An application of CCSDS archival standards to meet both submitter and archive needs during data ingest.

H. Kent Hills
Perot Systems
(NASA NSSDC)

Donald M. Sawyer
NASA GSFC

Patrick McCaslin
Perot Systems
(NASA NSSDC)

The National Space Science Data Center (NSSDC) is currently in the process of implementing new highly automated procedures for ingest of data into the archive, following the general guidelines of the CCSDS document "Producer-Archive Interface Methodology Abstract Standard (PAIMAS)', for the producer-archive interactions.

The main objective of the current development is to meet the concerns of our staff scientists about the difficulty of using the previous ingest process. We are actually in transition now, moving from our previous process to a new one, and various changes to our main database system will result. We need to make the acquisition of ingest information flow as smoothly as possible, and we need to effectively use as much automation as possible, to be cost-efficient in our operations, as well as to meet current and upcoming archival standards.

This presentation will briefly describe the evolution of the NSSDC archival ingest operations and will provide an overview of the development of the current process of gathering the necessary information relative to the automated ingest of data into the archive. The process is defined as a multi-phase one, as discussed in PAIMAS; in this presentation, we will consider mainly the Preliminary Definition Phase and the Formal Definition Phase.

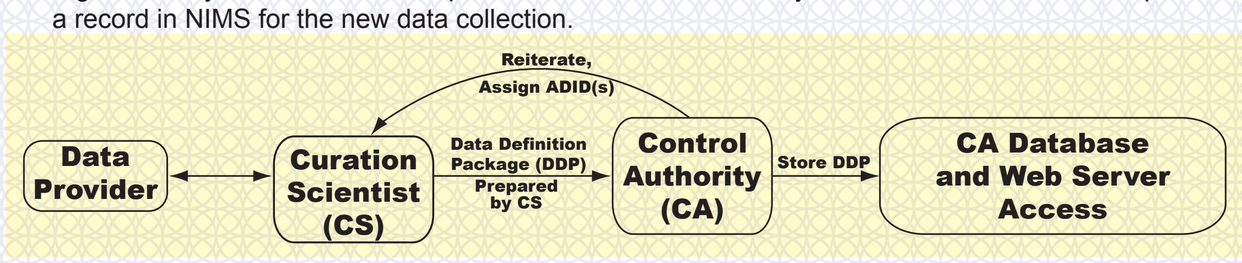
Previous Ingest Procedure

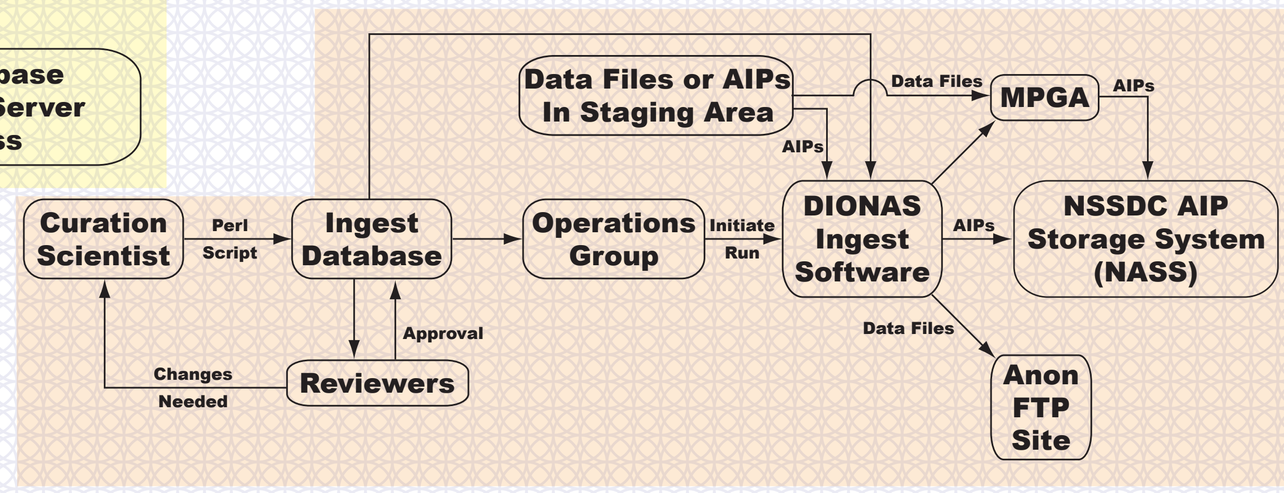
The previous procedure for ingesting and tracking of data into the archive evolved over the several decades of operation of the NSSDC. There came to be multiple separate steps to be completed, resulting in redundant entry of various information items, such as spacecraft or observatory name, experiment name, and personnel (and their contact information).

Initially, and up until about 2000, the sequence of actions depicted above was all that was required of the curation scientist involved in receipt and archival of data. Documentation was filed in hardcopy folders, and the Operations Group made appropriate copies of the data and made manual entries into an inventory file for tracking the data.

Current Procedure

The need to move to an archive form independent of both media and file systems, plus the need to get data off of obsolete optical disks, led to the implementation of AIPs for the archive. The MPGA software was developed to package data and attributes into AIPs, and a pipeline was created to semi-automate the ingest and archival storage processes. The AIPs are media-independent, however exact copies of the input data files or directory structure can be recovered. IMAGE was the first spacecraft project to use the MPGA software remotely to send AIPs to the archive. This resulted in the following additional two sequences of actions to be followed to archive a data collection, and to track and store documentation, using the Control Authority.document registration system. The initial sequence above is still necessary to insert information about spacecraft and experiments into the NSSDC Information Management System (NIMS), and also to open a record in NIMS for the new data collection.

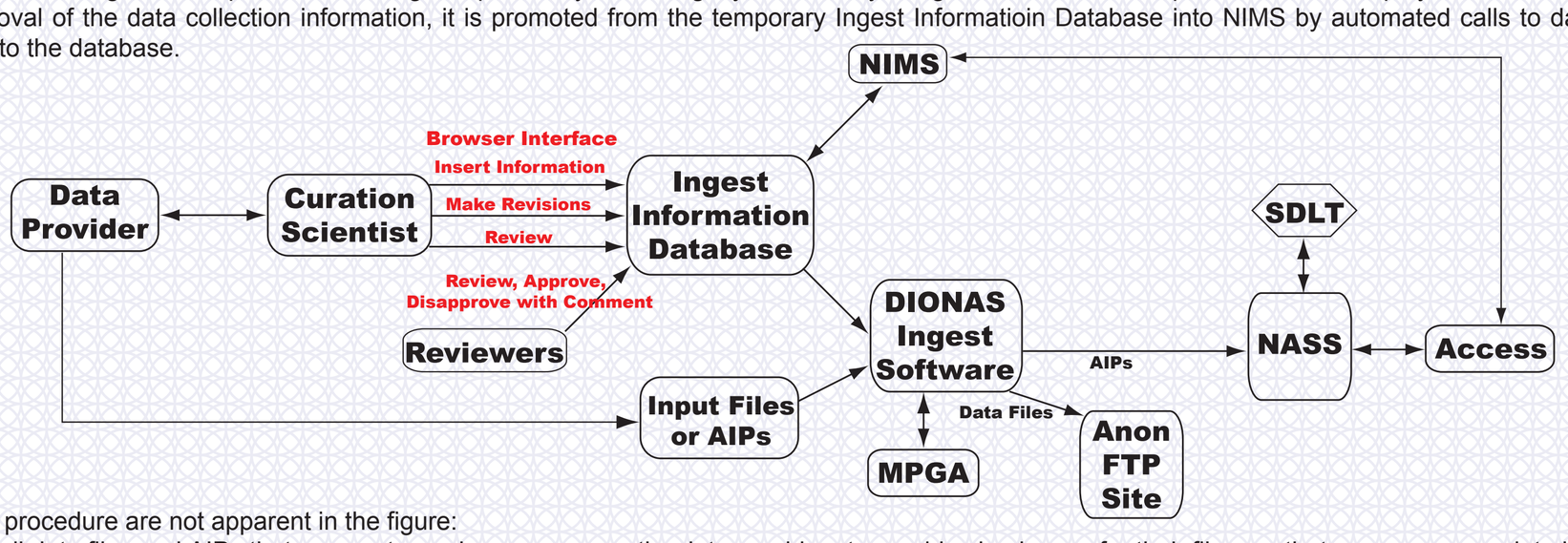




New Streamlined and Automated Procedure

The need is to make the overall process flow more smoothly, be more convenient for the ingest personnel, more cost-effective, and more robust. Therefore we are streamlining and combining these three processes, and automating them as much as possible, while still keeping the various functions separated in a modular logical structure. For now, we are addressing only the ingest of a data collection, and not entry of information about a spacecraft or experiment. Usually the Spacecraft (or Observatory) and Experiment names have already been entered into NIMS long before we are ready to enter data collections into the archive. Eventually we will include that process as well.

Adopting terminology from the CCSDS/ISO PAIMAS standard, there are certain attributes of a candidate data collection that are required early, in the Preliminary Phase of interaction between producer and archive, and certain other attributes aren't required until later, in the Formal Definition Phase. The newly developed ingest process, shown below, uses a browser interface to acquire all of the information required for the Preliminary Phase of archival, and soon will include the Formal Phase information as well. A Curation Scientist (or eventually perhaps the data provider and the Curation Scientist)) can enter the information items when available, and can go back at any time to modify items or to insert missing items. If the spacecraft or experiment IDs are known, they can be used instead of the name, allowing easier input and reducing the possibility for ambiguity. When fully integrated with NIMS, a pick list will be displayed to facilitate accurate spacecraft and experiment entry. After approval of the data collection information, it is promoted from the temporary Ingest Informatioin Database into NIMS by automated calls to database stored procedures, eliminating manual entries into the database.



Two important features of the procedure are not apparent in the figure:

a. Checksums are used on all data files and AIPs that we create, and we encourage the data providers to provide checksums for their files, so that we can ensure data integrity from end to end;
b. The AIPs contain various attributes, and the Access Function (aided by NIMS) can request the appropriate AIPs from NASS.

The attributes required in the Preliminary Phase and in the Formal Phase of ingest are shown below, together with the current implementation of the user interface, which now only addresses the Preliminary Phase attributes. The interface provides help via pull-down menus and hyperlinks to text guidelines for various fields.

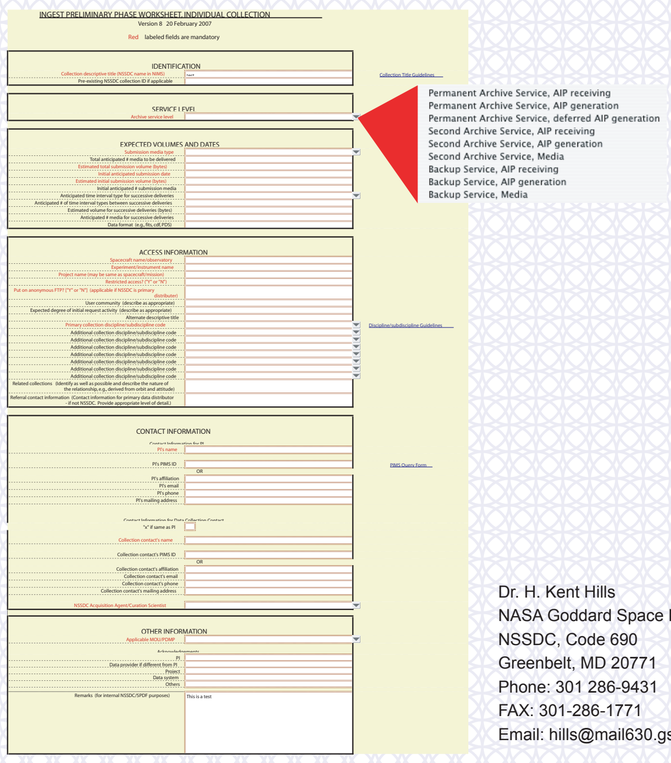
Attributes for Preliminary Phase

- Identification
  - Collection descriptive title
  - Collection ID Prefix
- Service Level
  - Archive service level
  - Service sub-level
- Expected Volumes and Dates
  - Submission media type
  - Total submission volume
  - Initial submission volume and date
- Access Information
  - Spacecraft name/Observatory
  - Experiment name/Instrument
  - Restricted Access (Yes/No)
- Contact Information
  - PI
  - Data collection submitter
  - Data collection contact
  - NSSDC Curation Scientist
- Other Information
  - Applicable MOU/PDMP
  - Preliminary agreement approved

Plus a number of optional attributes

Attributes for Formal Phase: Permanent Archive

- Location of Data & Attributes
  - wild card for directory use
  - input data filename structure
  - input data directory structure
  - input data other attributes and location
- Populate NSSDC Attribute Obj.
  - output data filename as it should appear
  - file type (binary/ascii)
  - NOST format ID(s) (NSSDmmmm)
  - Encoding used
  - NOST IDs of encoded format (NSSDnnnn)
  - Identify algorithm mapping file names,directory, etc. to ADIDs
- FTP Placement Information
  - Project name (editable)
  - Experiment name (editable)
  - Data set short name (editable)
  - Root for pathname
  - Algorithm for FTP insertion
- Processing Instructions
  - Ingest Priority Agreed
  - Special processing instructions
- DDP Registration
  - ADIDNAME: format is NSSDnnnn
  - TITLE: (free text)
  - DESCRIPTION: (free text)
  - DDP Personnel (Originator and Revisors)
    - Last name, first name of person
    - Position of person
    - Affiliation
    - email address
    - telephone number (if available)
    - FAX number (if available)
    - postal address
    - Project
    - Mission (spacecraft)



Dr. H. Kent Hills
NASA Goddard Space Flight Center
NSSDC, Code 690
Greenbelt, MD 20771
Phone: 301 286-9431
FAX: 301-286-1771
Email: hills@mail630.gsfc.nasa.gov