



# NADC Metadata Management for Astronomical Data: standards and system

Yihan Tao, Zhiying Huo, Boliang He, Linying Mi, Chenzhou Cui  
@ IVOA Northern Spring Interop meeting, 25-28 May

# Outline

- Metadata standards
  - Data categories
  - Metadata
  - Multifaceted taxonomy
  - Metadata granularity
- Metadata Management & Data Submission System
  - Data catalogue
  - Research project data submission
- Discussion - Data quality assessment and metadata



# Data Categories in NADC

Observation data  
191

Mirror data  
19

Simulation data  
1

Data from funded project  
5

Paper data  
23

EPO data  
56



# Metadata

- Mostly in consistent with Resource Metadata for VO
- Essential
  - Identity metadata
    - title, shortname, identifier (IVOID, DOI, cstr)
  - Curation metadata
    - Authors (creator), creation date(date), updated date, publicated date, share method, share scope, application procedure
  - General content metadata -> important for data discovery
    - description, keywords (subject), tags of data types, number of files, filesize, bibcode (source), URL (ReferenceURL), acknowledgement (rights)
- Specific
  - Paper information
  - Project information

**LAMOST Data Release 8 V1.0** Publish Time:2021-03-31

LAMOST Data Release 8 V1.0 includes spectra and catalogs obtained by LAMOST low/medium resolution survey during October 24th 2011 and May 25th 2020. For the low resolution survey, there are 5,207 plates observed, 10,388,423 stellar spectra, 219,776 galaxy spectra, 71,786 quasar spectra and 534,091 unknown object spectra. For the medium resolution survey, there are 6,038,218 spectra, among them 1,479,127 non time-domain spectra, 4,599,091 time-domain spectra. In addition, there are dozens of catalogs list spectral parameters for the low/medium resolution survey respectively.

**Data access**

[Data Website](#)

**INTRODUCTION**

Data format	application/fits , text/csv , image/png
Data volume	11 tables, 69936029 rows, 1881250.07 MB
Sharing method	online
Sharing scope	Share with conditions
DOI	10.12149/100361
CSTR	11379.11.100361
VO Identifier	ivo://China-VO/data/LAMOST/DR8/V1.0

**CONTACT AUTHOR**

Author	LAMOST Operation and Development Center
Email	support(at)lamost.org
Institution	National Astronomical Observatories, Chinese Academy of Sciences

**CONTACT PUBLISHER**

Email	support(at)china-vo.org
Phone	+86-10-64807973
Institution	National Astronomical Data Center

**TAGS**

<b>waveband</b>	Optical
<b>telescope or project</b>	Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)
<b>subject</b>	Cosmology and Galaxies Star and Interstellar Matter Astronomical Technology and Methods
<b>data type</b>	Spectrum Data Catalog Data

# Multifaceted Taxonomy

- Reflected in the metadata as tags of data types

- **Waveband**

( reference: Resource Metadata for the Virtual Observatory Version 1.12)

- Radio, Millimeter, Infrared, Optical, Ultraviolet, Xray, Gamma-ray

- **Telescope/Project**

- LAMOST, FAST, MUSER, The Tianlai project, BASS, NVST, Tian Ma Telescope...
- can add on to the list as needed

- **Subject**

- Cosmology and galaxies, Stars and interstellar matter, Heliophysics, Planetary science, Basic astronomy, Astronomical techniques and methods

- **Data Product Type**

( reference: IVOA Observation Data Model Core Components and its Implementation in the Table Access Protocol)

- Image, Cube, Photometry, Spectrum, SED, Timeseries, Visibility, Event, Measurements, Log, Ambient, Digitalised, Accessories



# Multifaceted Taxonomy

- **Production Age**

- Ancient, 1900-1950, 1950-1980, after 1980

- **Process Level**

( reference: IVOA Observation Data Model Core Components and its Implementation in the Table Access Protocol)

- Level 0 (raw data), Level 1, Level 2, Level 3, level 4

- **Content Type**

( reference: Resource Metadata for the Virtual Observatory Version 1.12)

- Animation, Archive, Artwork, Background, BasicData, Bibliography, Catalog, EPOResource, Education, Historical, Journal, Library, Organisation, Other, Outreach, Photographic, Press, Project, Registry, Simulation, Survey, Transformation

- **Content Level**

( reference: Resource Metadata for the Virtual Observatory Version 1.12)

- Research, General, Amateur



# Metadata Granularity

- Followed the hierarchical structure of data collections
- Dataset, sub-datasets and tables

LAMOST Data Release 8 V1.0		Publish Time:2021-03-31
◆ Sub-datasets and Tables ◆		
LAMOST Data Release 8 V1.0 -- Low Resolution		
LAMOST DR8 v1.0 LRS General Catalog	11214076 rows	1952.57 MB
LAMOST DR8 v1.0 LRS A, F, G and K Type Star Catalog	6478063 rows	1472.58 MB
LAMOST DR8 v1.0 LRS A Type Star Catalog	629041 rows	186.87 MB
LAMOST DR8 v1.0 LRS M Star Catalog	773721 rows	292.48 MB
LAMOST DR8 v1.0 LRS Observed Plate Information Catalog	5207 rows	0.28 MB
LAMOST DR8 v1.0 LRS Input Catalog	20824500 rows	1788.50 MB
LAMOST DR8 v1.0 LRS Multiple Epoch Catalog	2054302 rows	394.98 MB
LAMOST Data Release 8 V1.0 -- Medium Resolution		
LAMOST DR8 v1.0 MRS General Catalog	22356885 rows	3604.84 MB
LAMOST DR8 v1.0 MRS Parameter Catalog	1243895 rows	441.42 MB
LAMOST DR8 v1.0 MRS Observed Plate Information Catalog	1089 rows	0.08 MB
LAMOST DR8 v1.0 MRS Input Catalog	4355250 rows	323.94 MB

Antarctic AST3 Telescope Data Release 1		Publish Time:2019-06-30
◆ Sub-datasets and Tables ◆		
Antarctic AST3 Telescope Data Release 1 - Image		
Antarctic AST3 Telescope Data Release 1 - Catalog		
Antarctic AST3 Telescope Data Release 1 - Light Curve		



# Metadata Management System and Data Catalogue

All categories (Total number of datasets: 50) Hide categories

- waveband** Radio (5) Optical (42)
- telescope or project** Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) (30)  
Five-hundred-meter Aperture Spherical radio Telescope (FAST) (5) Beijing-Arizona Sky Survey (BASS) (4)  
New Vacuum Solar Telescope (NVST) (1) Antarctic Survey Telescope (AST3) (1) Solar Magnetic Field Telescope (SMFT) (1)  
Lijiang 2.4m Telescope (1) China Astronomical Plates Data (1)
- data type** Spectrum Data (36) Catalog Data (38) Accessories Data (1)

published time ↑

## FAST Scientific Data Release 5

Publish Time: 2021-05-01

DOI: 10.12149/100400

CSTR: 11379.11.100400

## LAMOST Data Release 7 V1.3

Publish Time: 2021-04-20

DOI: 10.12149/100381

LAMOST Data Release 7 V1.3 includes spectra and catalogs obtained by LAMOST low/medium resolution survey during October 24th 2011 and June 8th 2019. For the low resolution survey, there are 10,640,255 spectra obtained totally, including 9,881,260 stellar spectra, 198,393 galaxy spectra, 66,406 quasar spectra, and 494,196 unknown object spectra, and related catalogs published. For the medium resolution survey, there are totally 992,669 coadded spectra obtained by the non time-domain survey, and 2,861,575 single exposure spectra obtained by the time-domain survey, and related catalogs published. All data products are available from the website <http://dr7.lamost.org/v1.3/>.

## LAMOST Data Release 8 V1.0

Publish Time: 2021-03-31

DOI: 10.12149/100361

CSTR: 11379.11.100361

LAMOST Data Release 8 V1.0 includes spectra and catalogs obtained by LAMOST low/medium resolution survey during October 24th 2011 and May 25th 2020. For the low resolution survey, there are 5,207 plates observed, 10,388,423 stellar spectra, 219,776 galaxy spectra, 71,786 quasar spectra and 534,091 unknown object spectra. For the medium resolution survey, there are 6,038,218 spectra, among them 1,479,127 non time-domain spectra, 4,599,091 time-domain spectra. In addition, there are dozens of catalogs list spectral parameters for the low/medium resolution survey respectively.

## LAMOST Data Release 8 V1.0

Publish Time: 2021-03-31

LAMOST Data Release 8 V1.0 includes spectra and catalogs obtained by LAMOST low/medium resolution survey during October 24th 2011 and May 25th 2020. For the low resolution survey, there are 5,207 plates observed, 10,388,423 stellar spectra, 219,776 galaxy spectra, 71,786 quasar spectra and 534,091 unknown object spectra. For the medium resolution survey, there are 6,038,218 spectra, among them 1,479,127 non time-domain spectra, 4,599,091 time-domain spectra. In addition, there are dozens of catalogs list spectral parameters for the low/medium resolution survey respectively.

### Data access

Data Website

### INTRODUCTION

Data format	application/fts, text/csv, image/png
Data volume	11 tables, 69936029 rows, 1881250.07 MB
Sharing method	online
Sharing scope	Share with conditions
DOI	10.12149/100361
CSTR	11379.11.100361
VO Identifier	ivo://China-VO/data/LAMOST/DR8/V1.0

### Sub-datasets and Tables

Dataset Name	Rows	Size
LAMOST Data Release 8 V1.0 -- Low Resolution		
LAMOST DR8 v1.0 LRS General Catalog	11214076 rows	1952.57 MB
LAMOST DR8 v1.0 LRS A, F, G and K Type Star Catalog	6478063 rows	1472.58 MB
LAMOST DR8 v1.0 LRS A Type Star Catalog	629041 rows	186.87 MB
LAMOST DR8 v1.0 LRS M Star Catalog	773721 rows	292.48 MB
LAMOST DR8 v1.0 LRS Observed Plate Information Catalog	5207 rows	0.28 MB
LAMOST DR8 v1.0 LRS Input Catalog	20824500 rows	1788.50 MB
LAMOST DR8 v1.0 LRS Multiple Epoch Catalog	2054302 rows	394.98 MB
LAMOST Data Release 8 V1.0 -- Medium Resolution		
LAMOST DR8 v1.0 MRS General Catalog	22356885 rows	3604.84 MB
LAMOST DR8 v1.0 MRS Parameter Catalog	1243895 rows	441.42 MB
LAMOST DR8 v1.0 MRS Observed Plate Information Catalog	1089 rows	0.08 MB
LAMOST DR8 v1.0 MRS Input Catalog	4355350 rows	333.94 MB

### CONTACT AUTHOR

Author	LAMOST Operation and Development Center
Email	support@lamost.org
Institution	National Astronomical Observatories, Chinese Academy of Sciences

### CONTACT PUBLISHER

Email	support@china-vo.org
Phone	+86-10-64807973
Institution	National Astronomical Data Center

### TAGS

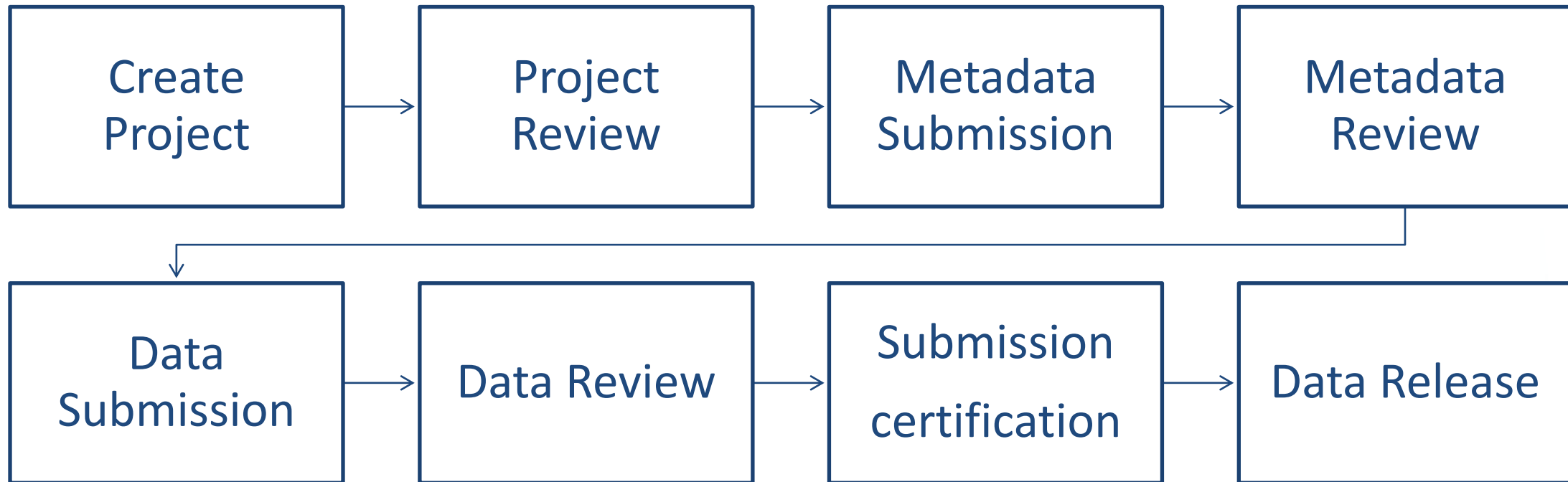
waveband	Optical
telescope or project	Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)
subject	Cosmology and Galaxies Star and Interstellar Matter Astronomical Technology and Methods
data type	Spectrum Data Catalog Data
production age	1980-
content level	Research





# Project Data Submission System

The system is aimed at supporting the process of data submission of national key research and development project, which is one of the important responsibilities of NADC.



# Research Project Data Submission

MMDCS Guide

芯片原子钟技术 Metadata

[Add Metadata](#) [Message](#) [Return to Project List](#)

ID	Title	Last Modified	Status	Operations
100395	芯片原子钟输出时间和频率数据 Time and Frequency Output Data of the Chip Scale Atomic Clock	2021-04-21	Data under review	<a href="#">View Metadata</a>
100399	数据计算辅助软件 Assistant Software for Data Calculation	2021-04-21	data approved	<a href="#">View Metadata</a>
100396	芯片原子钟输出秒脉冲与UTC的差值数据 Difference between the 1PPS Output of the Chip Scale Atomic Clock and the Coordinated Universal Time (UTC)	2021-04-21	data approved	<a href="#">View Metadata</a>
100397	芯片原子钟16天输出频率数据 Frequency Data of the Chip Scale Atomic Clock Measured for 16 Days	2021-04-21	data approved	<a href="#">View Metadata</a>
100398	芯片原子钟5000秒输出频率数据 Frequency Data of the Chip Scale Atomic Clock Measured for 5000 Seconds	2021-04-21	data approved	<a href="#">View Metadata</a>



# Data Quality Assessment ?

- Expert committee review
- Detailed Criteria?
- Best practices?

*DataQuality* (char)

Definition: An overall assessment of the integrity, consistency, and level of documentation concerning uncertainty estimates and calibration procedures, of the data provided by the resource. We suggest 3 general grade levels, plus codes for unknown or undocumented cases:

- A Data are fully calibrated, fully documented, and suitable for professional research.
- B Data are calibrated and documented, but calibration quality is inconsistent. Users are advised to check data carefully and recalibrate.
- C Data are uncalibrated.
- U Data quality is unknown. If a resource does not provide a data quality assessment, class U should be assumed.

( reference: Resource Metadata for the Virtual Observatory Version 1.12)

