

# **NAIS Management Processes**

**(Document Version: V1.0)**

**Release Date: 12 February 2020**  
**National Space Science Data Center, China**

# 1 INTRODUCTION

## 1.1 Purpose and Scope

This document describes the data management processes of the Archival Information System (NAIS) that is managed by the National Space Science Data Center (NSSDC). We outline the various processes and pipelines and how these are managed by the staffers of the NSSDC.

## 1.2 Document concept & architecture

Chapter 1 provides purpose, scope, document concepts & architecture, definitions of this document. Chapter 2 provides applicable and reference documents.

Chapter 3 provides the major concepts involved in an NAIS Archive. It provides a view of the OAIS environment applied to NAIS and the roles played by those who interact with it. It also defines the requirement of NAIS archives, and NSSDC preservation and management responsibilities.

Chapter 4 describes the notion, model and detailed process in NAIS. 4.1-4.2 provides notions and model views needed for a detailed understanding of an NAIS processes. 4.3 defines the requirements for submission information, archival information and dissemination information. 4.4-4.9 provides model views needed for a detailed understanding of an NAIS. It breaks down the NAIS into a number of functional processes, and also provides detailed process views.

Chapter 5 describes the responsibilities required in NAIS.

Chapter 6 defines the procedures in use.

## 1.3 Definitions

Some components of OAIS definitions relevant to this document are listed below:

**OAIS (Open Archival Information System):** The term ‘Open’ in OAIS is used to imply that the OAIS Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the Archive is unrestricted.

**Archive:** The term covers the whole of the activities, from the initial archival management to the preservation and access to archived files. Therefore, the archive ensures sustainability of data up to preservation and access to archived files but also preservation, with data, of all necessary information to its comprehension and use (metadata).

**“Long-Term”:** A period of time long enough to be concerned about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in an OAIS. This period extends into the indefinite future.

**OAIS Information Package:** An OAIS Information Package is a conceptual container of two types of information called Content Information and Preservation Description Information (PDI).

**Content Information:** A set of information that is the original target of the preservation or that includes part or all of that information.

**Preservation Description Information:** The information which is necessary for adequate preservation of the Content Information. It includes the data and metadata (required for the understanding) and information required to rebuild the information system.

**SIP (Submission Information Package):** Data delivered by the producer of the data. These data contain Content Information and usually some level of Preservation Description Information, i.e. parts of the metadata that are useful for archival purposes.

**AIP (Archival Information Package) :** In the OAIS model, one or more SIPs are transformed into one or more Archival Information Packages for preservation. The AIP has the complete set of PDI for the associated Content Information.

**DIP (Dissemination Information Package):** Data provided to a consumer in response to a request. This may be all or part of an AIP. A DIP may include collections of AIPs, and it may or may not have complete PDI.

## **2 APPLICABLE AND REFERENCE DOCUMENTS**

### **2.1 Applicable documents**

AD1 : Operation and maintenance procedures of Chinese space science data center NAIS.

### **2.2 Reference documents**

RD1: Reference model for an Open Archival Information System (OAIS), CCSDS 650.0-M-2

RD2: Information security technology—Implementation guide for classified protection of information system, GB/T 25058-2010

RD3: Information technology equipment Safety Part1: General requirements, GB 4943.1-2011

RD4: Information technology equipment safety Part23: Large data storage equipment, GB 4943.23-2012

RD5: Information security technology—Baseline for classified protection of cybersecurity, GB/T22239-2019

RD6: Quality management systems requirements, GJB9001C

## **3. OAIS environment applied to NAIS**

Diagram 1 shows the environment model of an OAIS archive. This section provides the top level description of the model components as applied to the NAIS archive.

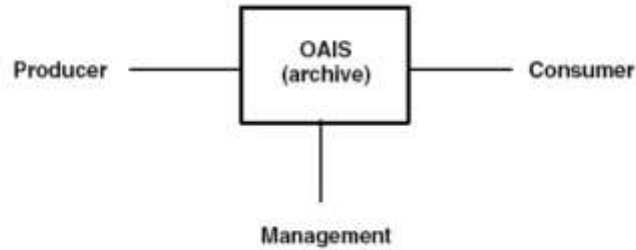


Diagram1. OAIS environment model

### 3.1 The producer

Producer is the role played by those persons, or client systems, which provide the data to be preserved. The producers can be:

- The ground-based observatories or space-based projects who have prepared science data products as the result of their observations.
- The agencies or research groups who have prepared science data products as the result of their experiments or models.
- Authors of publications with datasets to be ingested into NAIS.

### 3.2 The archive – NAIS

The NAIS consists of elaborated scientific data (science-ready) from journals or data from ground-based observatories/experiments or space missions. Diagram 2 shows the data management model of NAIS.

#### 3.2.1 Data Model

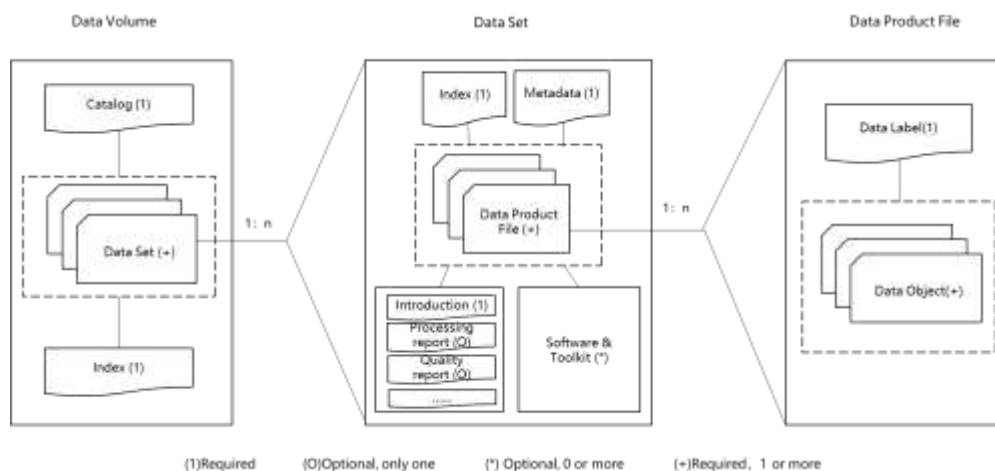


Diagram2. NAIS data management model

The data management model describes the hierarchical organization relationship of scientific data products. In order to standardize the management of data resources, the model defines and

constrains a series of entities and elements. The data management model defines three levels of entities: data product files, data sets, and data volumes.

- Data product file consist of a data label and the data objects that it describes. It is one component of data set and the smallest entity of data management;
- Data set is the main form for data release. It is an aggregation of data product files with a common origin, processing, application or related. A data set also includes index, metadata, software, tools and documentation needed to understand and use the data. A data set has a unique data set identifier (DOI or PID) for management, release and retrieval.

Data volume is composed of a series of related data sets and it is used for long term storage.

### **3.2.2 Data File**

Data product file consist of a data label and the data objects that it describes. The data label is attached at the beginning of the data product file, just before data objects.

Data objects are a set of results from a scientific observation, experiments or analysis. Data objects are main body of data product file and usually described using table, series, image, etc. The data format specifications of data objects are as follows:

- Image is usually stored in a standard image format, such as png, jpg or FITS developed by IAU.
- Series can be stored in standard format, special format or customized text format, such as CDF, RSF, Text and so on.
- Table can be stored in standard format, such as FITS, CSV, etc.

Data label is required for describing the contents and format of each individual data product file. According to the data objects format, data label has different specification of content and structure.

- If data object stored in standard format and special format, the data label shall comply with the above format definition requirements;
- Data labels of data objects stored in customized text format shall include but not limited to the following items:
  - data product file name
  - data source
  - definition, description, format, spatiotemporal range, spatiotemporal resolution, threshold, unit, version, etc. of data objects

### **3.2.3 Data Set**

A data set is an aggregation of data product files, index, metadata, auxiliary documents and software. The auxiliary documents and software are optional.

The dataset index describes the structure of the dataset directory, which is the file list of the dataset. It is stored in standard text format and consists of:

- Directory name, file name, storage pointer information.
- Information that allows a user to locate the data file of interest in data set.
- Identification and description of the function of each file

Metadata describes the contents of the dataset in detail, and stores them in XML format according to the metadata specification of SPASE (v2.0.0). The main contents of metadata include:

- Information of dataset, such as dataset name, unique identification, content description, keywords, acquisition method, time and space scope, etc.
- Information of observation station/mission, i.e. name, brief introduction, contact person, etc.
- Instrument information, such as instrument name, brief introduction, contact person, associated observation station, etc.
- Other information, such as service organization, Registry organization, repository organization, etc.

Auxiliary documents are supplementary or ancillary reference materials which are usually included with archive products to improve their short- and long-term utility. These documents provide further assistance in understanding the data and accompanying materials. These may include:

- References to science papers or the papers themselves
- Supplementary and ancillary information to assist in understanding and using the data products, such as data usage notes, data processing report, data quality report, etc.
- Documents describe the mission, spacecraft, instrument, and/or data set

Dataset contains a collection of software and toolkits which are used for data processing, analysis, visualization and other application. It may contains following items:

- the software libraries, utilities, or application programs supplied for accessing or processing the data descriptions of processing algorithms, such as user's manual.

### **3.2.4 Data Volume**

Data volume consists of scientific data set, calibration data set, auxiliary data set, catalogue and index. The calibration data set and the auxiliary data set are optional.

The data volume index describes the structure of the data volume directory, which is the datasets list of the data volume. It is stored in standard text format and consists of:

- Directory name, datasets name, storage pointer information.
- Information that allows a user to locate the data file of interest in data volume.
- Identification and description of the function of each file

The catalog identifies and describes the function of data volume. It is stored in XML format and may consist of:

- Name, contents, theme, quality and other information of data volume
- Introduction and version of the data processing
- Information of auxiliary data and calibration data related to data application
- Information of the mission, instrument, datasets, etc.

## **3.3 The management**

NSSDC is hosted by the National Space Science Center, Chinese Academy of Science, and is labelled as "National Data Center" by the Ministry of Science and Technology PRC.

NSSDC undertakes the responsibilities of long-term preservation and open sharing of space science data in China, including the following aspects

- Undertake the task of integrating and archiving space science data
- Responsible for space science data classifying, processing, analyzing and mining

- Safeguard scientific data, and promote open sharing of space science data according to laws and regulations
- Strengthen the exchange and cooperation in space science data at home and abroad

The predecessor of the National Space Science Data Center (NSSDC) was the China Space Science Data Center (CSSDC), which was established in 1988. NSSDC is also the China Space Science Center of the World Data Center (ICSU-WDC). In 2003, NSSDC became one of the regular members of World Data System. In 2019, NSSDC became one of the 203 data repository recognized by American Geophysical Union.

NSSDC is the only national data center in the field of space science in China. Since 2016, NSSDC has provided data services for nearly 100 scientific research and application projects. NSSDC also has implemented data production and management for major national space science project, including Strategic Priority Program on Space Science, Lunar and Deep Space Exploration, Meridian Project etc. In addition, NSSDC is a data service provider for dozens of units such as scientific research institutes, enterprises and institutions, and domestic colleges and universities, NSSDC has cumulatively supported the publication of more than 800 academic papers.

NSSDC data resources cover the fields of space physics, space astronomy, etc. The data resources can be searched and discovered on the NSSDC website (Chinese version [www.nssdc.ac.cn](http://www.nssdc.ac.cn), English version [www.nssdc.ac.cn/eng](http://www.nssdc.ac.cn/eng)), the sharing service Platform of National Science and Technology Infrastructure as well as other third platforms such as the Data Cloud of Chinese Academy of Sciences.

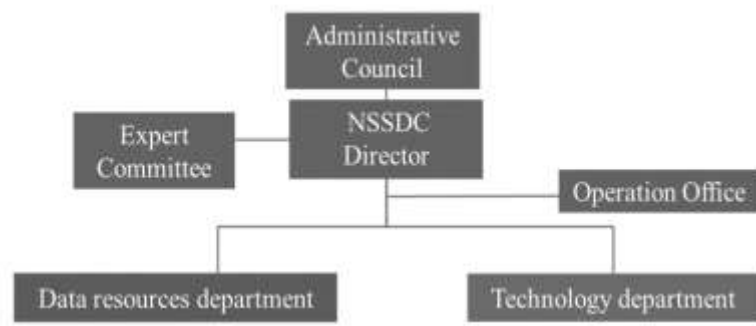


Diagram3. NSSDC Organization Structure

- **Administrative council:** is the decision-making organization for major issues of the National Space Science Data Center. It is responsible for formulating the overall planning of NSSDC, coordinating and making decisions on major technical issues. Administrative council also provides guidance and supervises the operation and service of NSSDC. The chairman of administrative council represented by the legal representative of the National Space Science Center, CAS
- **Expert Committee:** is the business guidance and consulting team of the NSSDC. It is composed of experts in the field of space physics, space astronomy, planetary science, space geoscience, computer science and data application technology, including 9 academicians of CAS and Chinese Academy of Engineer, and 7 winners of National Science Funds for Distinguished Young Scholars. The Expert Committee meeting is held once a year to consult and comment on major issues such as the work progress and work plan of the NSSDC
- **NSSDC Director:** responsible for the routine operation and management of the NSSDC,

and regularly hold regular meetings of the directors.

- **Operation Office:** Responsible for supporting the routine operation of the NSSDC
- **Data Resources Department:** is implementation department of the NSSDC's business. It is responsible for data integration, system operation, services, publicity and promotion, data security, international cooperation and personnel training.
- **Technology Department:** is the implementation department of the NSSDC's key tasks. It is responsible for the strategic planning of NSSDC, the development of the data resource system, the framework development of the data standard system, the improvement of NSSDC infrastructure the development of application tools, etc.

NSSDC is deeply involved in science data curation and sharing activities at home and abroad. NSSDC regularly participates in relevant academic conference such as International Data Week, WDS regularly meeting, International Planetary Data Alliance conference, China Space Science Assembly etc. Besides, NSSDC actively participates in various science data working groups, such as the WDS Harvestable Metadata Service working group and the science data management standard system expert working group of the Ministry of Science and Technology.

### **3.4 The consumer**

Consumer is the role played by those persons, or client systems, that interacts with the NAIS services to find and acquire preserved data or information of interest. The NAIS makes the data being presented in terms of the Knowledge Base of the space scientific community. The NAIS services incorporate a range of standard formats used by the space scientific community (For example, ASCII, CDF, FITS, PNG, SAO, MCS format, and SPASE data model and data dictionary), which allow the data to be exploited by the commonly used software and algorithms of this discipline.

## **4. Processes of NAIS**

### **4.1 Principles and notions**

The following concepts can be applied to the operations of NAIS.

- NAIS processes is inspired by the CCSDS1 Open Archival Information System (OAIS) organizational model.
- NAIS is an important resource of Chinese Science and Technology Resource Sharing Service Platform and Data Cloud of CAS.
- NSSDC is strongly engaged with the Space Science research community via relationships with the major academic Solar-terrestrial physics journals, with agencies which run observation instrument, with large projects which produce data, and also with other data centres. NSSDC also participates in the major national and international events and demonstrates and discusses the services during sessions and at demonstration booths.
- User feedback is an important driving force for management of NAIS. NAIS benefits from well organised customer service which allows a direct link between the NSSDC staffs and



the users of services. Requests and suggestions from users are collected and are taken into account as much as possible.

## 4.2 NAIS pipeline

This section describes the organization of the NAIS pipeline. A visual summary of the processes, including the use of OAIS terms is shown in the diagram 4 below.

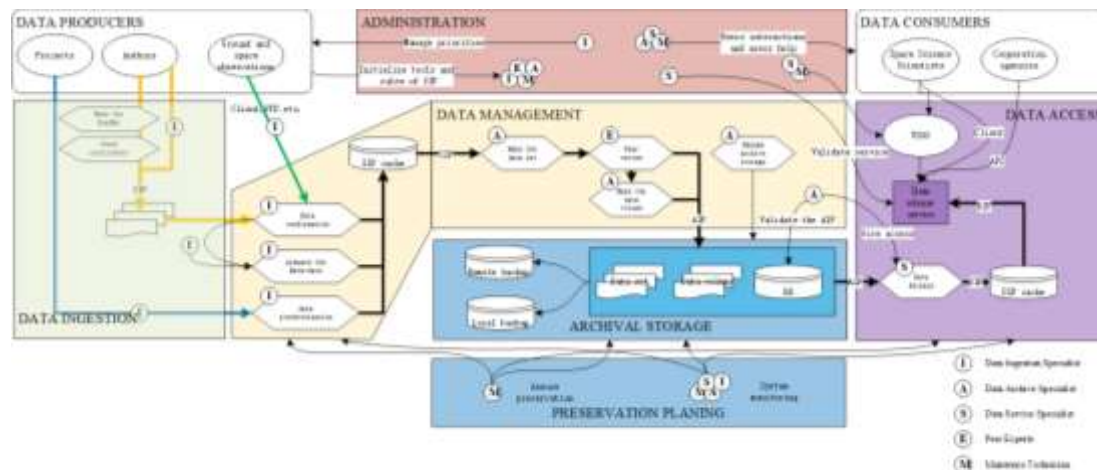


Diagram4. NAIS pipeline

The mainly inputs for Ingestion is the data that received by projects teams or ground/space observations. All these designated community submit the data in a Submission Information Package (SIP). After accepting a submission information package, quality checks will be performed, the consistency between the content and notification will be checked and the content's header will be validated. Based on the SIP, CSSDC will create the quality report, metadata and different levels of data product, all the information will be assembled as an Archive Information Package (AIP).

The main functions of Data Management are providing the service for maintaining the data set and data volume. Archive Storage provides the services and functions for the storage, maintenance and retrieval of AIPs. Its functions include receiving AIPs from Data Management and adding them to permanent storage, managing the storage hierarchy, refreshing the tape library on which AIPs are stored, performing routine and special consistency checking among different storage areas, providing disaster recovery capabilities.

Access provides the services to the space science data consumers. Its main functions include generating Dissemination Information Packages (DIPs) based on the AIPs sent by Ingestion entity, managing the DIPs and generate the query response. In case there is some queries from consumers asking for the previous AIPs, Access entity could send a request to Data Management entity. The DIP defined in CSSDC includes content (compressed or part of AIPs) and dissemination list (file index), the data volume and time period vary depending on consumers' orders.

Administration entity provides services for the overall operations of the OAIS. Its functions include performing access control, operations monitoring, managing the configurations and defining strategies and standards for ingestion.

## **4.3 Information package**

The notion of OAIS information packages distinguishes the data format used in the exchanges between the NAIS and the external participants (producers, users). There are three kinds of information package in NAIS.

### **4.3.1 Submission Information Package (SIP)**

The Submission Information Package (SIP) is the data set which is provided by the producer. In the NAIS, the SIP consists of one or several data product files, metadata related to data processing, software, toolkits, processing reports, data quality reports and other supporting documents. The format of these contents is described on the website: <http://www.nssdc.ac.cn/eng>.

The metadata is initially provided by the producer (see section Data Set) or by NSSDC staffers in collaboration with the producer. This information is checked and completed by authorized NSSDC staffers (AIP data).

The producer can edit the metadata by using the metadata generation tool and submit SIP to the data center for approval through the online registration system.

### **4.3.2 Archival Information Package (AIP)**

One or more SIPs are transformed into one or more Archival Information Package (AIP) for data release and preservation.

In the NAIS, the AIP is the data set and data volume which are prepared by competent and authorized NSSDC staffers. The constitution of data set and data volume is listed in Section 3.2.3 and Section 3.2.4.

The ingestion process and analysis builds new meta-data, index and supporting files used by the metadata generation tool and archive production software in the NAIS. All contents will be organized into a standard directory structure.

At the end of the ingestion process, the final AIP may contain the following items:

- complete set of data product files, includes auxiliary data files and calibration data files
- complete metadata
- software and tools related to data processing and application processing reports, data quality reports and other supporting documents

### **4.3.3 Dissemination Information Package (DIP)**

The Dissemination Information Package (DIP) is the data set that NSSDC ultimately distributes to the consumer. DIP may be all or part of AIP, even a set of several AIPs. The DIP may include the following items:

- data product files, includes auxiliary data files and calibration data files
- metadata

- software and tools related to data processing and application
- processing reports, data quality reports and other supporting documents

The NAIS data are made openly available to the community via the NAIS service with interactive interfaces (<http://www.nssdc.ac.cn/eng>). The interface allows for searching, filtering, obtaining of the DIP.

## 4.4 Data ingestion

Data producer organizes data file and metadata into SIPs (Submission Information Package), then submits the SIPs to NSSDC. Data producer is responsible for the quality of data content and metadata content, who also in cooperation with NSSDC, figures out the data quality problems happens in data SIPs submission and ingestion processes. Then, NSSDC receives the SIPs, audits and examines the SIPs data quality. Data quality examination results would be fed back to data producer; those SIPs not pass the quality examination would be revised by data producer and submitted again till the NSSDC approves the SIPs data quality. SIPs approved by NSSDC would be ingested into NSSDC, managed in dataset form: data entities be stored in database and metadata be stored in information database. Datasets in NSSDC would be organized into data volume form further, being stored in archive base.

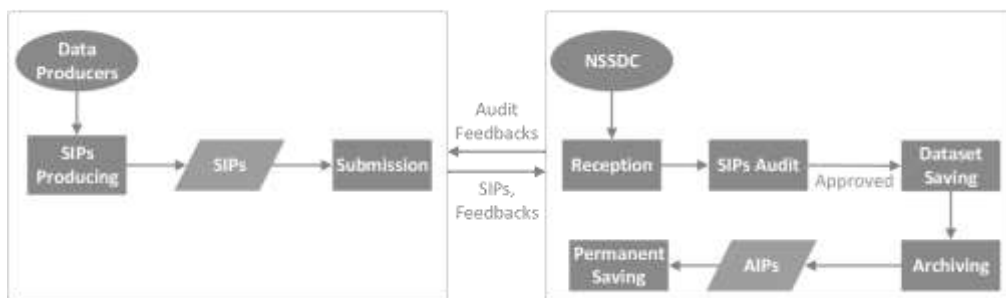


Diagram5.Data Ingestion Pipeline

Responsibilities of members in SIPs submission and ingestion:

- Data producer (person, research team or institutes) is responsible for data products and metadata generation, SIPs organization in accordance with NSSDC SIPs requirements and SIPs submission into NSSDC. Data producer is also in charge of SIPs data quality, ensuring the readability, specification, integrity, consistency, and availability of the data and metadata, figuring out the quality problems fed back by NSSDC in data ingestion process.
- Data Ingestion Specialist is responsible for SIPs reception, SIPs data quality audition, examination (validate the automatically examination results and examine the SIPs manually further) and examination results feeding back. Data Ingestion Specialist would organizes peer experts to validate the precision and availability, organizes the training and popularization of data model, metadata standards and technical system.
- Data Archive Specialist is responsible for dataset and data volume generation, designing the catalogs and attributes of dataset and data volume by considering different sub-discipline characteristics and storing dataset and data volume into database.
- Maintenance technicians is responsible for the operation, maintenance and upgrading of

technical systems in NSSDC, such as data ingestion web service, metadata editor tools, data archiving system and et al.

- Peer expert is responsible for data evaluation such as data generality, usability and accuracy.

## 4.5 Data management

NAIS manages and retrieves AIP and metadata through product database and archive database:

- After AIP preparation, the information of data product files and data labels is stored in the product database. The organization information and metadata information of data set and data volume are stored in archive product database. All the information is managed and stored uniformly.
- Data archive specialist copies, queries, analyses, transforms and checks data product files, data sets and data volumes through product and archive management software;
- As the content or version of data and metadata is updated, the data archive specialist will update the data product file information, metadata, data description documents, etc. in the product database and archive database.

NAIS performs real-time or scheduled backup of database tables, indexes, views and other information. Real time backup is realized by the database backup tools (such as RMAN) and the scheduled backup is realized by the backup script (expdb, mysqldump) written by the maintenance technician. The backup frequency adopts real-time incremental backup, daily full backup and weekly full backup according to the increment and importance of data.

Responsibilities of personnel in data management activities:

- Data archive specialists are responsible for metadata and AIP management, such as data update, long-term storage, disaster recovery, statistical analysis of archived data, regular inspection of data and metadata to ensure data security, usability and long-term accessibility;
- Maintenance technicians maintain information systems and application softwares, such as data product library, archive database, tape library, etc. to ensure the safe and stable operation of the system.

## 4.6 Archival storage

NAIS is composed of long-term storage system, data backup system and data disaster system. Based on the OAIS storage model, the diagram of NAIS process is as follows:

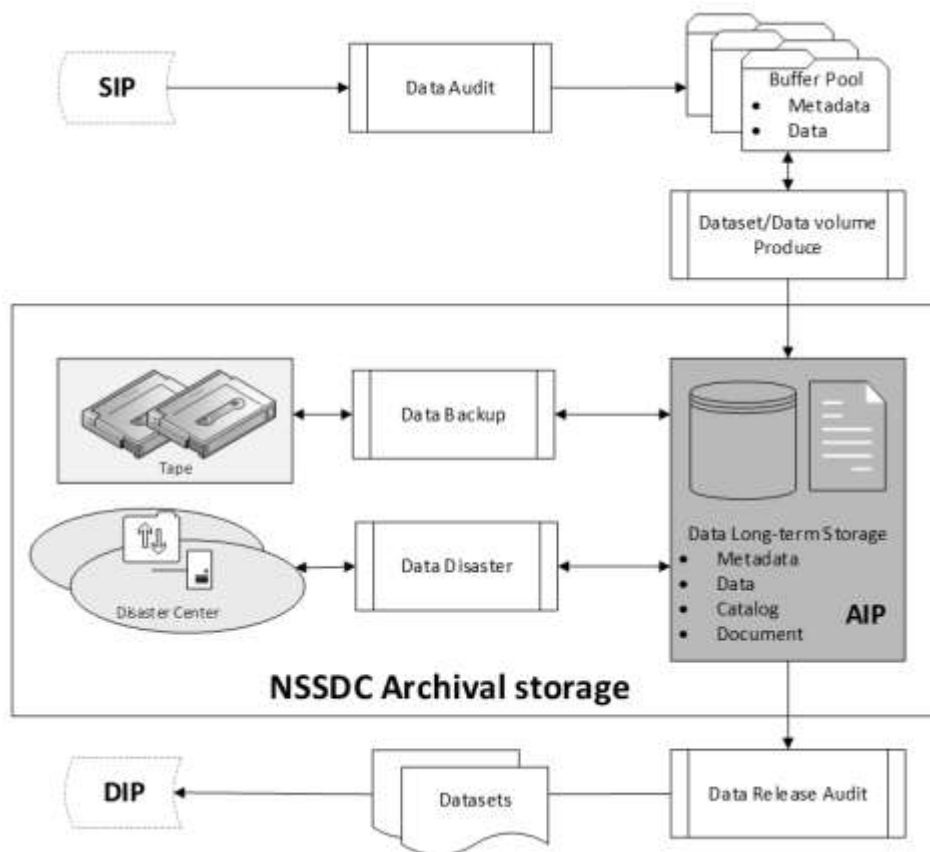


Diagram6. NAIS Archival Storage Process

Each stage involve several participants who have the following responsibilities:

- Data Ingestion Specialist: Responsible for the initial review of the integrity of the SIP submitted by the producer
- Data Archive Specialist: Responsible for restructuring directory and producing AIP after supplementary information according to discipline or specification
- Data Service Specialist: Responsible for composing AIP to SIP which is demanded by consumer;
- Maintenance Technician: Responsible for monitoring and upgrading the hardware and software of NAIS
- Peer Experts: Professional advice on NAIS’s archive, storage, etc.

#### 4.6.1 Digital storage

NAIS uses online-near line-offline multilevel storage. Online storage using SAN storage and distributed storage, near line and offline storage using tape library.

SAN storage consists of HP3PAR SAS array and switches, with a total capacity of about 300TB, which is mainly used to support business cache and database cluster. Distributed storage is composed of HUAWEI, DELL storage server and distributed file system. Distributed storage is divided into three categories: distribution cache, release cache and long-term archive storage, with capacities of 500TB, 500TB and 1PB respectively. The tape library uses the Quatumn i6000 with a maximum capacity of 1500 tapes and eight LTO6 drives to store inactive data in archival data.

All the hardware and software in NAIS like server, tape and file system are equipped with corresponding management system, which can achieve real-time monitoring and maintenance.

## **4.6.2 Data duplication**

NAIS adopts different methods to copy and store data. Level 3 data (highest important) is defined as the initial digital sample data obtained based on sampling technology, which use local backup and remote disaster recovery; the data processed on the basis of Level 3 data can be divided into level 2 (important) and level 1 (general) according to processing complexity and application scope. Level 2 data and level 1 data use local backup and RAID for storage respectively. Besides, level 3 data requires additional remote disaster due to non-renewable data.

NAIS uses tape library as the local backup system. At present, the tape library can support 1500 tapes with a total capacity of 3PB. NAIS selects Kunming and Dongguan as remote disaster recovery centers which is in the south of China (distance from the data center are more than 700km). Each disaster recovery center is equipped with good facilities, with an effective storage capacity of more than 200tb.

NAIS adopts data driven way to execute data backup and disaster after generating data entities / data sets / data volumes. NAIS uses tape's Application Programming Interface (API) to trigger backup process and transfer data to disaster recovery centers through virtual private network (VPN). In addition, the data center will regularly check MD5 values of data stored in local and remote to ensure data consistency.

## **4.6.3 Recovery Service**

When data service specialist or data archive specialist find occurrence of data loss, damage or not accessible. They will inform maintenance technician to check the system status. After technician confirms the abnormal data, they will activate recovery task and NAIS will execute operation of moving tape data or recovering the disaster data automatically.

## **4.7 Administration**

### **4.7.1 NAIS operation and maintenance**

NAIS consist of hardware and software components. The hardware components are the infrastructures include computer, network, firewall, etc. The software components functions include data acquisition and ingest, data processing pipeline, data quality control, data management, data archive storage and data access service.

NAIS released a document to define operation and maintenance procedures. The following contents was defined in the document.

- Operation and maintenance modes, lifecycle and also the methods for measuring service quality.
- Service architecture and management model of operation and maintenance.

- The objects and daily activities about operation and maintenance.
- Events management procedure.
- Operation organization architecture and its responsibilities.

For details, please refer to AD[1].

## 4.7.2 Roles and responsibilities

Different roles were set up for the NAIS operation and maintenance, include data ingestion specialist, data archive specialist, data service specialist, peer experts, maintenance technician, etc.

Each role needs to carry out operation and maintenance activities according to the corresponding workflow and operation specification, the following activities needs to be done by the operating personnel.

- The events found during operation shall be notified to the relevant, and dealt according to the operating procedure and failure plan.
- Monitoring and routine maintenance should be done according to operation plan.
- Record the events of operation and maintenance activities.
- Collect and archive operation and maintenance documents.
- Carry out analysis and evaluation regularly, and put forward suggestions for optimization and improvement.

The detailed information could be seen in AD[1].

## 4.7.3 Management activity

The NAIS processes is subject to continuous maintenance and evolution. Operations, maintenance and evolutions are discussed in dedicated meetings:

- Weekly meetings of the NAIS team (data archive specialist, data ingestion specialist, maintenance technician, and researchers) are dedicated to discuss the questions of data set management.
- Regular internal NSSDC follow-up meetings gathering the whole NSSDC staffers.

Evolutions are also discussed with partners: The NSSDC staffers maintains relations with data producers (spaces agencies, observatories, journal editors)

The report of the scientific council which meets annually gives high level guidelines and recommendations to the NSSDC.

The NSSDC has an email contact point ([nssdc\\_service@nssc.ac.cn](mailto:nssdc_service@nssc.ac.cn)) with users of the NSSDC services. It is the contact to help users with the NSSDC services. It is also the contact point to provide feedbacks or report on content or access malfunction or propose evolutions. The NSSDC takes in charge the reports and answers to the caller in brief delay.

The catalogue service provides also an annotation service to give the possibility for users to annotate catalogues. The annotations are regulated by an astronomer of the NSSDC.

A dedicated contact is available for the submission process. The NSSDC contact point is the data ingestion specialist which in charge of the data ingestion.

## 4.8 Data access

The NAIS system designs a publishing database and a data publishing service system in the data access pipeline to ensure that data consumers can understand, request and obtain data.

Responsibilities of members in data access pipeline:

- Data Service Specialist: Responsible for tracking user requirements, answering user questions, processing user orders and providing data products to users;
- Maintenance technicians maintain information systems and application softwares, such as publishing database, VSSO and NSSDC website, etc. to ensure the safe and stable operation of the system.

### 4.8.1 Data release

To ensure data security, NAIS provide data access functions and services based on the Release Database, a mirror of the Archival Database, rather than directly on the Archival Database. The Data Service Specialists are responsible for the administration of the Release Database.

NSSDC release data into two categories, open-online data and open-offline data, depending on the data properties.

For open-online data, all of the content data objects and description information, namely the scientific data and ancillary data files, the metadata file and introduction document, are available on the Internet. Consumers can download the DIPs by self.

For open-offline data, only the metadata file and introduction document are available on the Internet. Consumers should submit data application forms to apply for scientific data and ancillary data files. NSSDC Service Specialists will deliver the data via FTP service, e-mail or hard copy within 5 working days, if the application is approved.

### 4.8.2 Data retrieval and access

The following data retrieval and access services are available at any time to NAIS consumers:

- Online data access service: NAIS has built the Virtual Space Science Observatory (with its Chinese site <http://vssso.nssdc.ac.cn> and the English site <http://www.nssdc.ac.cn/eng/>) to provides 7\*24 online data access services. Users can retrieve, browse or download data through the website, as well as view the metadata, unique identifiers, data licenses and other attached information online. The upgrade information of release data will be state on the website, if any. VSSO has been registered on re3data.org to promote the findable and accessible of NSSDC archival data.
- Database access API service: Other client systems can access to the Release Database via this API service, to query or acquire data in the authorized DB tables.
- Co-operation access service: NSSDC metadata are also available via data catalogue on the Science and Technology Resource Sharing Service Platform (<https://test.escience.org.cn/>) and that on Data Cloud of CAS (<http://www.csdb.cn/>). Consumers may find and access NSSDC data via these two websites.



### **4.8.3 Recovery Service**

Both database and VSSO have configured backup policies. Databases operate in dual-master mode to ensure high availability. Use database tool or scripts to implement daily backup. VSSO adopt active-standby deployment by using KeepAlive services. When one host goes down, the service can be automatically switched to another host to ensure the normal external service of NSSDC's portal and retrieval system.

## **4.9 Preservation planning**

The NSSDC plans for very long-term preservation of the data archives. This requires continuous actions on both data management and information system management which need to be kept fully up to date to avoid their obsolescence.

### **4.9.1 Data sustainability**

NSSDC implements data management based on the data model. In terms of data files and metadata, NSSDC refers to field or industry standards, tracks changes in standards. If necessary, the version of data files and metadata will be upgraded to ensure the sustainable use of data.

- Data Model: it defines the organization requirements of data file-data sets-data volumes to ensure that the data packets in each stage of NAIS processes can be manage and organize in a unified manner, see 3.2.1.
- Data File: are usually stored in CDF, TXT, GIF, PNG, ASCII and other formats, which are common formats in the field or industry. The advantage of these formats is that they can be widely recognized by the industry. When the format standards are upgraded, there are open communities to discuss, and even some open source code or tools can be used. When necessary, NSSDC will formulate a data format upgrade plan and perform batch processing and upgrade processing on data files. The NSSDC data version change event will invite the data producer to participate. The data in the new format after the change will be used as a new data set version, and the changes will be described in the metadata and documents.
- Metadata: NSSDC has designed metadata with reference to the SPASE 2.0.0 standard. Metadata files are stored in XML format, which can be used to mark data and define data types. It is also a source language that allows users to define their own structures and marks. The XML format has nothing to do with the operating system and the development platform of the programming language; it can realize the data interaction between different systems. When the metadata standard changes, because the metadata file is stored in XML format, it is very easy to upgrade and convert. The XML format has nothing to do with the operating system and the development platform of the programming language, it can realize the data interaction between different systems. When the metadata standard changes, it is very easy to upgrade and convert because the metadata file is stored in XML format.
- Auxiliary data or files: stored in common formats, such as SQL, PDF, XLSX and other formats. These formats are supported by professional manufacturers, and easy to upgrade and

maintain.

## 4.9.2 Continuity of access

The development of IT information technology and changes in data user needs make NAIS need to track and maintain sustainability. NSSDC will collect system transformation requirements and upgrade the system through online user feedback mechanisms, consultation of subject experts and IT technical experts. System upgrades usually take place every 3-5 years. The upgrade of NAIS will be conducted by professional software companies in accordance with the requirements of software engineering management. The upgrade of NAIS will be conducted by professional software companies in accordance with the requirements of software engineering management.

Before the new version of NAIS is launched, system testing will be conducted to ensure the safety and reliability of the system. When the new version of the system is deployed, it will be implemented according to the migration plan to ensure the integrity and consistency of data and related information.

### 1. Fund guarantee and support from affiliate institutions

NSSDC is a national-level science data center recognized by the Ministry of Science and Technology (MOST) of the People's Republic of China, and a disciplinary data center recognized by the Chinese Academy of Sciences(CAS). NSSDC operates on the basis of the National Space Science Center (NSSC), CAS. NSSDC can obtain fixed operating funds from the Ministry of Science and Technology, CAS and NSSC every year. Most of NSSDC's staffers are employees of permanent government positions at the National Space Science Center, which can effectively ensure the stability of the NSSDC's staffers. In addition, NSSC also provides an 800-square-meter computer room and workplace for NSSDC. The above support conditions can ensure that NSSDC continues to provide data access services.

### 2. Tracking and evaluation of the development trend of data management standards and data specifications.

NSSDC staffers have joined IPDA, IVOA and other standard working groups, for example, China National Standardization Management Committee subject standard committee (such as space science, etc.). NSSDC tracks the development trends and new requirements of relevant data policies, standards, and management demands through working groups meetings and standard committee meetings. These tracking results provide important references for NAIS related standards (such as AIP, SIP requirements) and data policy. The NSSDC tracking list is as follows:

- Simulation Extensions for the Space Physics Archive Search and Extract Data Model
  - Planetary Data System Standards Reference
  - Planetary Data System Data Dictionary
  - Common Data Format
  - Flexible Image Transport System Data Format
  - GB/T32843-2016 Science and technology resource identification
  - GB/T 36344-2018 Information technology—Evaluation indicators for data quality
  - GB/T 37846-2019 Definition of archived data format for space-based high-energy astronomy
  - GB/T 30114.1-2013 Terminology for space science and application—Part 1:General
- .....

### 3. Tracking and evaluating the development trend of the technology system.

NSSDC information system construction refers to or follows the following national standards and international standards. NSSDC conducts long-term follow-up on the development of the following standards and evaluates technical system upgrade needs based on the follow-up results.

- GJB 5000A-2008 Capability maturity model for military software development
- GB/T 25058-2010 Information security technology—Implementation guide for classified protection of information system
- GB 4943.1-2011 Information technology equipment Safety Part1: General requirements
- GB 4943.23-2012 Information technology equipment safety Part23: Large data storage equipment
- GB/T22239-2019 Information security technology—Baseline for classified protection of cybersecurity
- GJB9001C Quality management systems requirements
- ISO 14721 Reference Model for an Open Archive Information System
- ISO 20652-2006 Producer-archive interface — Methodology abstract standard

#### 4. Tracking of user service requirements and product technologies

In order to improve services, NSSDC defines two ways to track the service requirements and product technologies of users (producer and consumer).

- Tracking requirements during Data Archive Plan preparation. Through the preparation process of the *Data Archive Plan*, the NSSDC 's archive specialist negotiates with the producer on the format, version, dedicated software tools for archiving data, archive arrangement, and archive interface, etc. The archive specialist also confirms these contents during each archiving process. This process is one of the effective ways for NSSDC to track data producer's service and product technology needs.
- User feedback and survey activities. NSSDC tracks the needs of users (producer and consumer) through academic conferences, website feedback pages, and user survey activities.

## 5. Responsibilities for managing NAIS

This section describes the related various aspects of the NAIS management processes to the list of the mandatory responsibilities for an OAIS archive.

### 1. Communicate with the Producers to obtain more accurate information

The NSSDC keep in touch with agencies, projects and journals from which the data set come from through the high-level dialog. Via this methods the NSSDC signs the data archive plan with the producers to agree upon the data deposit rules, and a clear understanding or definition of the NSSDC responsibilities for archiving the data set.

### 2. Perform the sufficient control of the data set reception to ensure Long Term Preservation.

The NAIS data reception process is operated by the NSSDC data archivist and validated by space science researcher. Meta-data can be built entirely by the NSSDC or assigned in advance by data producer and then completed by NSSDC archivist.

The NSSDC is accountable for the data content it archives and releases: which gives it the rights to modify the storage format based on the technological novelty or obsolescence. Data and meta-data are stored into long term preservation formats that enables to recreate the database.

The NSSDC provides a metadata editor for producers who submit data into NAIS. (The data ingestion guideline see [http://www.nssdc.ac.cn/eng/ingestion\\_guideline.html](http://www.nssdc.ac.cn/eng/ingestion_guideline.html)) The documentation web page explains the metadata standard and sample, as well as the sample of data introduction document expected. (Details see

[http://www.nssdc.ac.cn/eng/standard\\_specification\\_list.html](http://www.nssdc.ac.cn/eng/standard_specification_list.html))

See data and meta-data contents description in Sections 4.3.1 Submission Information Package and 4.3.2 Archived Information Package (AIP)

3. Ensure that the information to be preserved is independently understandable to the related research community. The research community should be able to understand the data without needing special help such as the assistance of the archivist who produced the data.

The NSSDC archives are primarily for professional space science researcher. Space science research institutions, groups or individuals are the target community of NSSDC. NSSDC needs professional feedback from the community to improve its technique and service levels. (The data service flow see [http://www.nssdc.ac.cn/eng/service\\_flow.html](http://www.nssdc.ac.cn/eng/service_flow.html))

Meta-data enhance the readability of data set and allow an indexation in Virtual Space Science Observatory (VSSO) or other data service website in line with the needs of research.

The VSSO provides meta-data formats according to the SPASE 2.0.0 (details see [http://www.nssdc.ac.cn/eng/Spase-2\\_0\\_0.pdf](http://www.nssdc.ac.cn/eng/Spase-2_0_0.pdf)). Space science scientist and software engineers participate in the SPASE for the standard definition, evolution and implementations, hence ensuring the relevance of these standards.

See also Section 4.3.3 Dissemination Information Package (DIP)

4. Follow documented policies and procedures which ensure that the data and information is preserved against all reasonable unforeseen reasons.

See Sections 4.6 Archives storage and 4.9 Preservation planning

5. Make the preserved data and information available to the space science community and enable the information to be traced to the original submitted Data Objects with evidence supporting its Authenticity.

Data and meta-data are available through the NSSDC services.