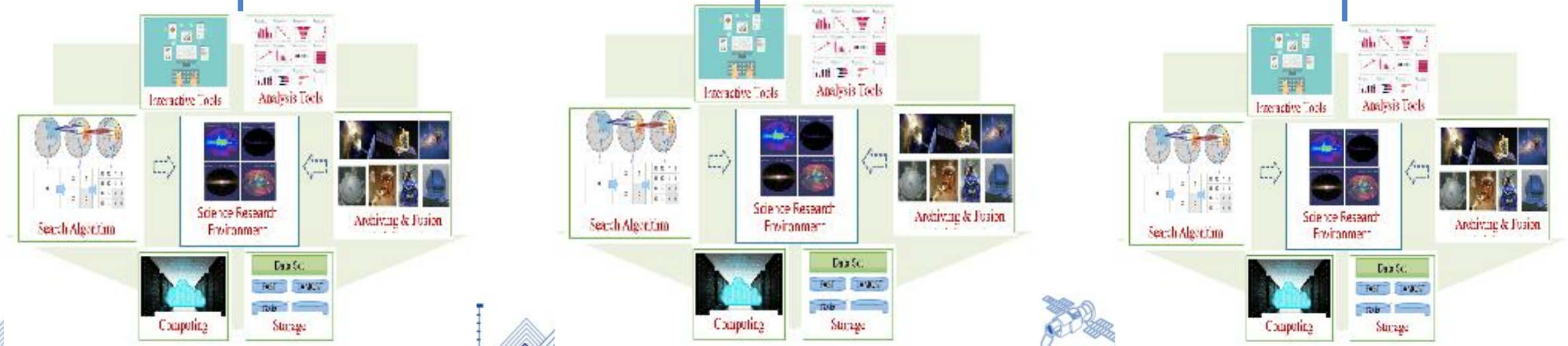
平台架构



Global Scheduler
(NadcSched)

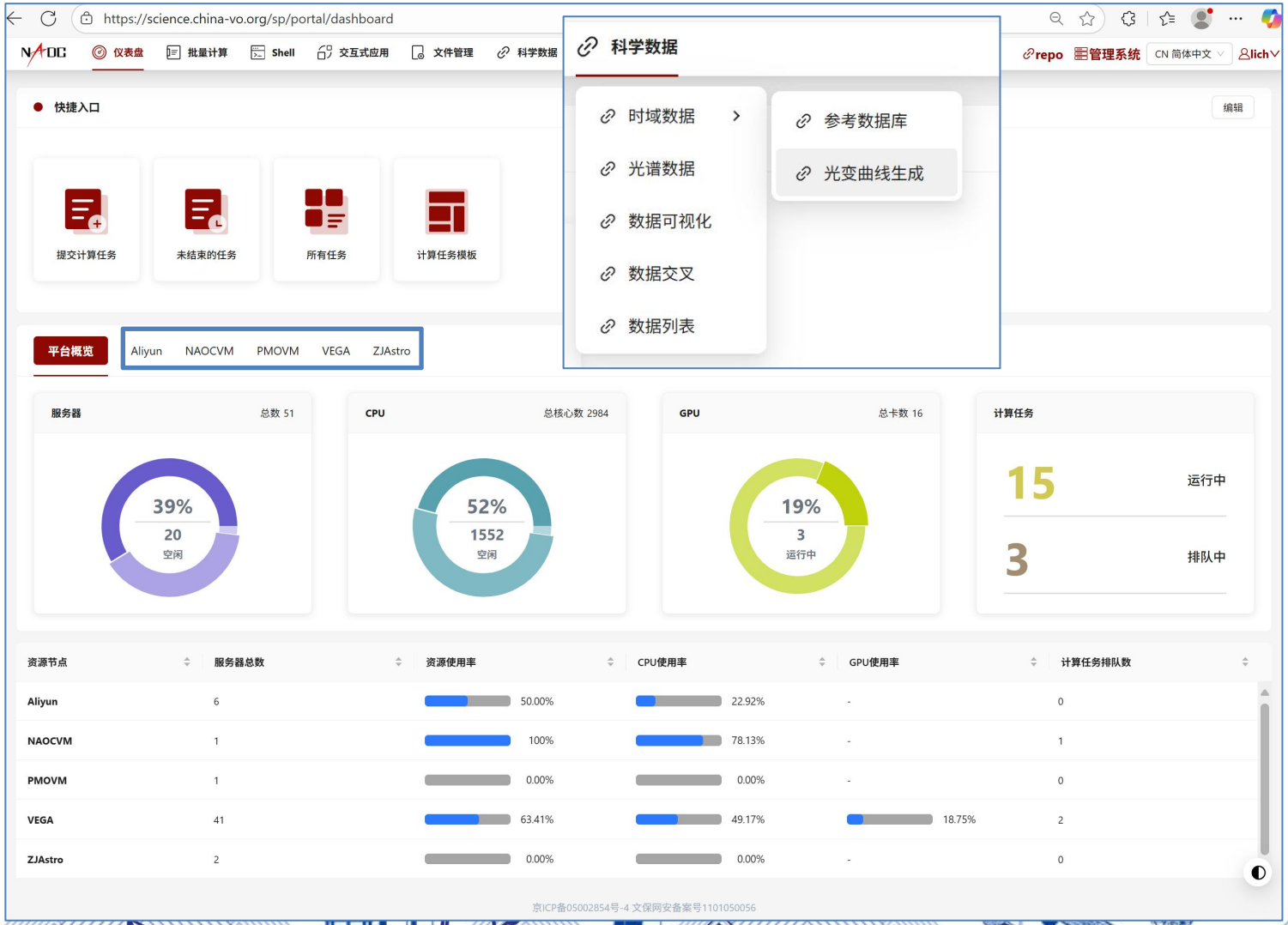


NADC



平台总览

National Astronomical Data Center
国家天文科学数据中心





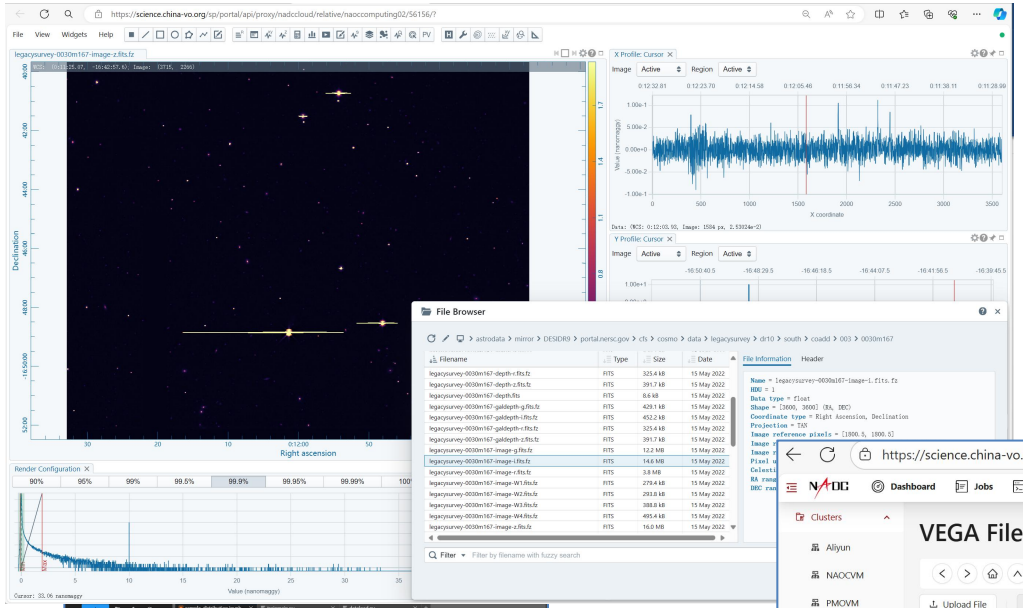
关键字: Search

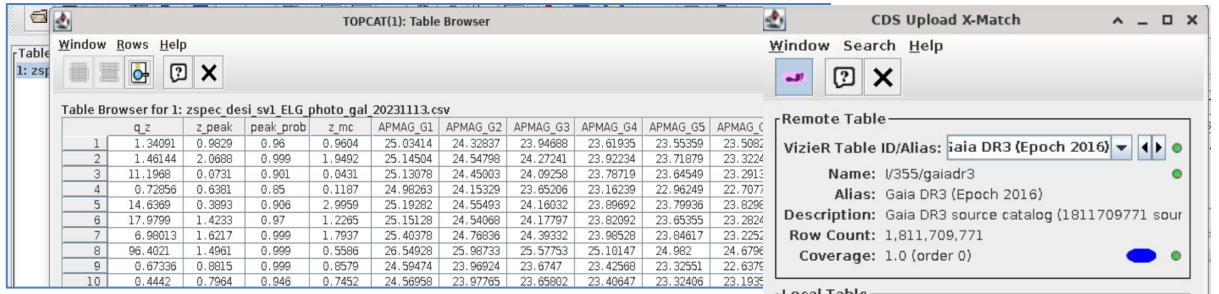
数据编号	数据名称	数据描述	资源节点
102045	ALMAGAL: ALMA银河系大质量原星团形成演化研究数据	本数据集是ALMA大型项目ALMA银河系大质量原星团形成演化研究（…	VEGA
102044	Euclid Q1 星表数据	Euclid Q1的星表数据包含6个目录，6个星表目录分别是: MER_FINA…	VEGA
100550	LAMOST光谱巡天第十次数据发布第〇版 第二批数据	LAMOST第十次巡天数据发布第〇版第二季度释放了2022年1月1日到…	VEGA
102034	eROSITA-DE 巡天第一次数据发布	德国 eROSITA 联盟 (eROSITA-DE) 于2024年1月31日公开了 SRG/eR…	VEGA
100878	SAGES第一版恒星物理参数增值星表	基于SAGES第一版数据以及Gaia DR3数据，我们提供了近2600万颗…	VEGA
100876	SAGES巡天第一次数据发布及其增补	元素丰度与银河系演化巡天（Stellar Abundance and Galactic Evol…	VEGA
100875	中国天眼中性氢巡天第一次数据发布	中国天眼中性氢巡天（英文缩写FASHI）旨在覆盖整个FAST可探测天…	ALIYUN
100669	南极AST3-2数据发布	南极AST3-2望远镜2016年3月至5月SN巡天数据i波段数据发布，数据…	VEGA

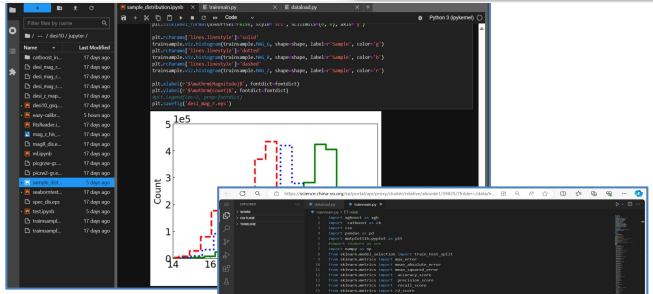
数据、计算与软件融合

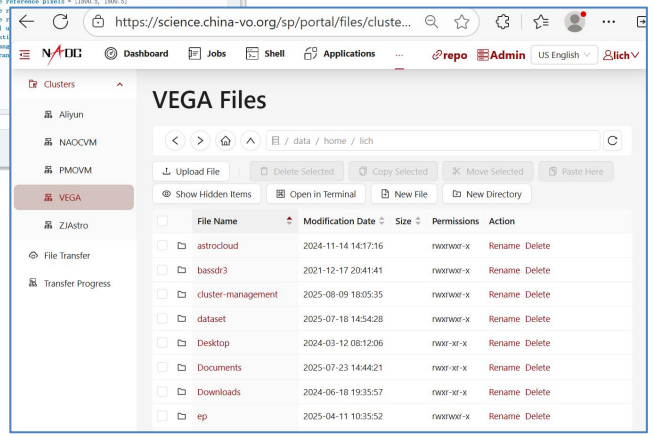
National Astronomical Data Center
国家天文科学数据中心

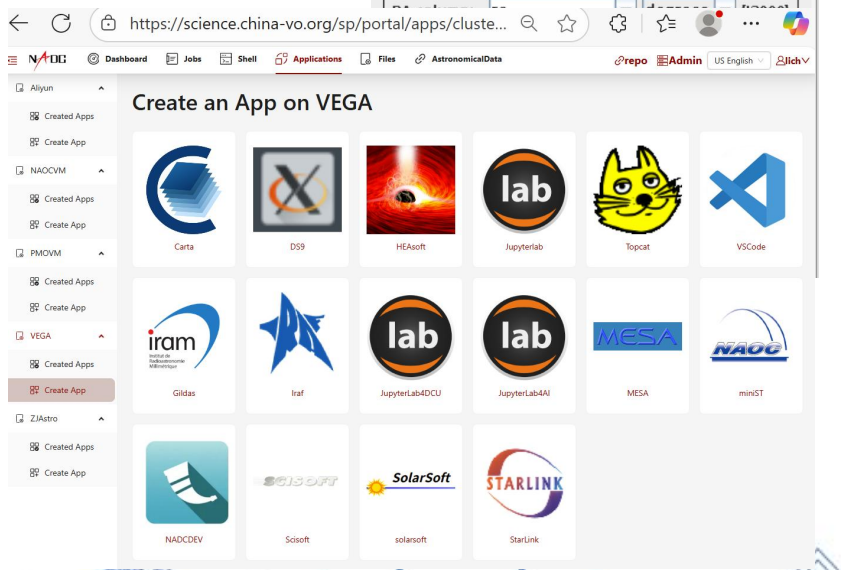
The fusion of computing, data and software tools

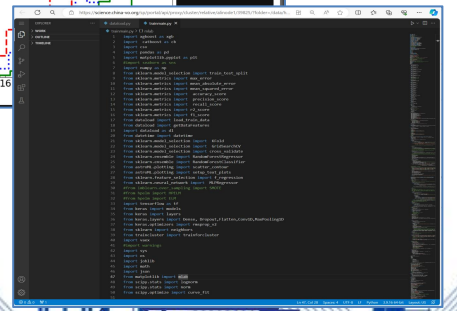










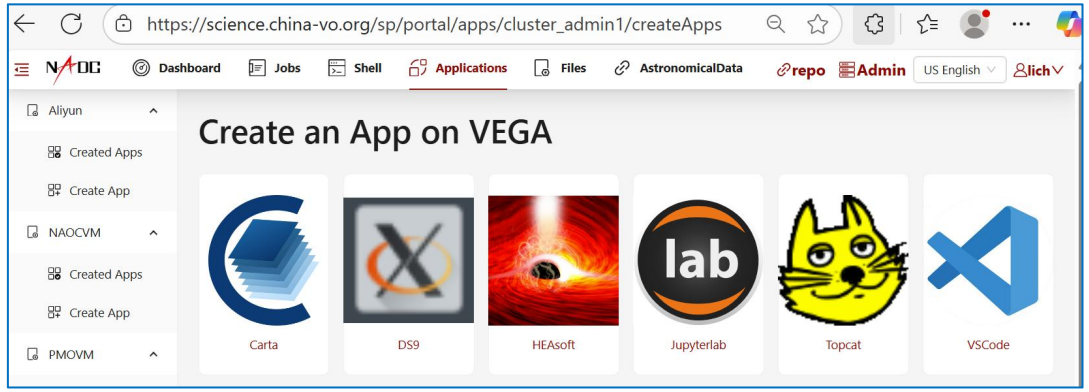


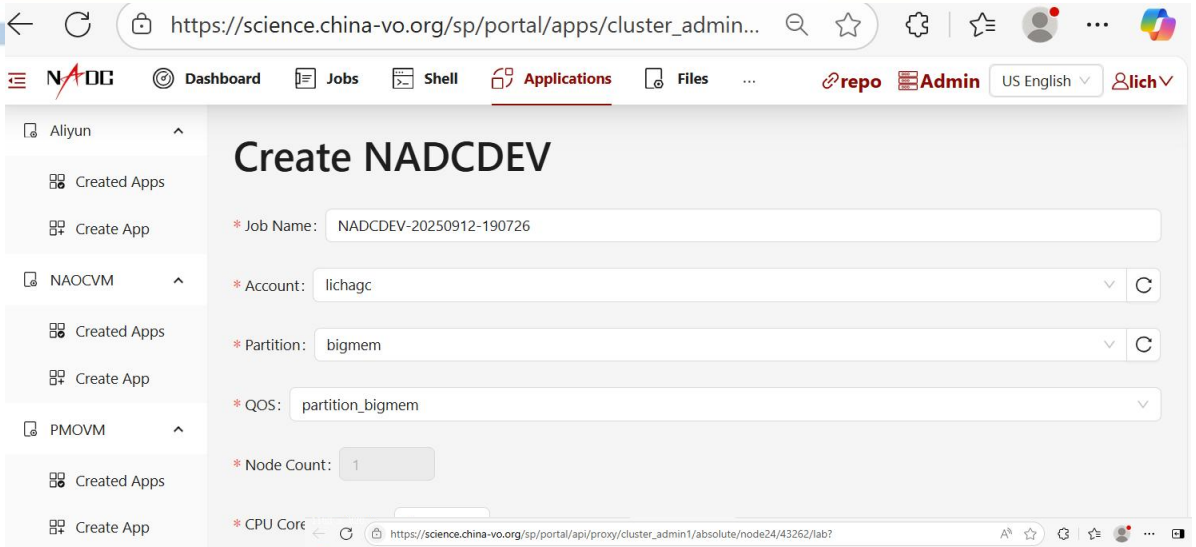
Personal Data/
ScienceData

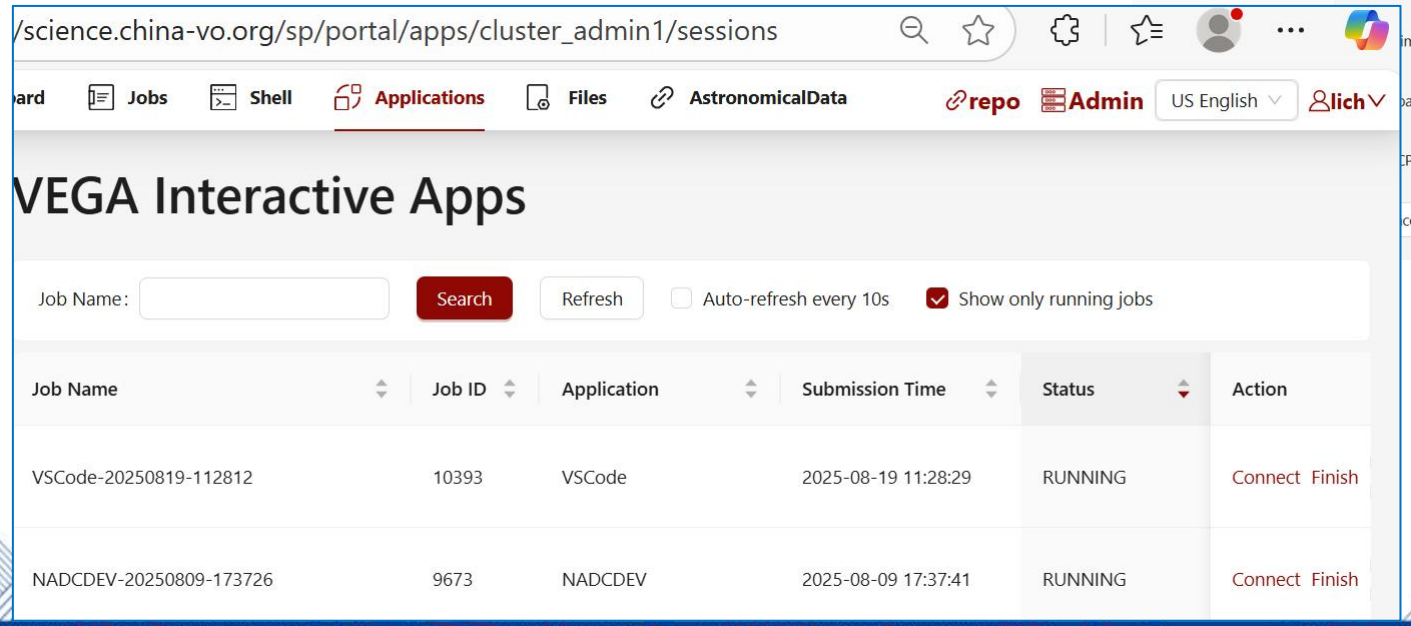
天文软件的一键部署

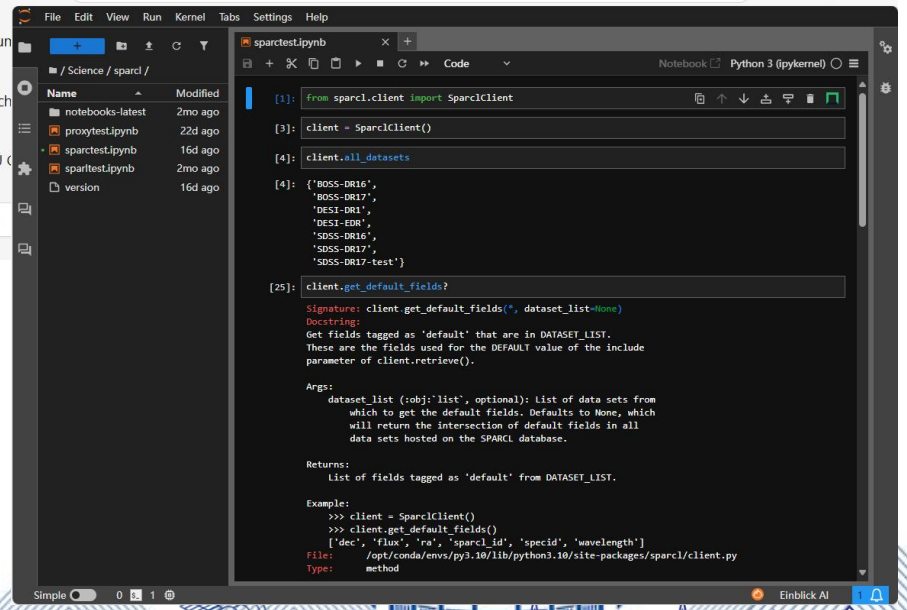
National Astronomical Data Center 国家天文科学数据中心

On-demand, one-click deployment





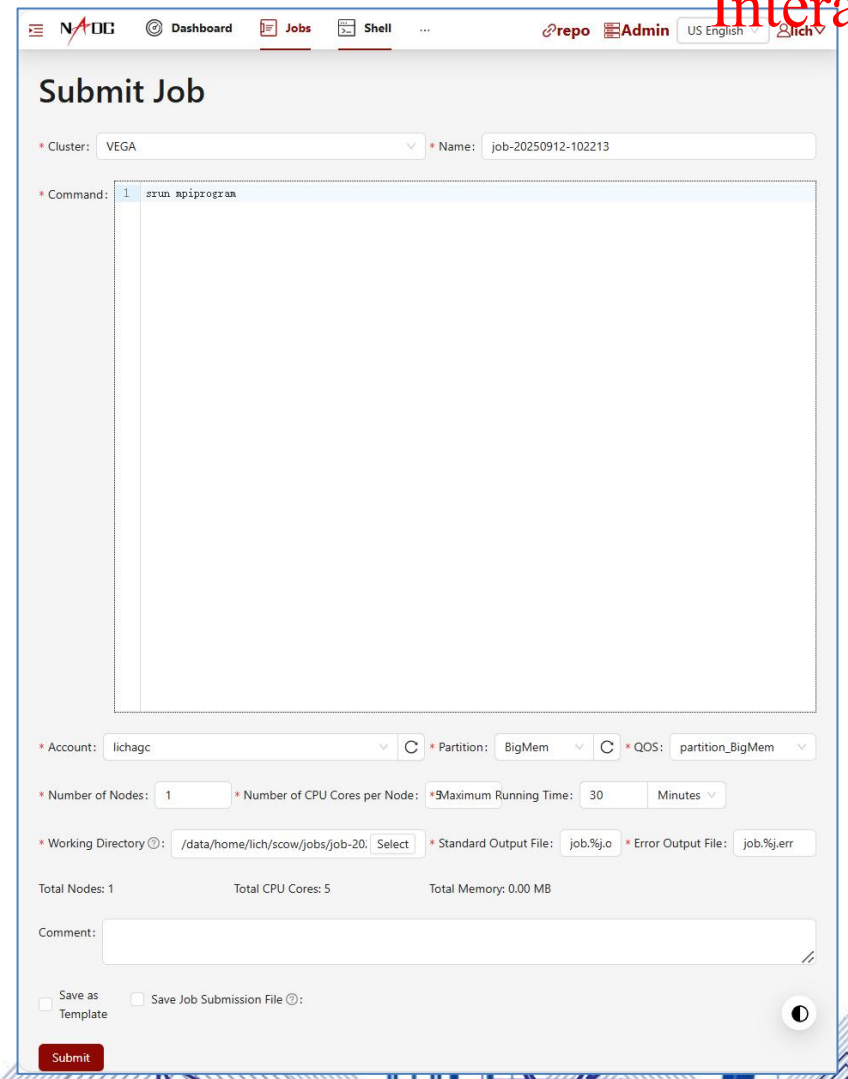


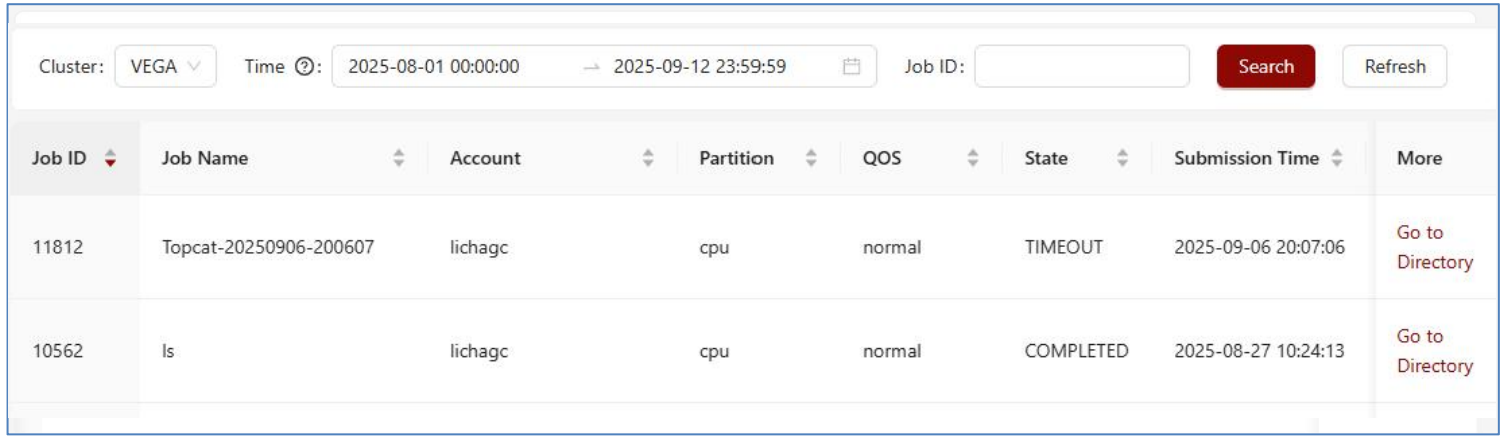


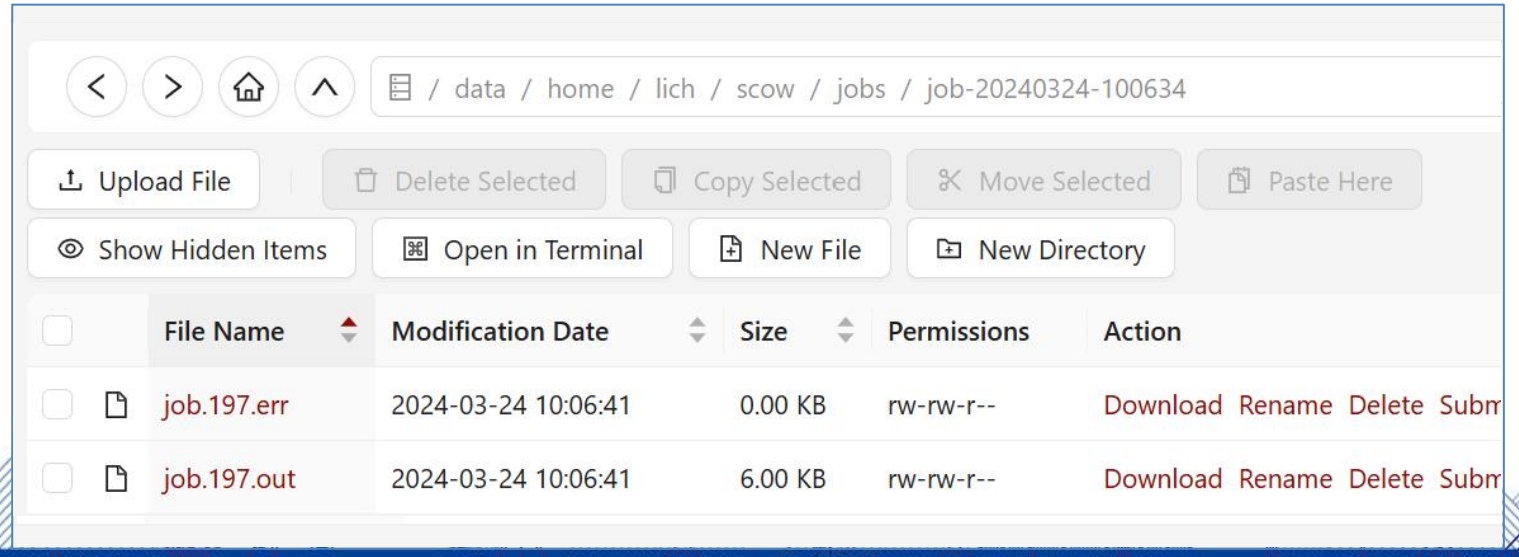
基于WEB的统一访问

National Astronomical Data Center 国家天文科学数据中心

Interactive applications and large-scale computing share resources





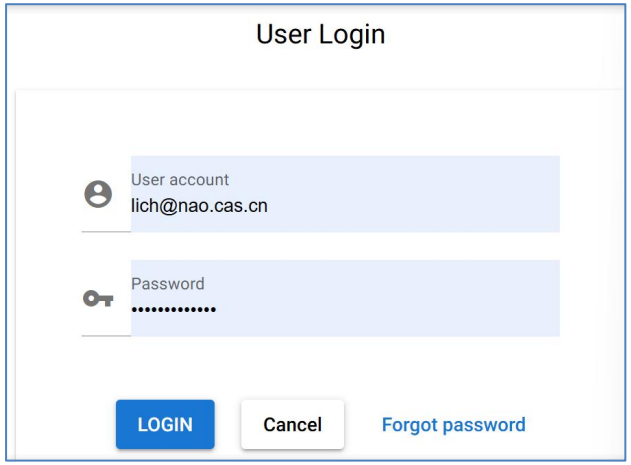


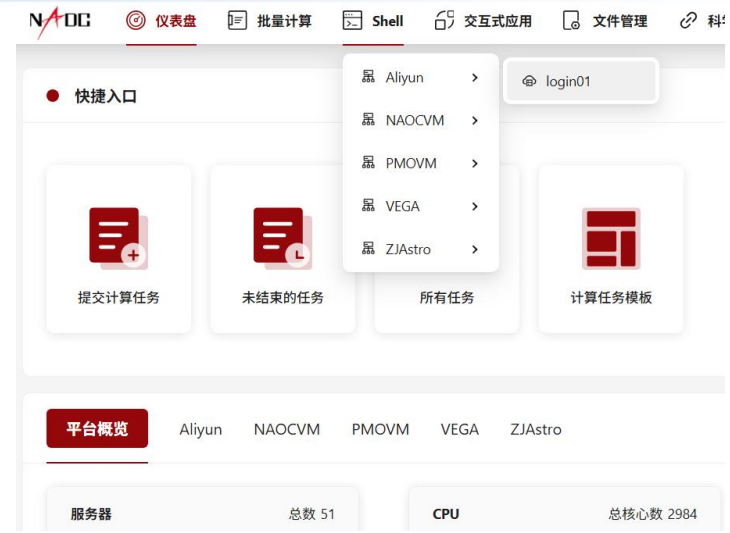
NADC 通行证整合 单点登录（OAUTH2.0+LDAP）

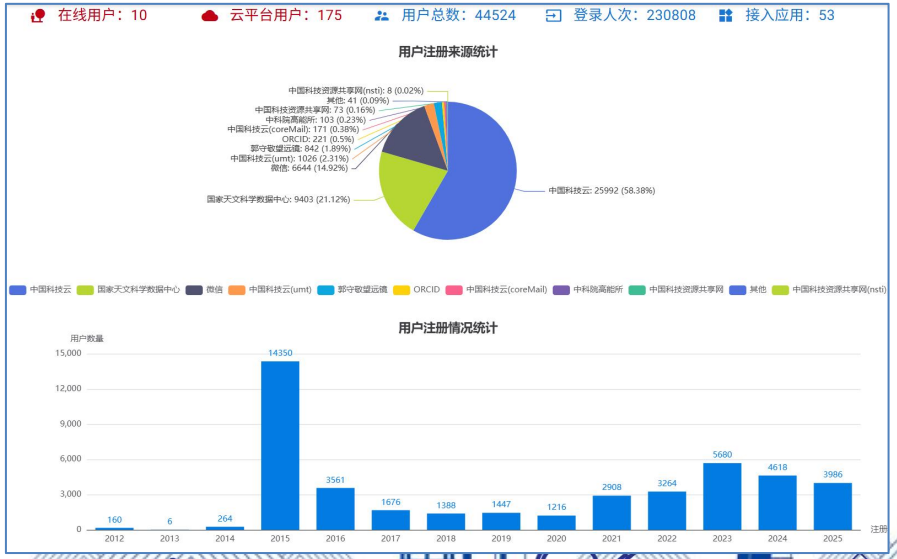
National Astronomical Data Center 国家天文科学数据中心

■ NADC用户通行证

- NADC业务系统的统一登录
- NADC内部资源的统一登录
- NADC业务系统的另一个集中入口
- 力所能及的全站统计分析

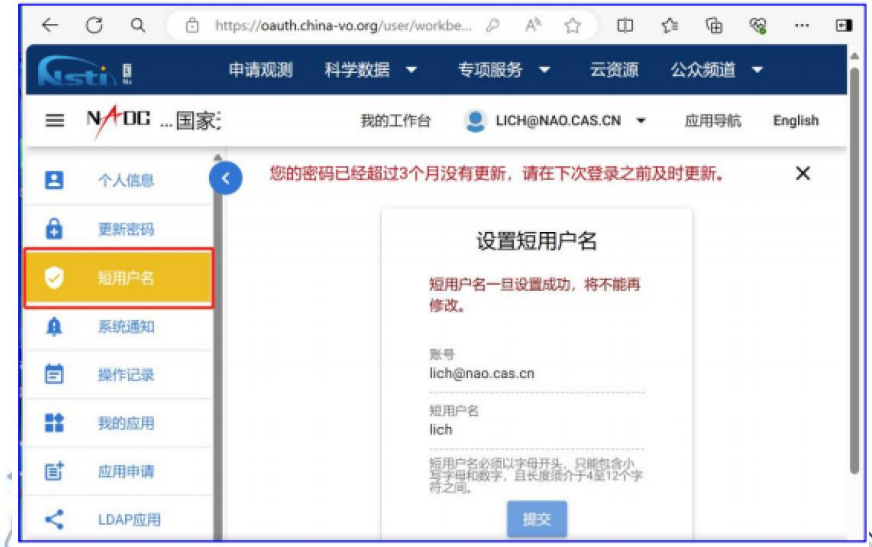






■ 平台帐号申请

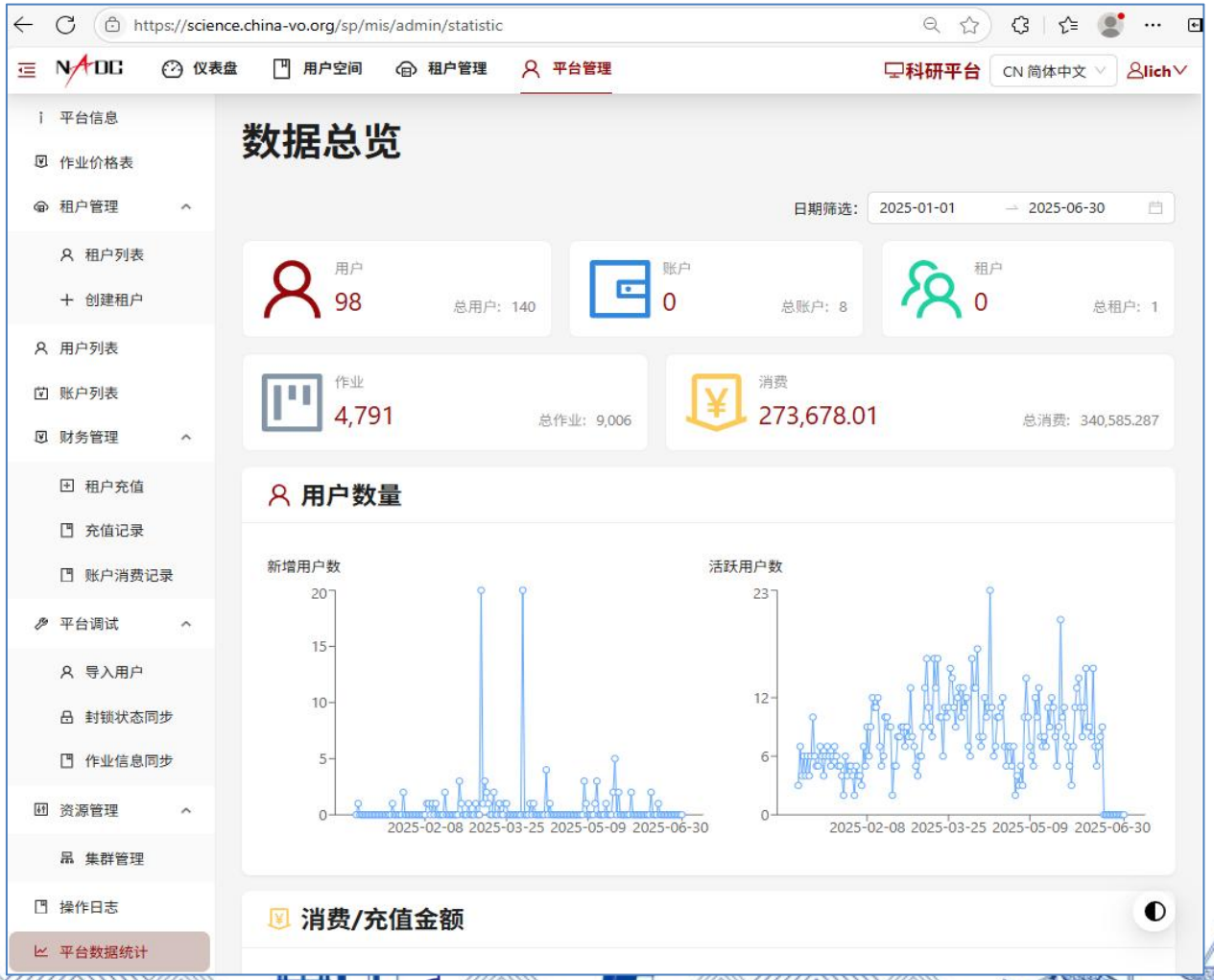
- ■ 注册NADC通行证
- ■ 设置短用户名
- ■ 提交帐号申请表
- ■ 等待管理员邮件通知

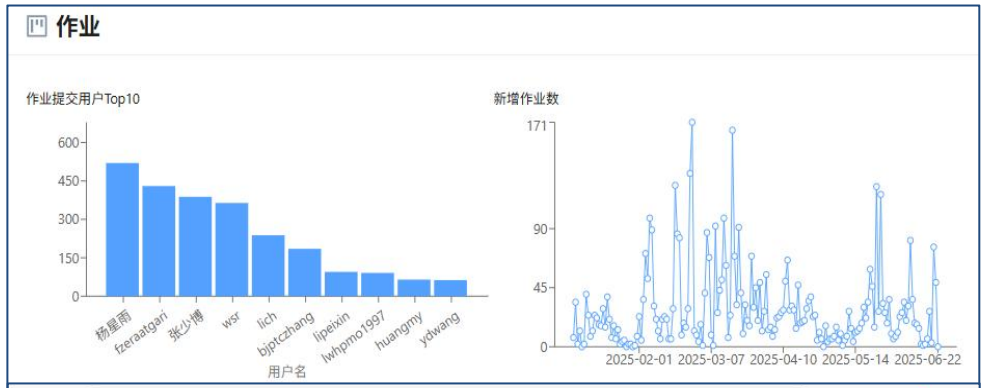


丰富的统计管理功能

National Astronomical Data Center 国家天文科学数据中心

Usage statistics and multi-level account management







主要内容

- 天文科学平台建设背景
- 天文科学平台发展趋势
- 天文科研云平台进展
- 天文科研云平台应用案例

应用场景1：面向基于AI的天文研究

National Astronomical Data Center 国家天文科学数据中心

- **AI algorithm**
  - https://sklearn.org/stable/
  - (Gradient boosting, nearest neighbors, MLP，ANN，random forest, logistic regression, k-Means, HDBSCAN, hierarchical clustering and more)
  - XGBoost, CatBoost
  - Deep Learning (CNN, RNN, LSTM, GRU, Transformer, etc.)

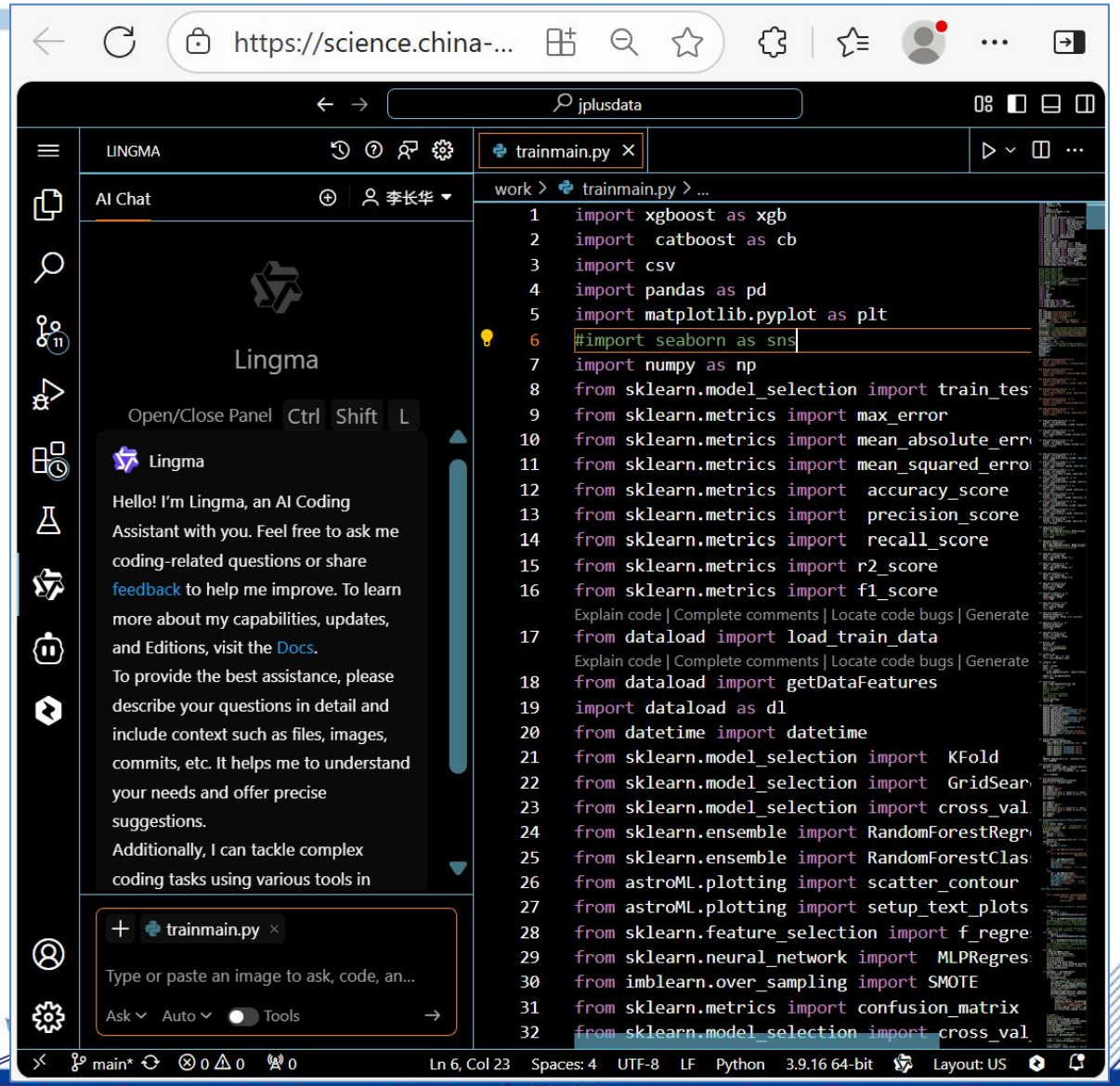
- **AI-Ready Datasets and preprocessing tools**
  - Spectrum data (LAMOST, DESI, SDSS)
  - Photometric data (DESI Imaging surveys, GAIA，...)

- **Computing power for AI** (3000 CPU cores and some GPU cards)

应用场景1：面向基于AI的天文研究

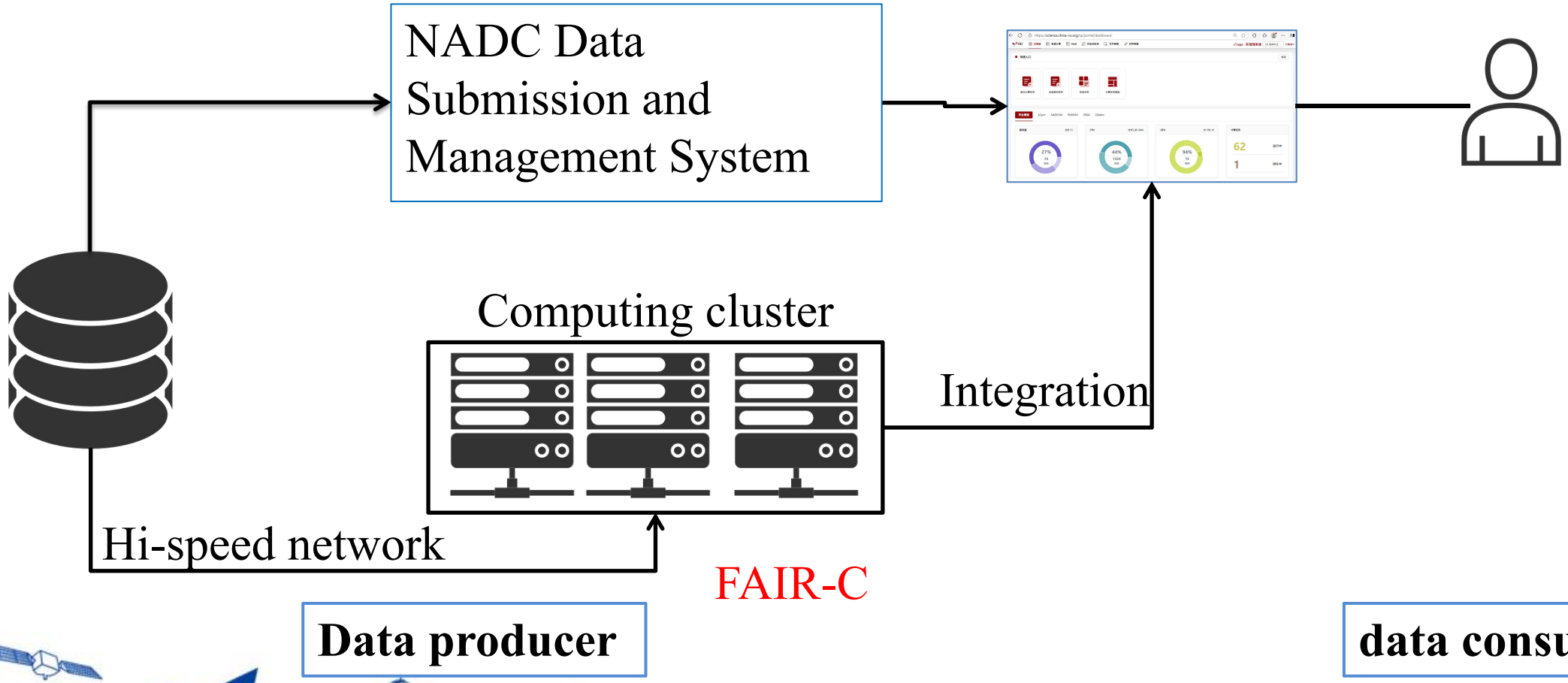
National Astronomical Data Center
国家天文科学数据中心

- **AI for development**
  - AI Coding Model
    - DeepSeek
    - Tongyi
    - Claude
  - VSCode
    - Cursor
    - Cline
    - Tongyilingma
    - CodeBuddy



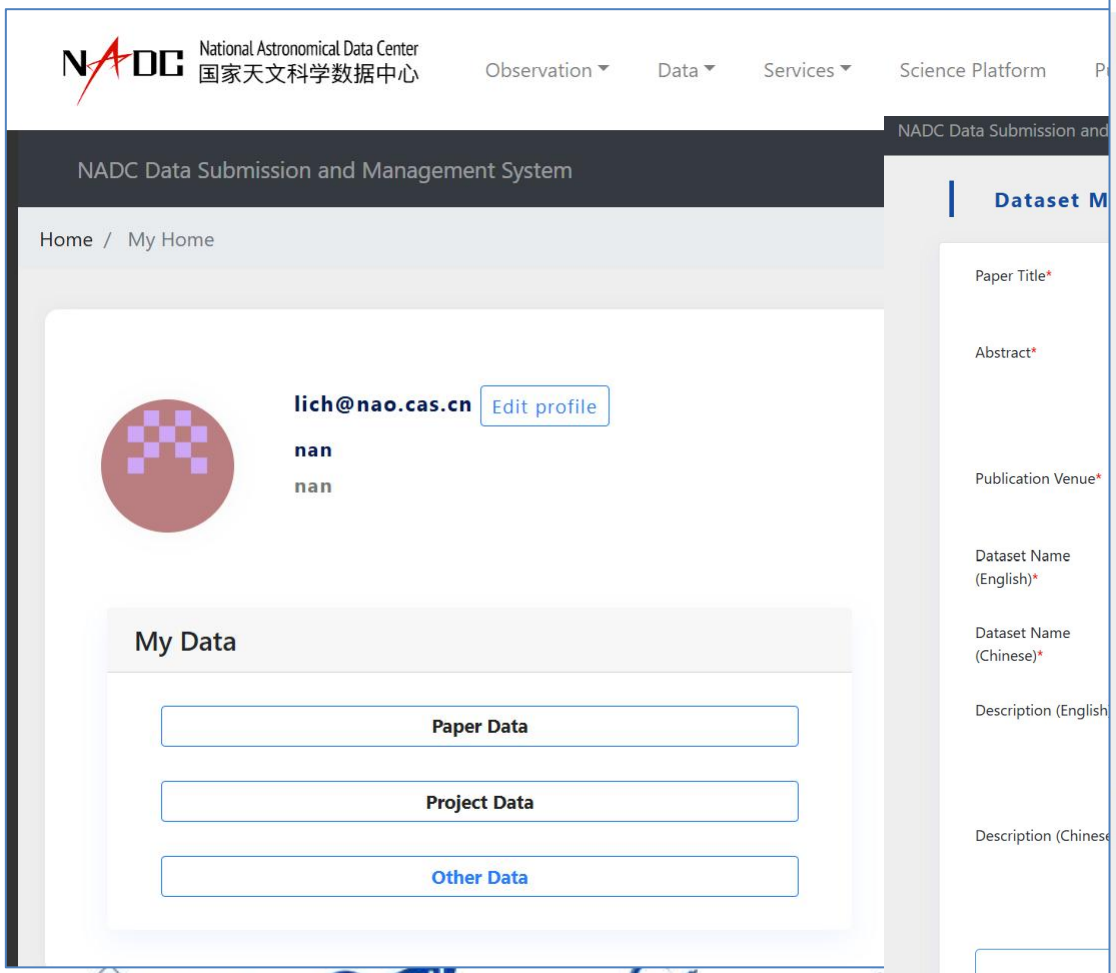
应用场景2：面向大规模数据集的科研

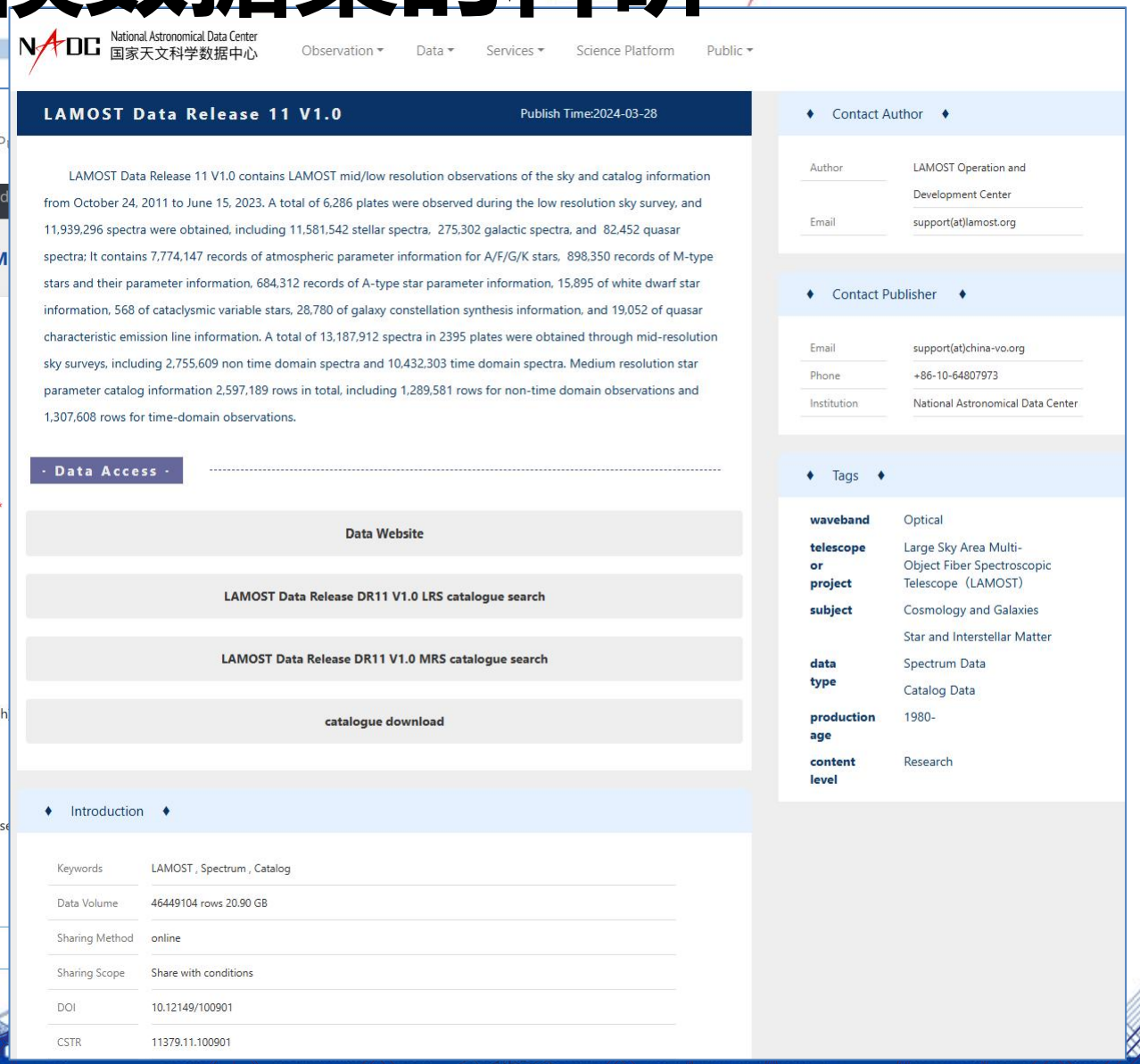
National Astronomical Data Center 国家天文科学数据中心



应用场景2：面向大规模数据集的科研

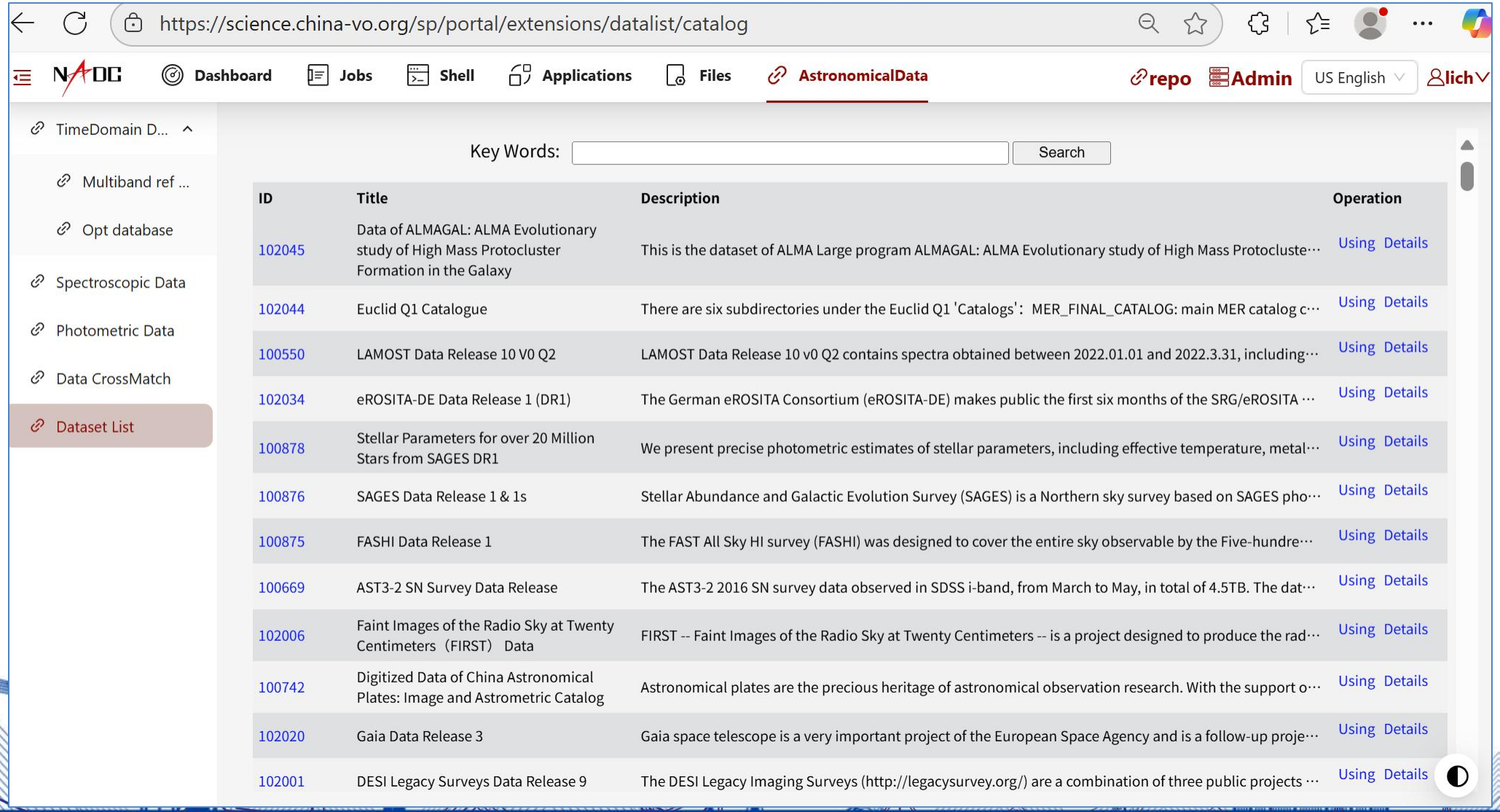
National Astronomical Data Center 国家天文科学数据中心





应用场景2： 面向大规模数据集的科研

National Astronomical Data Center 国家天文科学数据中心



ID	Title	Description	Operation
102045	Data of ALMAGAL: ALMA Evolutionary study of High Mass Protocluster Formation in the Galaxy	This is the dataset of ALMA Large program ALMAGAL: ALMA Evolutionary study of High Mass Protocluste···	Using Details
102044	Euclid Q1 Catalogue	There are six subdirectories under the Euclid Q1 'Catalogs': MER_FINAL_CATALOG: main MER catalog c···	Using Details
100550	LAMOST Data Release 10 V0 Q2	LAMOST Data Release 10 v0 Q2 contains spectra obtained between 2022.01.01 and 2022.3.31, including···	Using Details
102034	eROSITA-DE Data Release 1 (DR1)	The German eROSITA Consortium (eROSITA-DE) makes public the first six months of the SRG/eROSITA ···	Using Details
100878	Stellar Parameters for over 20 Million Stars from SAGES DR1	We present precise photometric estimates of stellar parameters, including effective temperature, metal···	Using Details
100876	SAGES Data Release 1 & 1s	Stellar Abundance and Galactic Evolution Survey (SAGES) is a Northern sky survey based on SAGES pho···	Using Details
100875	FASHI Data Release 1	The FAST All Sky HI survey (FASHI) was designed to cover the entire sky observable by the Five-hundre···	Using Details
100669	AST3-2 SN Survey Data Release	The AST3-2 2016 SN survey data observed in SDSS i-band, from March to May, in total of 4.5TB. The dat···	Using Details
102006	Faint Images of the Radio Sky at Twenty Centimeters（FIRST） Data	FIRST -- Faint Images of the Radio Sky at Twenty Centimeters -- is a project designed to produce the rad···	Using Details
100742	Digitized Data of China Astronomical Plates: Image and Astrometric Catalog	Astronomical plates are the precious heritage of astronomical observation research. With the support o···	Using Details
102020	Gaia Data Release 3	Gaia space telescope is a very important project of the European Space Agency and is a follow-up proje···	Using Details
102001	DESI Legacy Surveys Data Release 9	The DESI Legacy Imaging Surveys (http://legacysurvey.org/) are a combination of three public projects ···	Using Details

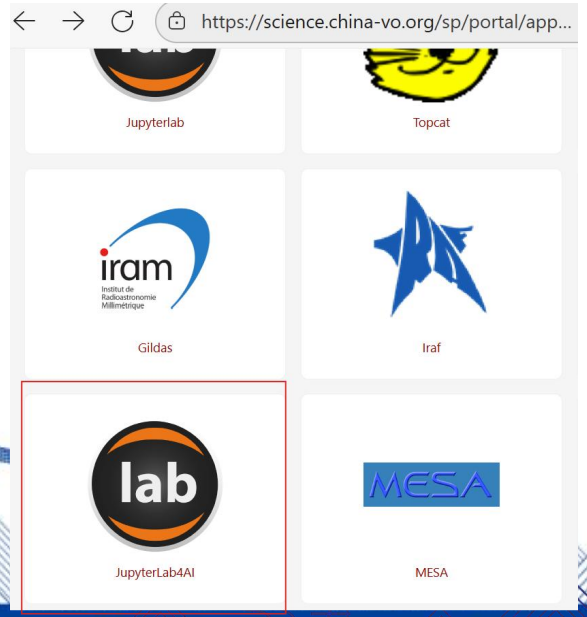
应用场景2： 面向大规模数据集的科研

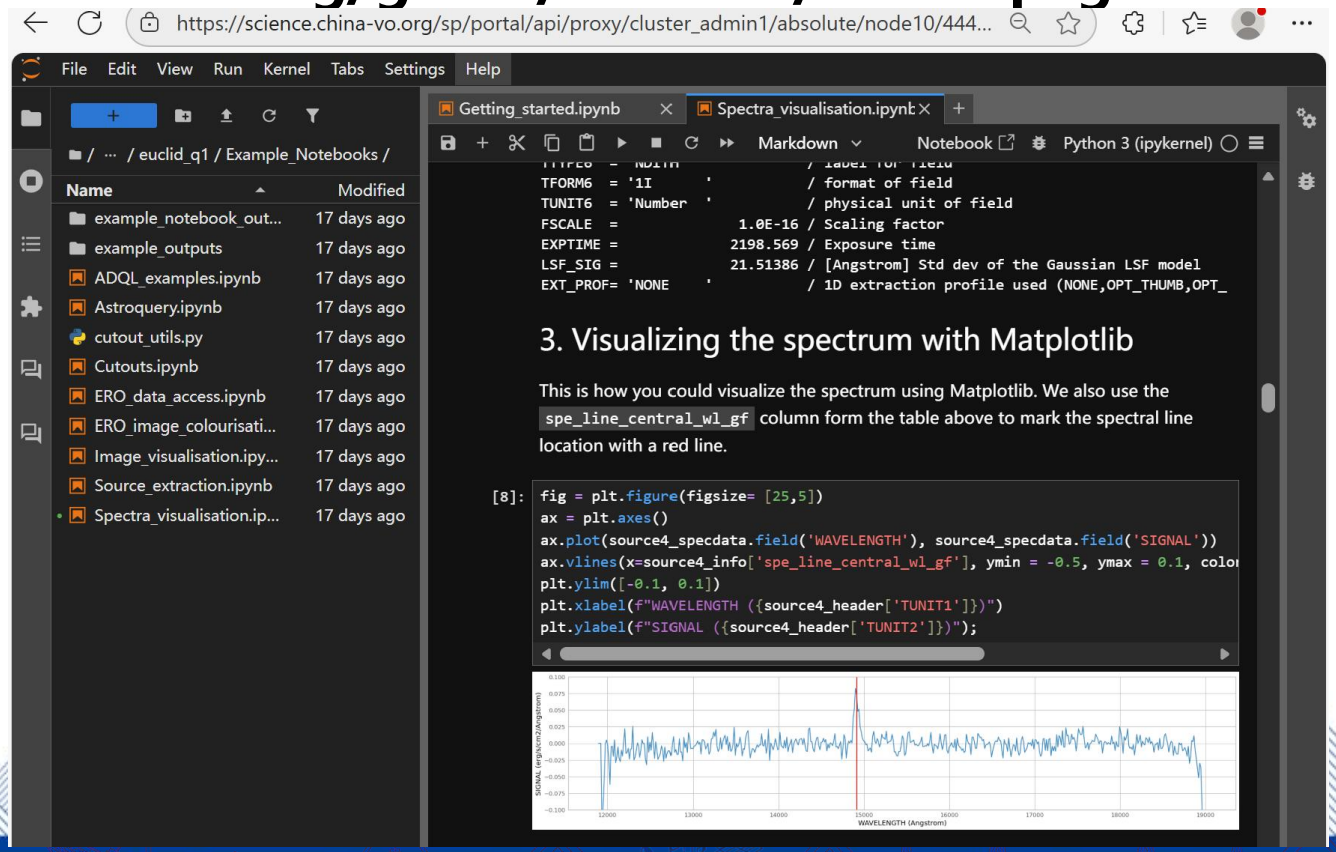
National Astronomical Data Center 国家天文科学数据中心

- 处理Euclid Q1数据
  - Euclid Q1 Sample:

    git clone https://science.china-vo.org/gitlab/external/euclidq1.git

  - 创建jupyterlab:





将来计划

- 持续优化科学平台的功能及使用体验
- 集成天文应用软件
- 标准化的节点配置及能力提升
- **典型科学案例**

https://science.china-vo.org/
https://science.china-vo.org/docs/
黄茂源海报

Thanks!

https://nadc.china-vo.org

vo@nao.cas.cn