

Release A CDR RID Report

Date Last Modified 10/9/95
Originator Chris Lynnes
Organization GSFC DAAC
E Mail Address lynnes@daac.gsfc.nasa.gov
Document Data Ingest Design

Phone No (301) 286-2260

RID ID	CDR 17
Review	SDPS/CSMS
Originator Ref	GD-CDR-CL-3
Priority	1

Section

Page

Figure Table

Category Name Ingest Design **Actionee** ECS

Sub Category Ingest

Subject Ingest over time of multi-part entities

Description of Problem or Suggestion:

Both granules and documents may be ingested in pieces, over time. The most likely example is the revision of one file making up part of a guide document (with 100% probability of occurrence). Other examples are documents delivered in pieces, or granules delivered in pieces.

Originator's Recommendation

Develop a mechanism for ingesting entities over time, or explain how current system provides this capability.

GSFC Response by:

GSFC Response Date

HAIS Response by: Jacob Eisenstein

HAIS Schedule 9/6/95

HAIS R. E. Carey Gire

HAIS Response Date 9/29/95

The "ingest over time" scenario for multi-part entities is handled by the current system design for the Ingest and Data Server software presented at the Release A CDR. The metadata supplied by the data provider and the underlying Data Server metadata data base model define the relationship between the multiple entities associated with a document or data granule. Note: each "part" in a multi-part entity is separately identified by inventory metadata. ECS is not supporting "accumulated" entities, where a complete entity is not accessible until all of its multiple parts are available. Instead, each of the multiple parts is individually accessible, in addition to the full entity.

The "ingest over time" scenario describes both the ingestion of portions of a document as well as the ingestion of data granules that are related. Document files ingested at different times (e.g., Guide document sections) and data granules ingested at different times (e.g., QA information related to a previously-ingested science data product) are handled similarly by ECS. The ingest process is handled in three general steps:

- 1) the data provider (e.g., an authorized SCF user) supplies metadata with the document section/data granule that a) describes the document/granule and b) relates the document section/data granule to the parent document/data granule which has previously been ingested. For GUI-based Ingest, an HTML form is provided to prompt the user for the relationship metadata described in b). Typically this will be the previously ingested Granule ID.
- 2) The Ingest software passes the document section/data granule and the accompanying metadata (in PVL format) to the Data Server software via an insert request.
- 3) After storing the document section/data granule information, the Data Server software loads the metadata into the Data Server metadata data base. The metadata data base is structured so as to use the relationship defined in 1b) above to relate the inserted document section/data granule to a previously-inserted document/data granule.

As an example, consider the ingest of XYZ Guide document sections. We assume that Guide documents are input to ECS by means of the ECS User Ingest HTML (GUI) forms mechanism. That mechanism is described in section 4.5.7 of the DID 305 document (305-CD-009-001).

An authorized SCF science user calls up an HTML form for Guide document ingest from the user's workbench. The form requests input of the the sub-data type ("XYZ Guide Document"), and the document section ("Section 1"). Note: the valid set of data sub-types and section identifiers will be defined and published. The user will be required to enter document-level metadata (document section, subtype, and date at a minimum) for the initial ingest of the document, or as an update to an existing document.

Date Printed: 11/14/95

Page: 1

Official RID Report

The HTML form contents are transferred to ECS, the contents are parsed by Ingest software, and the data type and document section are extracted. The Ingest software stores the sub-data type and document section in a companion PVL file (along with any other supplied metadata). The Ingest software invokes the Data Server insert service, supplying the data type and the names

Release A CDR RID Report

document.

The HTML form contents are transferred to ECS, the contents are parsed by Ingest software, and the data type and document section are extracted. The Ingest software stores the sub-data type and document section in a companion PVL file (along with any other supplied metadata). The Ingest software invokes the Data Server insert service, supplying the data type and the names of the document section and PVL files. The Data Server software stores the document section file and loads the PVL metadata into the metadata data base. Metadata is stored for the data type, the sub-data type, and the document section. The sub-data type and the document section metadata relate the inserted data to the XYZ Guide Document.

At Release A the document section file will be overwritten if it already exists. Similarly, any supplied document metadata will overwrite existing document metadata. We will work with DAAC personnel to determine what (if any) enhancements are required for long-term implementation at the DAACs post-Release A. In particular, multiple file versions are supported at Release B. DAAC policy will determine how that versioning capability is implemented for documents. Note: only authenticated and authorized science users are permitted to update documents.

As a second example, consider a QA file generated upon analysis of a previously-inserted science data granule. Note: the QA file is of a different data type (e.g., "QA Data") than the science data granule. The software generating the QA file generates a companion metadata file (in PVL format). The metadata file contains the identifier of the science data (Granule ID) upon which QA was performed. The QA data is ingested by one of four mechanisms (DAN, polling, User Ingest GUI, physical media) offered by the Ingest software. The data is prepared, inserted, and stored, and loaded as above. The identifier of the science data upon which QA was performed provides the relationship to the original science data.

At Release A data granules will be overwritten as described above, with potential enhancements post-Release A.

Status Closed

Date Closed 10/9/95

Sponsor Kobler

***** **Attachment if any** *****
