Archiving for Different Source Astronomical Data: Experiences and Lessons Learned

Youfen Wang , Yihan Tao
NADC Data Resource Curation Group
2024 Spring IVOA
May 23 2024, Australia, Sydney

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

Out Line

1. Five Types of Data
2. Meta Data
3. CSTR and DOI Registration
4. Summary

Experiences & Lessons

Five Type of Data

01 General Dataset

Data produced by Chinese astronomical telescopes and instruments

02 Paper Data

paper data accepted by astronomical journals

03 Scientific Project Data

Scientific project data supported by government funding

04 Mirror Data

Data from astronomical telescopes and instruments worldwide

05 Outreach Data

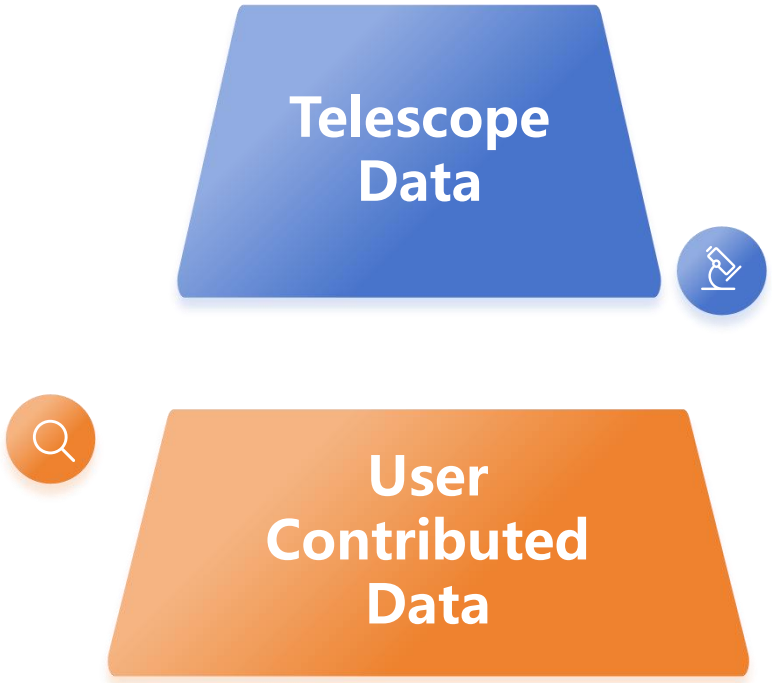
Astronomical science and education data

06 Statistics

Statistics of the data

General Dataset

- catalogs provided by the user
- not official release



- official release
  - update regularly
  - update irregularly

General Dataset-Telescope Data

01 LAMOST

Data save in NADC server each release with 10-20 metadata

Now have LAMOST DR1-DR5(Low resolution) , DR6-DR12(both low resolution and medium resolution)

There are Q1 Q2 Q3 V0 V1 V2 each release, V2 is for international access

02 FAST

Data saved in FAST server, programs and observed sources available

Update every three months

Now have DR1 to DR17 each release with one metadata

updata regularly upto now

General Dataset-Telescope Data

01 Antarctic AST3 Telescope

02 Beijing-Arizona Sky Survey

03 The South Galactic Cap u-band Sky Survey

04 Lijiang 2.4m Telescope

05 The First Release of the CSTAR

06 Digitized Data of China Astronomical Plates

More..... update irregularly

General Dataset-User Contributed Data

M subdwarf catalog	LAMOST Spectra Classification Dataset	Stellar Parameters for over 20 Million Stars from SAGES DR1	SAGES Data Release 1 & 1s	FASHI Data Release 1	The first comprehensive Milky Way stellar mock catalogue for the Chinese Space Station Telescope Survey Camera
1	2	3	4	5	6

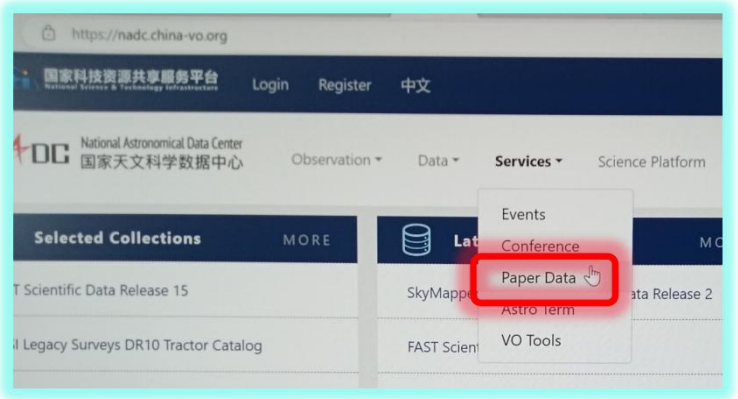
Paper Data

https://nadc.china-vo.org/

services

paper data

how access?



Chart

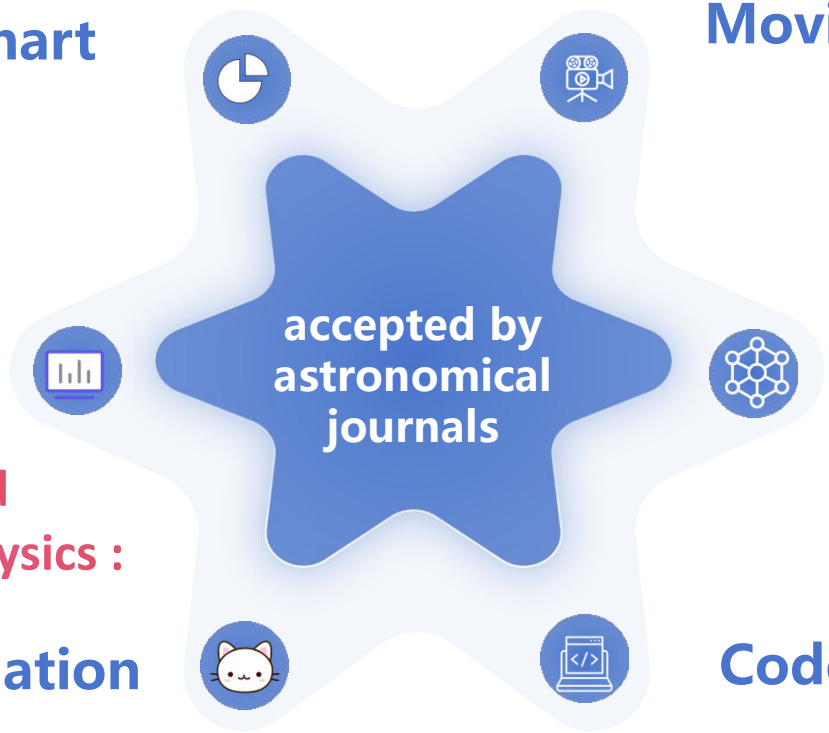
Movie

Data

Model

Animation

Code



AAS Journals :
(AJ/ApJ/ApJL/ApJS/RNAAS) and
Research in Astronomy and Astrophysics :
RAA

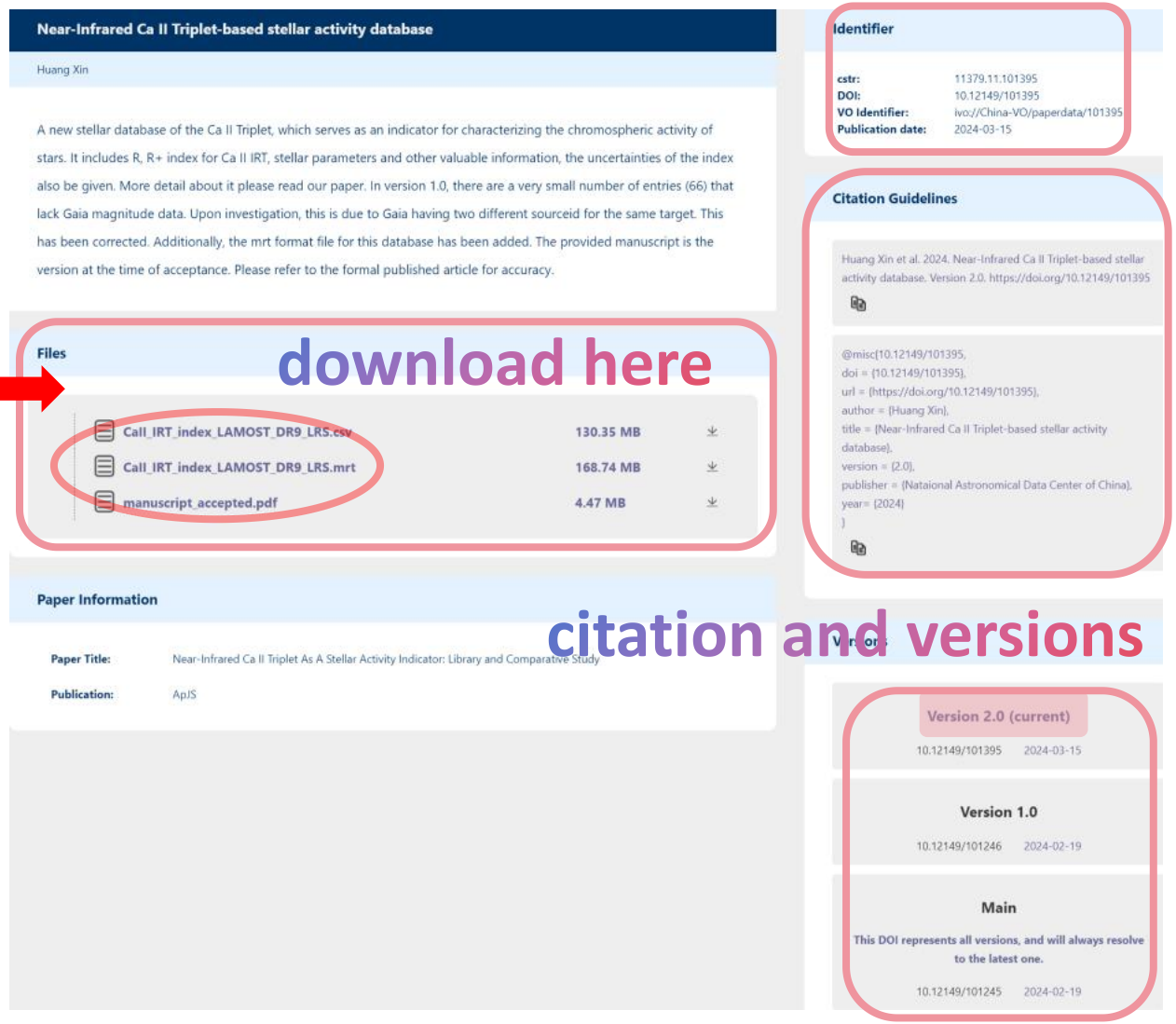
provides long-term storage and open access service

Paper Data

fill in metadata	title of data set and paper, data description, journal name
upload evidence	acceptance letter, full paper
calibrate data	standard format of the data(format asked by the journaled)
upload data	Vol. < 5GB on the web, Vol. > 5 GB rsync
registration of DOI CSTR	complete within two working days
update data version	if need to change data, update new version

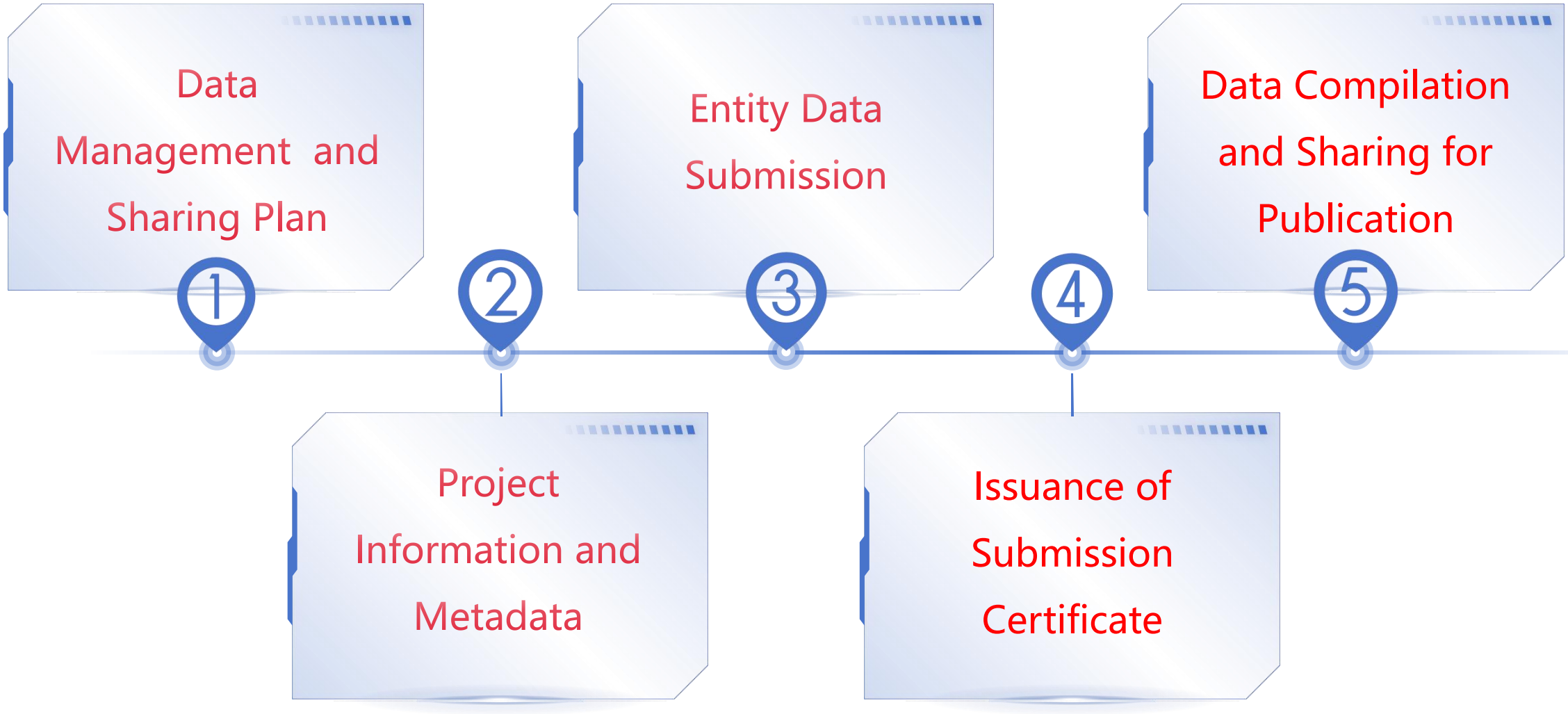
Paper Data

click ‘datasources’ get to the paper data



main version direct to the newest,
and the subversion direct to itself

Scientific Project Data



Scientific Project Data

01	02	03	04	05
scientific data meta data and assistent software	information of project, and metadata	upload and rating of data	certificate : number of files, data volume , number of dataset	after reach of publish date

10 scientific projects, involving 99 datasets, with a data volume of 329TB!

Outreach Data

nine datasets

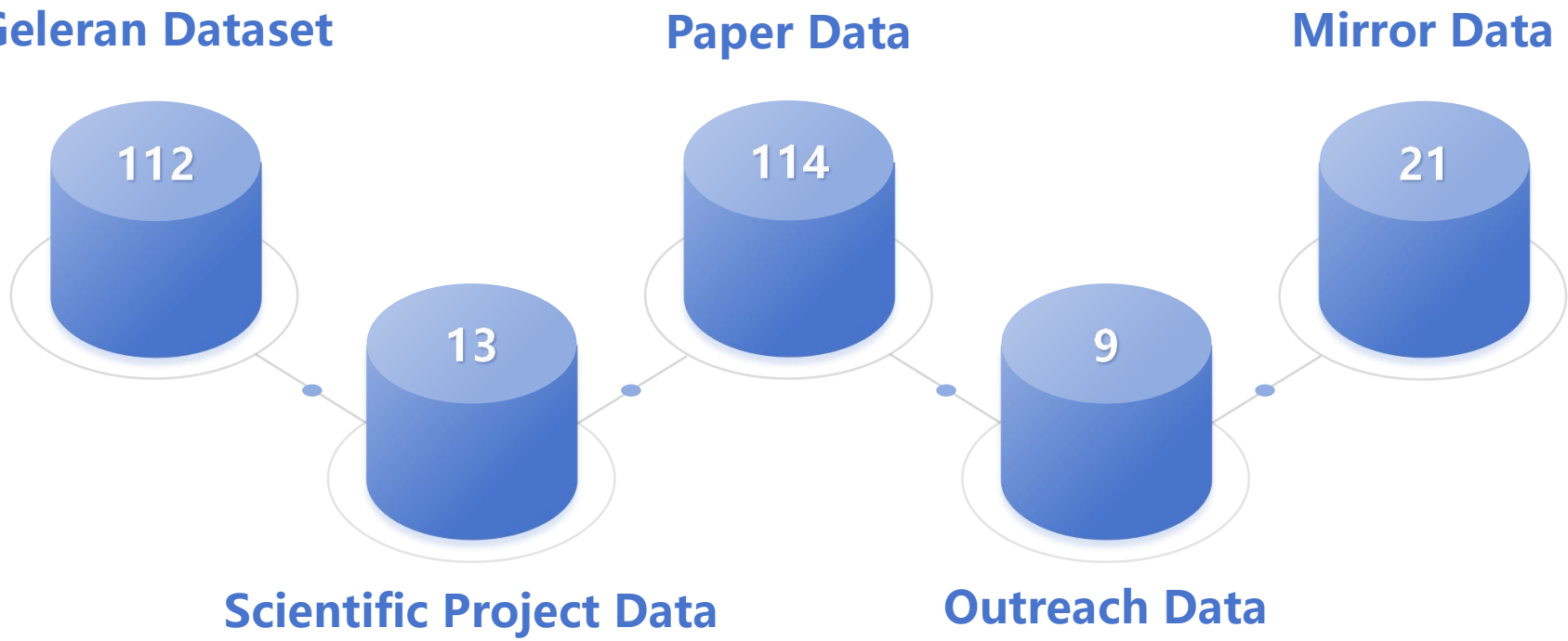
- Precursor—Guo Shoujing' s Astronomical Achievements
- Eclipse Calculator
- Recommended Chinese Names of Mars Features
- Recommended Chinese Names of Mars Features (2nd release)
- Atlas of remote sensing images of Jilin-1 of world-famous observatories
- Map of the surface of the Moon (the Second Edition)
- Four seasons star chart
- Multiband Astronomical Observations and Data Processing Graduate Course Optical Infrared Data Processing Data
- Map of the surface of the Moon (the Third Edition)

Mirror Data

21 datasets

- U.S. Naval Observatory CCD Astrograph Catalog(UCAC1-5)
- Wide-field Infrared Survey Explorer （AllSky All WISE CatWISE）
- Sloan Digital Sky Survey(DR12 DR14)
- The Pan-STARRS1 Data Release 1
- GAIA Data Release 1-DR2-EDR3 -DR3
- DESI Legacy Surveys Tractor Catalog(DR9 DR10)
- 2MASS All-Sky Data Release
- AAVSO Photometric All Sky Survey (APASS) Data Release 10
- Distances to 1.47 billion stars in Gaia EDR3 （2021）
- SkyMapper Southern Sky Survey Data Release 2

Statistics



first level

Meta data

43 items for the core metadata: including 22 necessary and 21 not necessary

will update metadate for different data type

e.g. paperdata

表 2-1: 核心元数据

字段	定义	必选	说明
title	VARCHAR	Y	数据集名称（中英文）
shortname	VARCHAR	N	短名称（英文）
description	TEXT	Y	说明文档（中英文）
create_date	DATE	Y	数据生产日期
update_date	DATE	Y	数据更新日期
publicate_date	DATE	N	（预计）发布日期
authors	VARCHAR	Y	数据作者、生产者信息（中英文）
affiliation	VARCHAR	Y	单位（中英文）
address	VARCHAR	Y	作者地址（中英文）
email	VARCHAR	Y	电子邮件
telephone	VARCHAR	Y	电话
contributors	VARCHAR	N	数据贡献者（中英文）
cb_affiliation	VARCHAR	N	数据贡献者单位（中英文）
cb_address	VARCHAR	N	数据贡献者地址（中英文）
cb_email	VARCHAR	N	数据贡献者电子邮件
cb_telephone	VARCHAR	N	数据贡献者电话
contacts	VARCHAR	N	数据联系人（中英文）
ct_affiliation	VARCHAR	N	数据联系人单位（中英文）
ct_address	VARCHAR	N	数据联系人地址（中英文）
ct_email	VARCHAR	N	数据联系人电子邮件

表 2-1 – 接上页

字段	定义	必选	说明
ct_telephone	VARCHAR	N	数据联系人电话
keywords	VARCHAR[]	Y	关键词（3-5 个）
type	VARCHAR[]	Y	数据实体类型
waveband	VARCHAR[]	Y	数据波段
project	VARCHAR	N	数据来源项目
facility	VARCHAR	N	生产装置
subjects	VARCHAR[]	Y	子学科
production_type	VARCHAR	Y	数据生产方式
data_type	VARCHAR[]	Y	数据格式类型
data_level	VARCHAR	Y	数据级别
production_age	VARCHAR	Y	生产年代
content_level	VARCHAR[]	Y	内容级别
data_class	VARCHAR	Y	数据发布分类
share_method	VARCHAR[]	Y	共享途径
share_scope	VARCHAR	Y	共享范围
application_procedure	TEXT	N	申请流程files
	INT	Y	文件数
filesize	INT	Y	文件量（单位字节）
catalogs	INT	N	星表数目
catalog_rows	INT	N	星表记录总数目
usage	TEXT	N	数据使用说明
bibcode	VARCHAR	N	相关的论文编码
URL	VARCHAR[]	N	相关网址
acknowledgement	TEXT	N	致谢文本（中英文）

CSTR and DOI Registration

China Science and Technology Resources

CSTR.11379.11.100436

- 11379 — Chinese Science and Technology Resource Code
- 11 — Registration Authority Code
- 100 — Codes for Types of Scientific and Technical Resources
- 436 — Unique ID in Data Base

Codes for Types of Scientific and Technical Resources

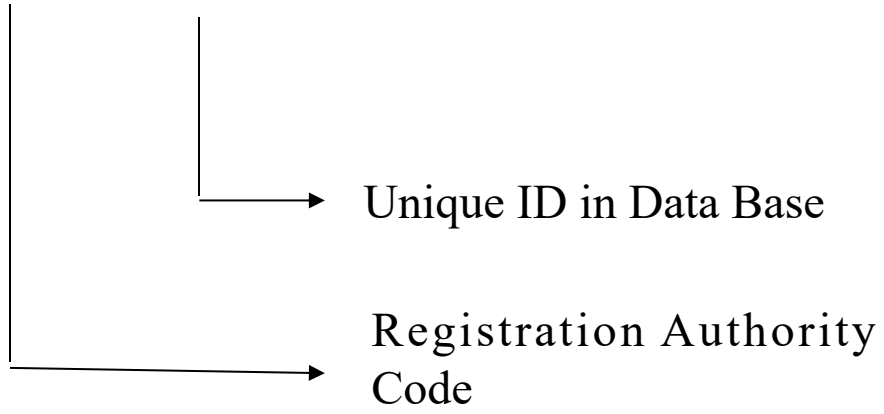
1	11	Scientific Data
2	19	Computer Software
3	21	Popular Science Resources
4	99	Other Scientific and Technical Resources

NO CSTR for mirror data!

CSTR and DOI Registration

Digital Object Identifier

10.12149/101179



- registration complete usually within two working days
- no DOI registration for mirror data

Summary

Five data types:Geleran Dataset, Scientific Project Data, Paper Data, Outreach Data, Mirror Data

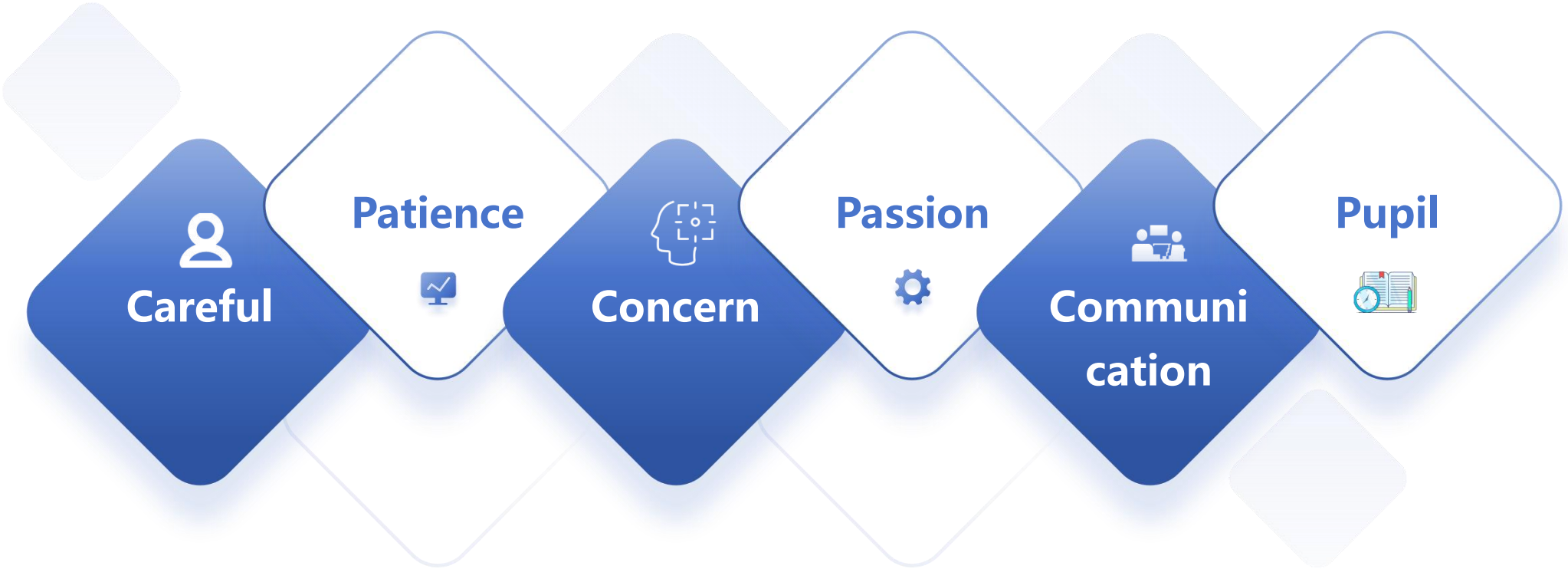
There are 269 datasets in totle: 112 13 114 9 21

There are 43 items for the core metadata

CSTR and DOI are provided, except for mirror data

1 2 3 4

Experiences & Lessons



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

email:yfwang@bao.ac.cn