

Resources and Services from National Astronomical Data Center

Chenzhou Cui

NAOC, CAS



It is not a talk about Science but one for Better Science

Contents

Introduction to VO and China-VO

VO, IVOA, China-VO, Science Platform

NADC Resources

data storage, data access, cloud computing, HPC, tools and information, WWT tours and curriculum

NADC Services

 data archiving, data management, software/system/platform design and development, cluster/HPC design, community outreach, education and public outreach

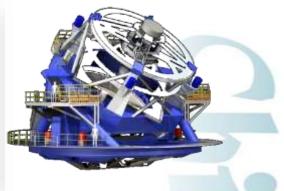


Heroes behind the Big Data Astronomy

















The Idea of VO

Vision of the VO:

- The Web is *transparent*. The goal of the Virtual Observatory is to achieve the same feeling for astronomical data - that it is all available to explore in a single transparent system.
- Like the World Wide Web, the VO is not a fixed system, but rather a way of doing things.
- Astronomical datasets, tools, services should work seamlessly together.
- The VO allows astronomers to interrogate multiple data centers in a seamless and transparent way, provides new powerful analysis and visualization tools within that system, and gives data centers a standard framework for publishing and delivering services using their data.
- 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.

-- my words

International Virtual Observatory Alliance

VO's vision is made possible by standardization of data and metadata, by standardization of data exchange methods, and by the use of a registry, which lists available services and what can be done with them.

IVOA:

- An organisation that debates and agrees the technical standards that are needed to make the VO possible
- A focal point for VO aspirations, a framework for discussing and sharing VO ideas and technology
- Promoting and publicising the VO

Basic Information about IVOA

- Created in 2002
- 21 member VO projects
 - Netherlands shows strong interests
- 6 Working Groups, 8 Interest Groups
- 2 Interoperability meetings per year
 - May
 - Oct/Nov with ADASS
- ~ 46 interoperability standards

































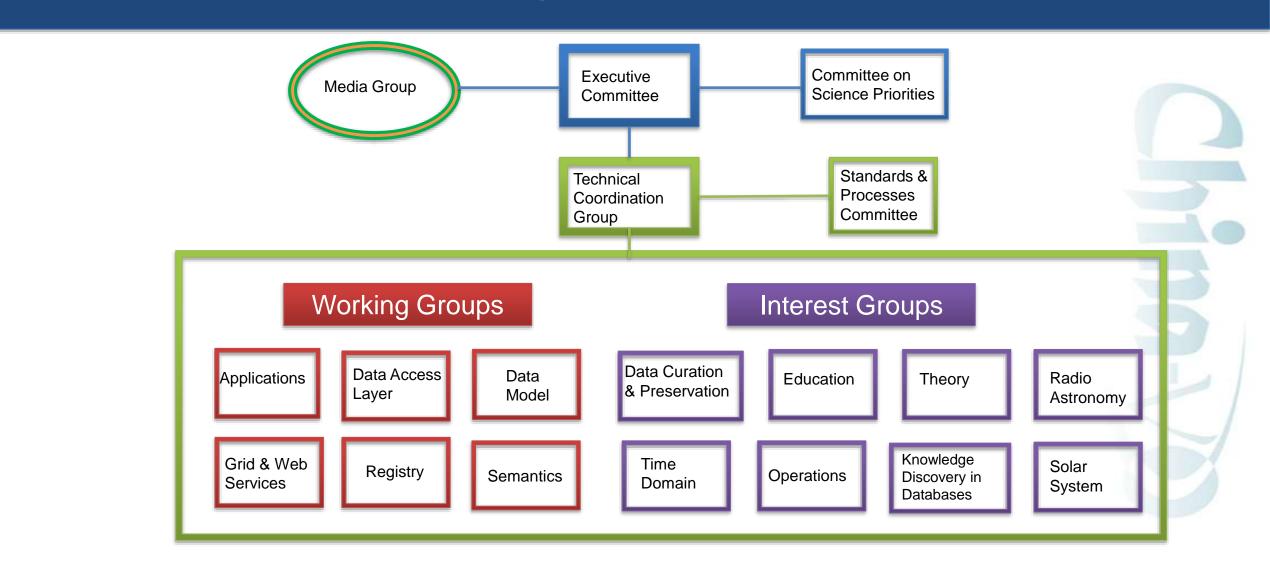








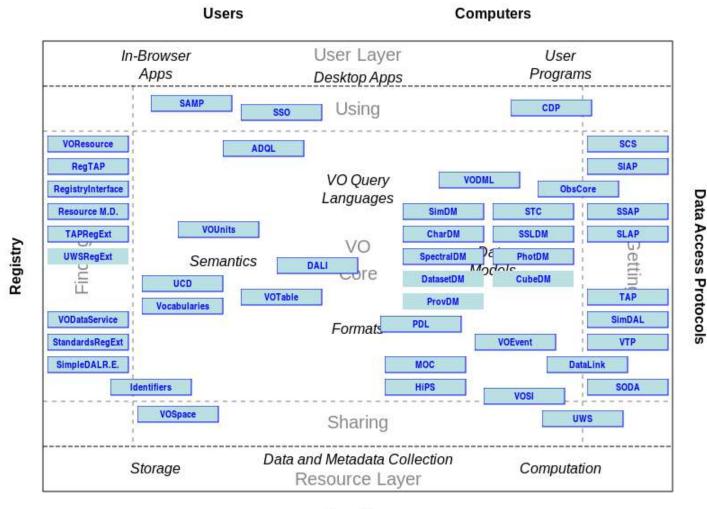
IVOA Organization Chart



Process

IVOA Standards



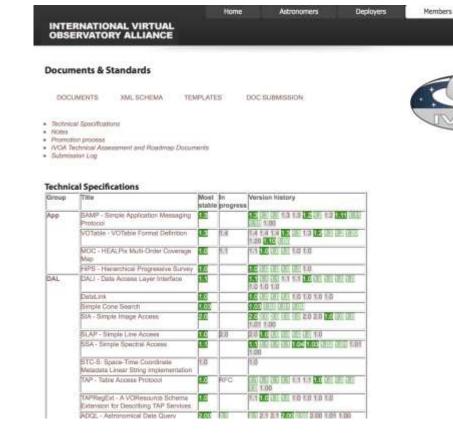


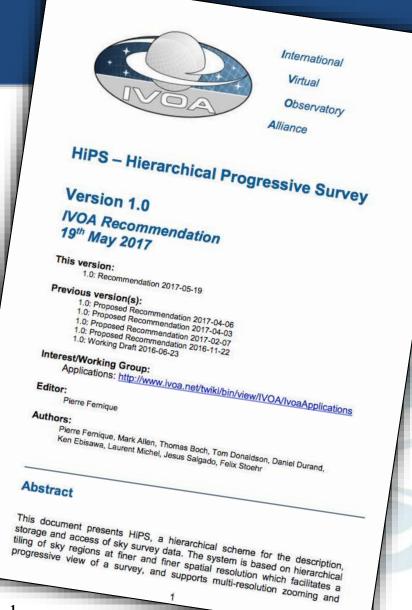
Providers

Standards documents

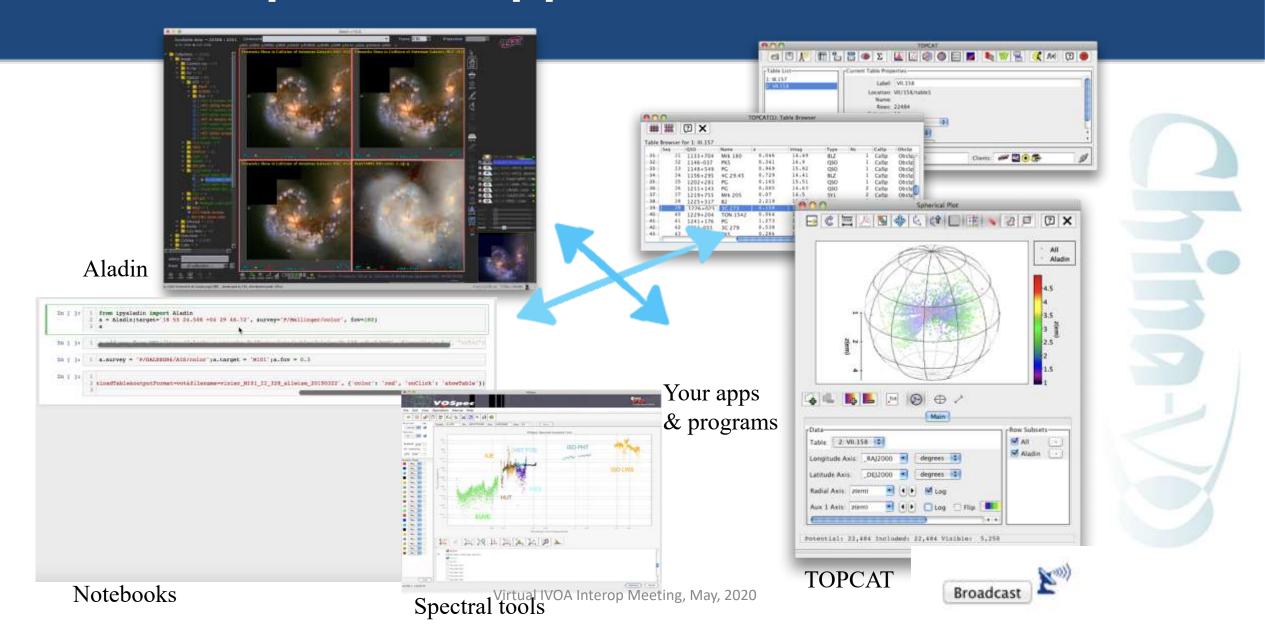
Recommendations, and 'in progress' all available on IVOA pages

Recommendations and Endorsed notes – also on ADS with DOIs

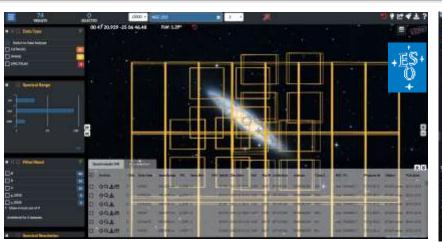




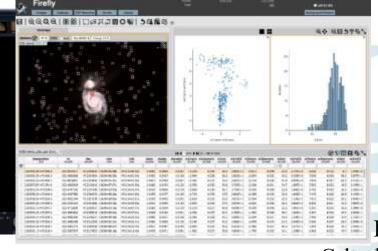
Interoperable applications and services



VO embedded in astronomy services

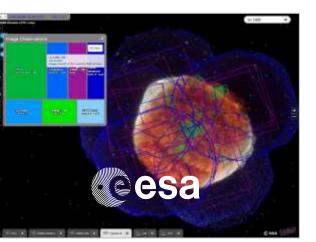






Firefly Caltech-IPAC

ESO Science Portal



Supprises Visional State And State A

WWT



SVO Filter Profile service

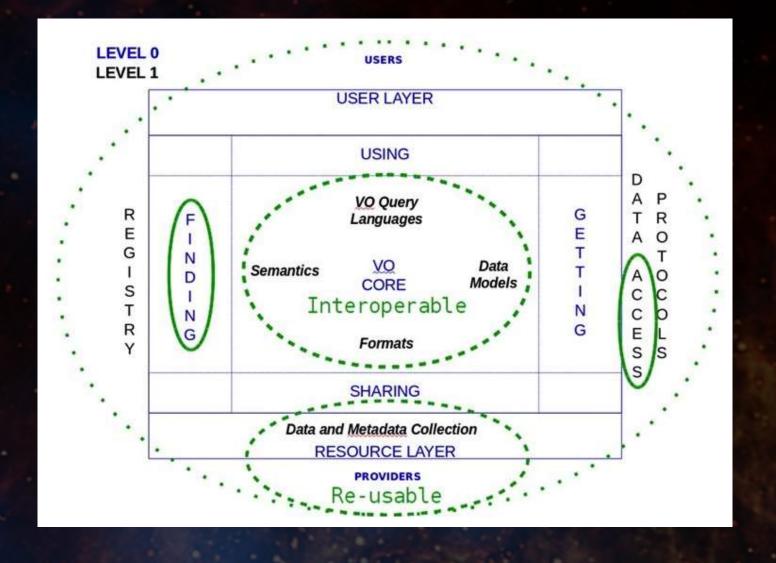
ESA Sky

Grav. wavesal IVOA Interop Meeting, May, 2020

VO is FAIR

Making data:

Findable
Accessible
Interoperable
Reusable





NASA VO & ESA VO

- The US Virtual Observatory Alliance (USVOA) is a US-wide open collaboration endorsed by the AAS to expand and promote VO tools and services. The USVOA is also the US project of the IVOA. So far, participating US institutions include:
 - Association of Universities for Research in Astronomy (AURA)
 - NOAO, LSST, TMT
 - Associated Universities, Inc. (AUI)
 - NRAO
 - California Institute of Technology (Caltech)
 - Harvard Smithsonian Center for Astrophysics (CfA)
 - HCO, SAO, CXC
 - NASA Astronomical Virtual Observatories (NAVO)
 - HEASARC, IPAC, NED, IRSA, STScl





VO-Driven Science Platforms

- The amount of astronomy data will increase greatly in the near future. Science
 platforms are being developed to allow researchers to efficiently analyze big data sets.
 These science platforms enable analysis close to the data, support online data mining
 and machine learning.
- Most science platforms in astronomy employ a similar architecture and technologies to provide an interactive data analysis environment. Basing on a Cloud computing platform, JupyterHub with JupyterLab are used as an interface for exploratory data mining and analysis. The interactive environment is generally deployed using container techniques (e.g., docker).





















Open Science Cloud Platforms

European Open Science Cloud

It is a trusted system providing seamless access to data and interoperable services. It supports the whole research data cycle, from discovery and mining to storage, management, analysis and re-use across borders and disciplines.

African Open Science Platform

The African Open Science Platform initiative (AOSP), funded by the South African Department of Science and Technology (DST) through the National Research Foundation (NRF), and implemented and managed by the Academy of Science of South Africa (ASSAf), is a pan-African project for Africa by Africa. Direction is provided by CODATA (ISC).

GÉANT

 GÉANT is a fundamental element of Europe's e-infrastructure, delivering the pan-European GÉANT network for scientific excellence, research, education and innovation.

Australian Research Data Commons (ARDC)

The ARDC is a transformational, sector-wide initiative, working with sector, government, and industry partners to build a
coherent national and collaborative research data commons. This will deliver a world-leading data advantage, facilitate
innovation, foster collaboration and enhance research translation.

Global Open Science Cloud

 The mission of GOSC is to connect different international, national and regional open science clouds and platforms to create a global digital environment for borderless research and innovation.

Pangeo, ...





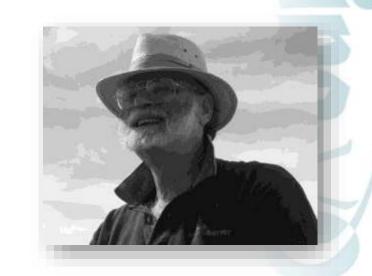


China-VO

- Chinese Virtual Observatory (China-VO) is the national VO project in China initiated in 2002 by Chinese astronomical community leading by National Astronomical Observatories, Chinese Academy of Sciences.
- China-VO became a member of the IVOA with the recommendation of Dr. Jim Gray

R&D Focuses

- China-VO Platform
- Unified Access to On-line Astronomical Resources and Services
- VO-ready Projects and Facilities
- VO-based Astronomical Research Activities
- VO-based Education and Public Outreach



NAOC – Alibaba Cloud Partnership









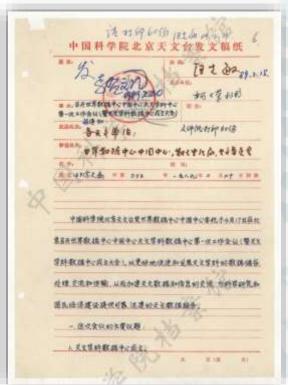


National Astronomical Data Center

National Astronomical Data Center (NADC) was announced by Chinese government in June 2019 as one of the first 20 national scientific data centers.

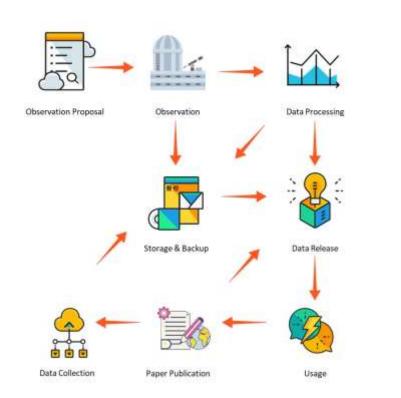




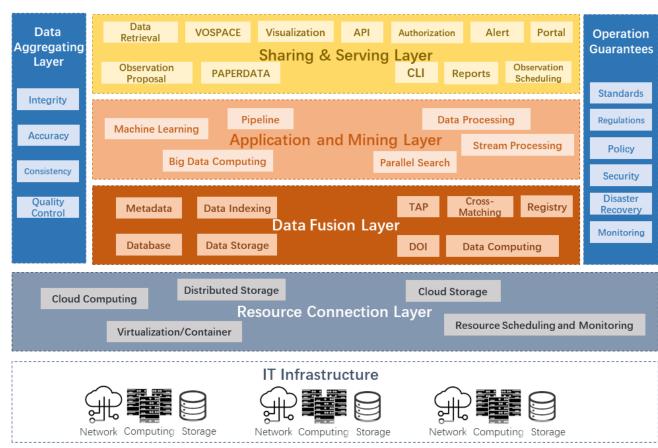


WDC-D for Astronomy was estibalished in April 1989.

Whole life-cycle management









Search

LAMOST DR8 vo --- STATES SCHOOL SECTION

DATA RELEASE 8 v0 Q1+Q2

Markey Barracker Space year Residence Search

Existing Download

FITO Dissesse

Law Resolution

*Instruction	Observed Plates	Total Spectru	Otter	5764 (5% _{2 (0.1} > 10)	Galaxy	990	Unksown	AFGK Star Catalog
(21 (21	121	165,997	139,613	136,331	5.023	1,090	20.271	109,521
(22 (28000)01-8001)	138	230,469	191,721	185,978	9,766	2,789	26,193	144,193

Medium Resolution

	Observed Plates	Total Single Exposure Spectra	Total Coadd Spectra	Total Single Exposure Spectra (SM+10)	Total Coadd Spectra (SM>16)	Parameter Catalog (Coodd)
Q1 (20(0)(00)(291)	144	2,976,882	672,488	1,304,318	458.294	198,259
(0000001-0001)	104	2,767,590	791,982	1.342,126	488,074	211,548

Standard acknowledgement

Supplemental Street Services (See Sept. 1998). See Sept. Sep Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and insuraged by the National Astronomical Observatories, Chinese Asserny of Sciences.

Copyright in 2006-2520; National Althonorecusi Observatoress, Chirarie Academy of Essenses Ader, 20% Descri Rosel, Chaoseny Delect, Beging, China P.C.: 100101 存にP第0500354号-4 文架研究智家号1位105000 E-mail sugnorpelamoning



NADC Resources

- Data storage
- Data access
- Cloud computing and HPC
- Tools and information
- WWT tours and curriculum

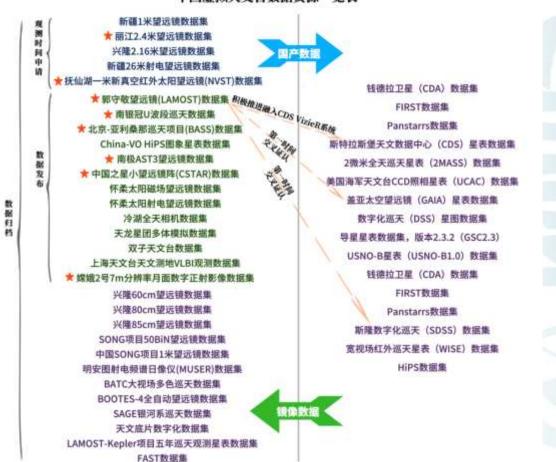




Data resources at the NADC

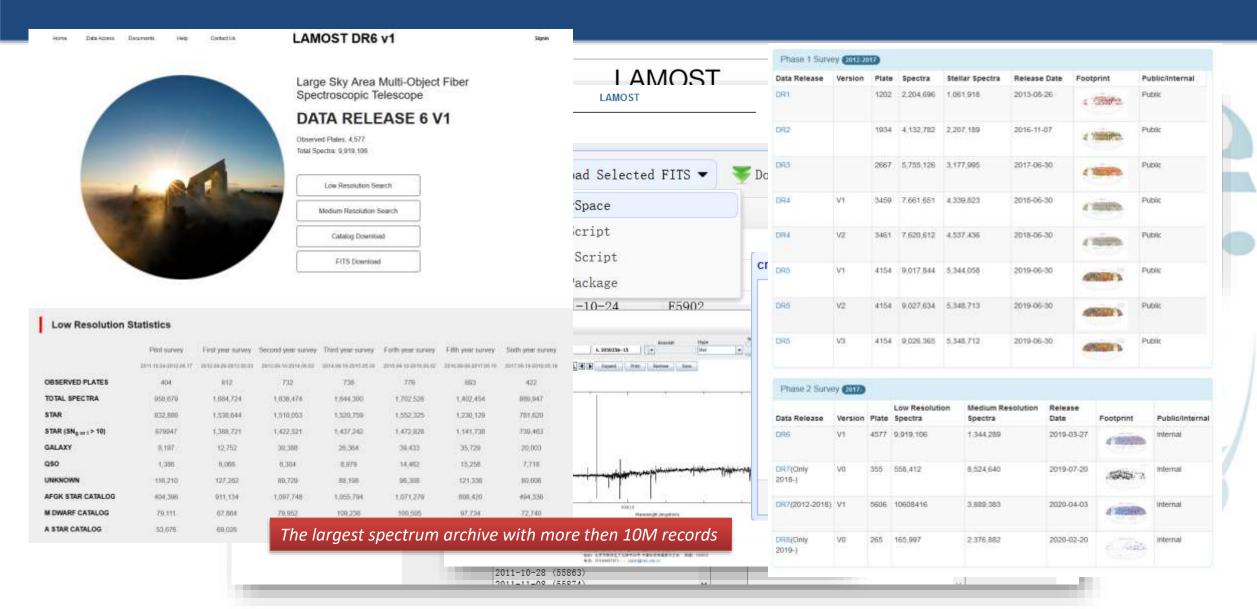


中国虚拟天文台数据资源一览表



注: 加星标的数据为曾获十大天文科技进展项目所产出的数据集

LAMOST Data Access



Cloud Computing and HPC

Based on CloudStack, we set up the cloud computing environment. It consists of 7
distributed nodes across the mainland of China. Based on GlusterFS, we built a
scalable cloud storage system. Each user has a private space, which can be shared
among different virtual machines and desktop systems.

At the headquarter of NAOC, there are 2 HPC clusters.

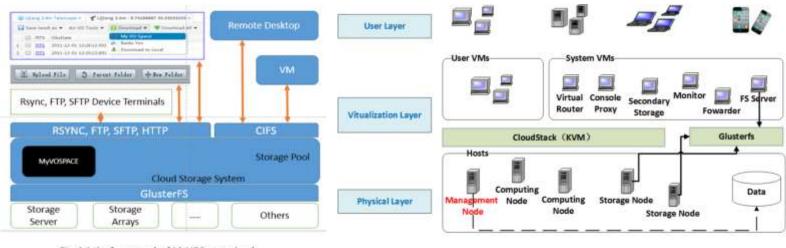


Fig. 4.1 the framework of MyVOSpace, cloud storage system and the relation with other components, technology.

Fig. 2.1 The Architecture of Cloud Computing Environments for AstroCloud



Global Cyber-infrastructure



AND WASTER TO THE PROPERTY OF THE PROPERTY OF

受益于国家天文台和阿里云战略合作,用户以战略合作名义采购阿里云资源,用户将得到所支付金额 20%的云资源代金券赠送。

联系我们: <u>support@china-vo.org</u>, 010-64807973



Tools and Information





欢迎加入<u>公众超新星搜寻项目</u> 开始您的超新星搜寻之旅









Data-driven Education and Public Outreach









Current Status (without Alibaba Cloud)

- Storage:
 - 3000TB
- Computing:
 - 700Tflops+472 cores
- Band width:
 - 155Mbps-300Mbps
 - 10Gbps/1Gbps
- User support:
 - 25 py

- Registered users: 23,200+
- Cloud nodes: 7
- VM instances: 1000+
- Telescope proposals: 1076
- Domestic datasets: 24
- Software environments: ~10



NADC Channels

Observation

- Telescope Proposal Submission and Management
 - 2.4m, 2.16m, Telescope Access Project (TAP), FAST

Data Discovery and Access

- Virtual Observatory
- Seamless astronomy

Services

- Conferences, PaperData, Astro-term, VO tools
- Cloud Computing
 - VM and Containers
 - data processing and analyzing environments
 - Cloud storage
- Public channel
 - Public Supernova search Project (PSP)
 - Worldwide Telescope
 - Eclipse caculator
 - CosmoStation



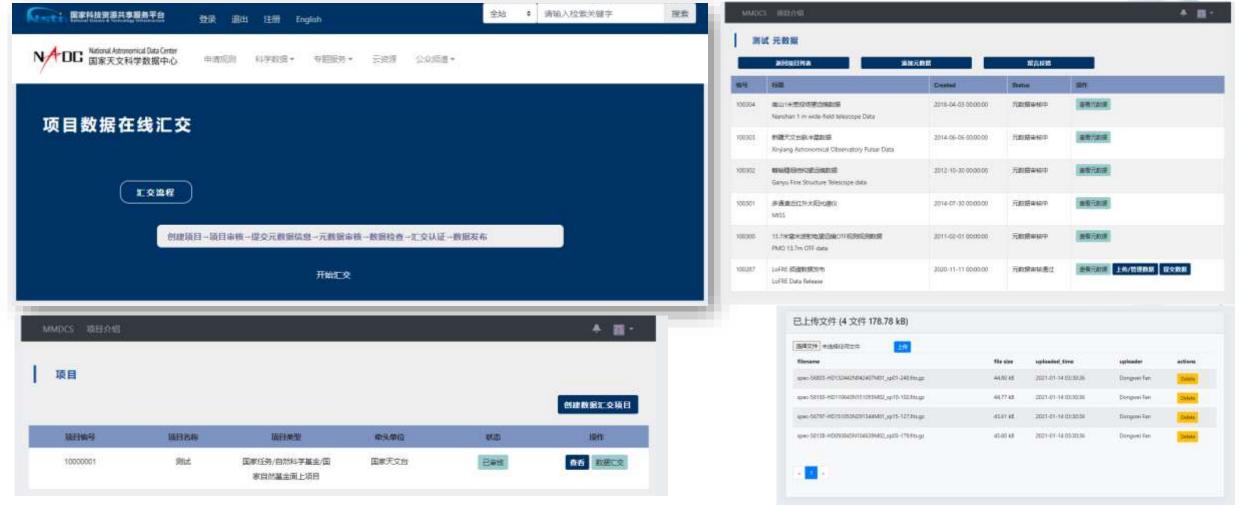
NADC Services

- Data archiving and management
- Data release and open access
- Software/system/platform design and development
- Cluster/HPC design
- Community outreach
- Education and public outreach



Data Archiving and Management

Metadata Management and Data Submission System (Alpha version)



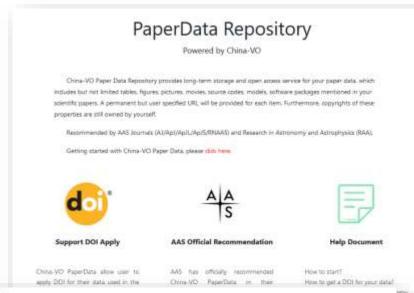
NADC PaperData

- A long-term storage and open access service for author's paper data
 - Stable storage
 - Long-term accessible URL
 - DOI support

Recognized Persistent Repositories for AAS

Journals





Astronomy Specific Repositories

The following astronomy archives offer data preservation platforms that mint DOIs for specific categories of data. These archives may or may not be available for your particular project:

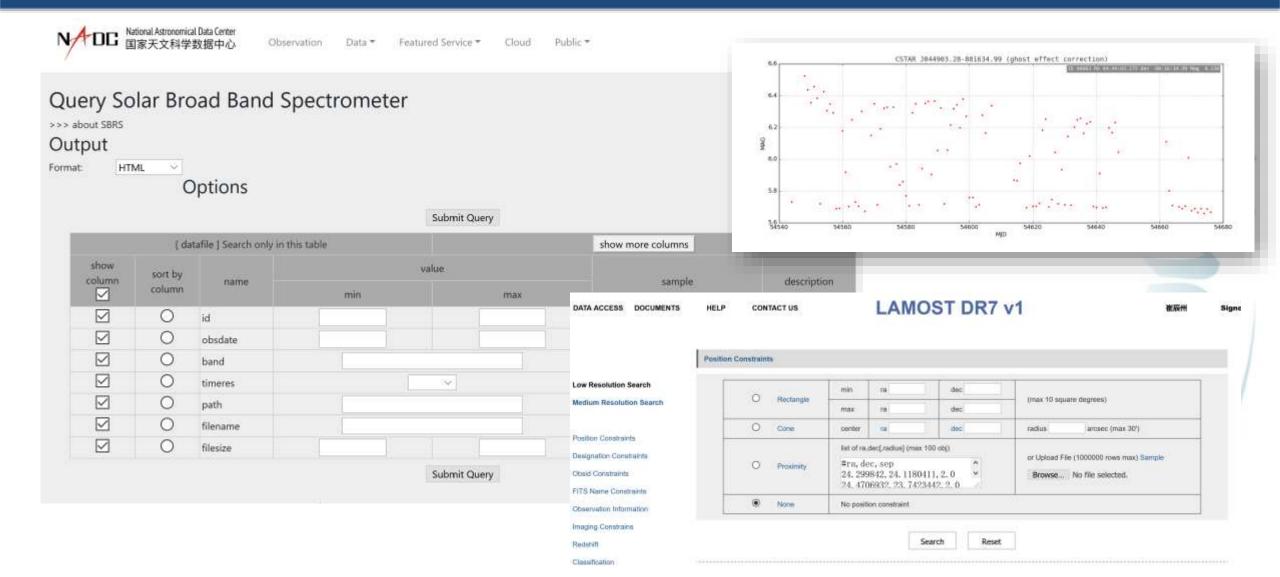
- . Caltech/NASA IPAC; [List of all released data DOIs]
- Canadian Advanced Network for Astronomical Research (CANFAR); [List of all released data DOIs]
- China-VO Paper Data Repository; [List of all released data DOIs]
- Mikulski Archive for Space Telescopes (MAST); Contribution Guidelines for HLSPs; [List of all released data DOIs]

Generic Repositories

There are a number of open "generic" data repositories that serve all fields from communities around the world. We curate collections of Journal related data in some of them:

- Zenodo (Contribute to the AAS Journals' Zenodo Community)
- Harvard Dataverse (Contribute to the AAS Journals' Dataverse)
- figshare

Data release and open access



software/system/platform design and development

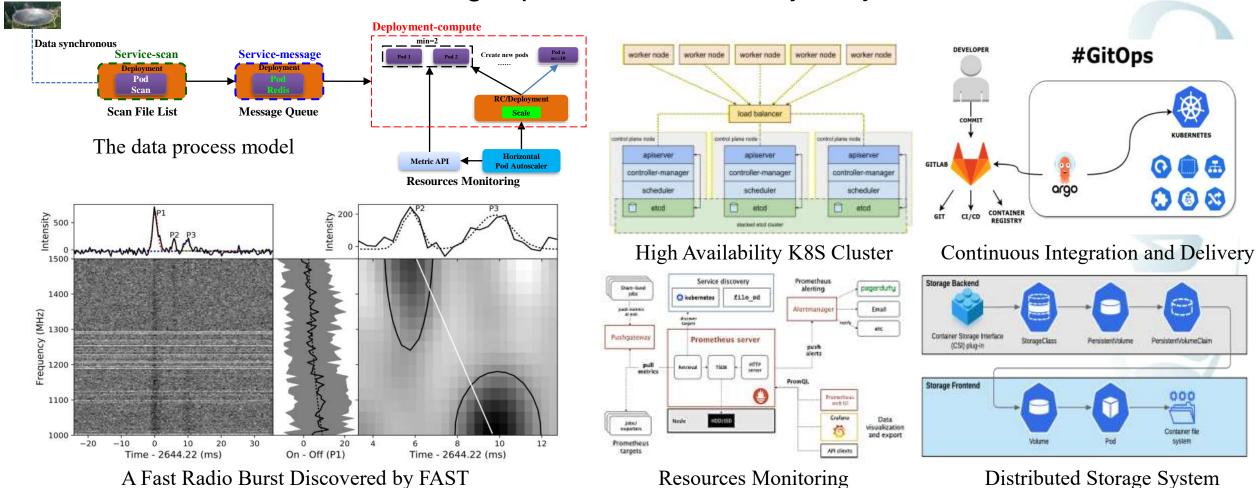
FAST Information System GW Alert Transient Identification **GW Host Galaxy Filter Image Subtraction** Resolving GW Localization Data OT Extraction **Database Retrieving Galaxies** OT Classification Coverage in GW Localization Region **Data Visualization** Filtering Host Galaxy Candidates **GW** Localization Telescopes Notification to Follow-up Observation Data Observation

The architecture and data processing stream of GWOPS



Docker-based HPC

Build Data Processing Pipeline for FAST Key Project and SVOM



Cost on Commercial Cloud Platforms

Computing is cheap, storage is expensive, bandwidth is very expensive

业务类型	核心需求	推荐星级
具有明确时限的计算任务	短时间算力	****
项目和团队网站	安全稳定的网络	****
协同工作环境	安全稳定的网络	****
大用户量但中小体量的数据发布(<500TB)	安全稳定的网络和数据环境	****
中小数据体量的在线服务系统(<100TB)	安全稳定的网络和数据环境	****
中小数据体量的算法和软件测试开发(<100TB)	算力和平台的灵活部署	***
计算密集型的数据分析处理	算力和一定的存储能力	***
大规模数据的分析挖掘(>500TB)	算力和大规模数据存储	**
国际数据大规模传输	国际带宽	**
PB+级别数据的长期存储	大规模数据存储	*
个人资料备份	安全的数据存储	*

Community Outreach



- China-VO Wechat
- NADC website



Education and Public Outreach













Big Data + Cloud Computing +



= science platform

It is not a talk about Science, but one for Better Science

FREE is good, but not always.

- https://nadc.china-vo.org
- ccz@bao.ac.cn

Wechat: China-VO



Special Thanks To

- MOST, CAS
- NSDC community, CNIC

• IVOA, WDS, CODATA, RDA

NADC builders and partners

