DRAFT

Interface Definition Document for the ESDIS Data Gathering and Reporting System (EDGRS)

August 2000

Deleted: April

National Aeronautics and Space Administration

Goddard Space Flight Center
Greenbelt, Maryland

DRAFT

Table of Contents

Deleted: 1-1
Deleted: 1-1
Deleted: 1-1
Deleted: 1-1
Deleted: 2-1
Deleted: 2-1
Deleted: 2-2
Deleted: 2-2
Deleted: 2-2
Deleted: 2-3
Deleted: 2-4
Deleted: 2-4
Deleted: 3-1
Deleted: 3-1
Deleted: 3-1
Deleted: 3-3
Deleted: 3-4
Deleted: 3-4
Deleted: 3-6
Deleted: 3-6
Deleted: 3-8
Deleted: 3-12
Inserted: 3-12
Deleted: 3-8
Deleted: 3-12
Inserted: 3-12
Deleted: 3-9
Deleted: 3-14
Inserted: 3-14
Deleted: 3-10
Deleted: 3-15
Inserted: 3-15
Deleted: 3-11
Deleted: 3-16
Inserted: 3-16
Deleted: 3-12
Deleted: 3-17
Inserted: 3-17
Deleted: 3-14
Deleted: 4-1
Deleted: 4-1
Deleted: 4-1
Deleted: 4-2

DRAFT

Deleted: 4-2

Deleted: 4-2

DRAFT

Section 1 - Introduction

1.1 Purpose

This document defines the file transfer methods and file formats used to provide data to the Earth Observing System Data and Information System (EOSDIS) Data Gathering and Reporting System (EDGRS). EDGRS is a tool for reporting metrics on science data processing and distribution within EOSDIS.

1.2 Organization

This document is organized as follows:

- Section 1—Purpose and organization of this Interface Definition Document (IDD)
- Section 2—High-level operations concept of the data ingest for EDGRS and the specific file formats
- Section 3—A description of the information needed for EDGRS metrics
- Section 4—High level description for file transfer to EDGRS

1.3 References

The following documents are referenced in or provide supplementary information to this document. They contain relevant supporting information.

1. Computer Sciences Corporation, Earth Science Data Information System (ESDIS) Data Gathering and Report System (EDGRS) White Paper Describing the EDGRS Analysis and Design Approach, Draft, September 1999
2. Raytheon Systems Company, 311-CD-500-001, Release 5A Data Management Subsytem (DMS) Specifications for the ECS Project, July 1999
3. Raytheon Systems Company, 311-CD-501-001, Release 5A Ingest (INGST) Database Design and Schema Specifications for the ECS Project, July 1999
4. Raytheon Systems Company, 311-CD-503-001, Release 5A Planning and Data Processsing (PDPS) Subsystem Database Design and Schema Specifications for the ECS Project, May 1999
5. Raytheon Systems Company, 311-CD-504-001, Release 5A Science Data Server Subsystem Database Design and Schema Specifications for the ECS Project, May 1999

6. Raytheon Systems Company, 311-CD-505-001, Release 5A Storage Management (STGMT) Database Design and Schema Specifications for the ECS Project, May 1999
7. Raytheon Systems Company, 311-CD-506-001, Release 5A Subscription Server (SUBSRV Database Design and Schema Specifications for the ECS Project, May 1999
8. Raytheon Systems Company, 625-CD-506-001, ECS Project Training Material Volume 6: Production Planning and Processing, July 1999
9. Raytheon Systems Company, 625-CD-508-001, ECS Project Training Material Volume 8: Ingest, July 1999
10. Raytheon Systems Company, 625-CD-509-001, ECS Project Training Material Volume 9: Data Distribution, July 1999
11. Raytheon Systems Company, 625-CD-510-001, ECS Project Training Material Volume 10: Archive Processing, July 1999

Section 2 - Operations Concept

2.1 Operations Overview

EDGRS is used to collect and process ingest, archive, and distribution metrics for the Earth Observing System Data and Information System (EOSDIS). The metrics collected will initially be limited to those from the EOS Distributed Active Archive Centers (DAACs) EOS Core System (ECS). In the future, data will be collected from other Earth Science Enterprise (ESE) systems such as the Langley TRMM Information System (LaTIS), the DAAC Version 0 systems and Earth Science Information Partners (ESIPs).

As shown in Figure 2-1, the EDGRS data collection scheme is designed to reside at each data metrics provider. This collection activity extracts data from operational tables into ASCII text flat files on a regular basis. The flat files are transferred via FTP to the centralized EDGRS location and the data is then imported into a database from which standardized reports are generated. Ad-hoc reports are supported through a Web-based user interface.

The flat files prepared for EDGRS are formatted so that one line in the file represents one record of information. The lines have field information in fixed width columns. For DAACs that use ECS, scripts will be provided to create these flat files. Other metrics providers may develop their own processes for creating flat files that conform to the formats specified in this document.

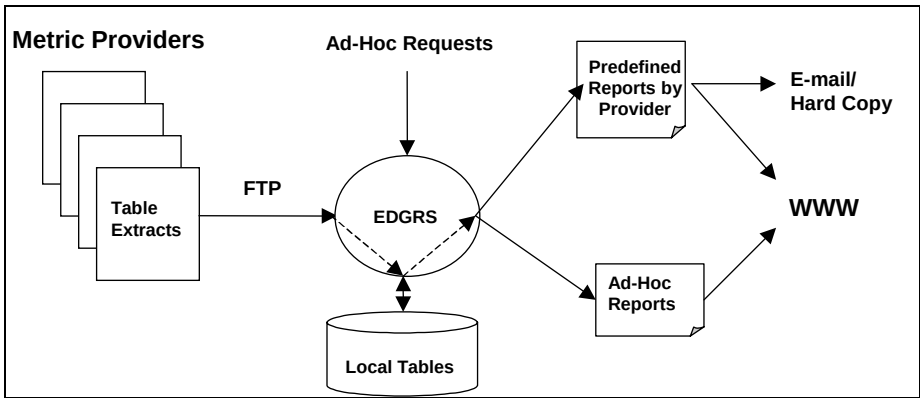

Figure 2-1. Operations Concept

The data extraction and data transfer should be accomplished automatically at the metrics provider location. Once the files arrive at the central EDGRS processing location, the import process is accomplished automatically using scripts on the EDGRS machines. Daily checks by EDGRS operations personnel will ensure that the data was received. The scripts are designed to

pull data with several days of overlap so that if the data transfer fails for a few days, data recovery will take place as soon as communication is restored.

2.2 Flat File Layout

Records from the metrics providers are received in flat files. These flat files are sent via FTP to EDGRS, where scripts read the files and insert the rows as records into the EDGRS database. The flat files must be formatted correctly for this process to work properly.

2.2.1 File Format

There are two types of flat files expected from the metrics data provider. They are data files and an action log file.

Data records from the metrics providers are grouped into ASCII text data files daily. These files conform to a defined format of assigning fixed column positions for each record field. Each field is separated from other fields by a single blank column, including a blank column at the beginning of each record. In addition, each file will have several lines at the top that function as a header. The description of the contents of these header lines follows in Sections 2.2.3 and 2.2.4 of this document, the description of the fixed length fields for each data record follows in Section 3 of this document.

An action log file should be generated that tracks a single execution of the set of scripts covering a single event period (i.e., one execution of all of the scripts). This log file should indicate what scripts were executed, and the date/time that it started/stopped. This is further described in Section 3.

For the present, the widths for each field in the data files correspond to the widths in the ECS files. In the future these may be reduced to reflect the expected data widths in order to reduce the size of the flat files.

2.2.2 File Name

The flat data files will be named using the following format:

"*xx_<mode><tablename>.YYYYMMDDHHMMSS.flt*"

and the action log file will be named using the following format:

"*xx_<mode>_action.YYYYMMDDHHMMSS.flt"*

where

xx_ designates the source of the data (e.g., g0 for Goddard, e0 for EDC), followed by an underscore character.

<mode>	designates the operations or test mode. Three valid values are defined as follows: 'OPS' means data taken from the operational production environment. 'TS1' means data is taken from an integration test environment. 'TS2' means data is taken from a system test environment.
<tablename>	designates the primary table for which the data is intended. See Section 3 for a description of the expected values for this parameter.
Period	A period (dot) follows the <tablename> in the data file format and the characters 'action' in the action log file name.
YYYY	designates the 4 digit year.
MM	designates the 2 digit month, 01 through 12
DD	designated the 2 digit day, 01 through 31
HH	designates the 2 digit hour in 24-hour time, 00 through 23
MM	designates the 2 digit minutes , 00 through 59
SS	designates the 2 digit seconds, 00 through 59

Note that HHMMSS is in 24-hour time and designates the time the flat file was produced. For example, the corresponding action file for a run at the Goddard DAAC production environment that takes place on November 19, 1999 12:25:47am would be represented as follows: "*g0_OPS_action.19991119002547.flt*". See Section 3 - Data Descriptions for sample file names.

Deleted: *19991119002547out*

2.2.3 File Version

The first line of the file indicates the version of the software or script that was used to generate and format the file. It has the following format:

<tablename><space>xx.yy.zz

where

<tablename>	designates the primary table for which the data is intended. See Section 3 for a description of the expected values for this parameter. This value should match the one used in the file name.
<space>	represents a single space
xx	designates the major version number. This should change only when the major content of the file is changed such as a change in the data source or the main structure of the file changes such as from blank filled to comma delimited.

Yy designates the minor version number. This should change only when a minor change is made, such as a field is added or removed but the data source remains unchanged.

Zz designates the fix number. This should change only when a field's length or format is adjusted.

2.2.4 Mandatory Data Marker

Since the data is sent in ASCII format, it is readable. This provides an easy method for checking the contents of a file. To assist in the reading of a file, it is preferred, but not necessary, to provide a descriptive name at the top of each column of data. Following this line of column names is a line of dashes and spaces, indicating the width of each column. While the column name is not required, the import program looks for a line containing dashes (at least 3 contiguous dashes) before it starts reading data. A line of text containing these dashes will trigger the import program to start reading data on the line immediately following the dashed line. There may be more than one line of dashes in the first 100 lines in the file; however, the last line of dashes within the first 100 lines triggers the data importing starting on the next line.

2.2.5 Data Types

Although the files are full ASCII text files, the values represented in each column can portray various data types, such as numeric values, dates, and text strings. A single blank separates each column from the next. Also, each data record of each flat file starts with a blank column.

Alphanumeric fields can contain any combination of letters, numbers or ASCII symbols. Loading is based on fixed width columns. Special formatting is specified. Blank-filled fields are preferred when no data is available; however, use of the "NULL" keyword may also be used.

Justification for all fields should be consistent within the field.

Number fields have no leading or trailing zeros unless specified. Number fields may be either left or right justified and are blank-filled. The time reference is local time for system generated dates and GMT for dates indicating temporal coverage of the science data. The date fields are formatted in either one of two styles.

Format 1: MMM DD YYYY HH:MM:SS XM where MMM is the first three characters of the month name (e.g., JAN for January) and X = A or P for AM or PM. The HH portion is in 12-hour time. The SS portion is optional. All other symbols follow the conventions defined in Section 2.2.2.

Format 2: MM/DD/YYYY HH:MM:SS The hours are in 24-hour or military time.

(e.g. 02/16/2000 02:45:00)

Deleted: ¶
¶
¶

Formatted: Bullets and Numbering

Section 3 - Data Descriptions

This section describes in detail the flat files used in EDGRS processing. Systems that are based on ECS will have these files generated by EDGRS scripts. Systems that are not based on ECS will need to generate files that follow these formats.

Each row in each flat file will be of a fixed length and will represent one record comprised of fixed-width fields. The rows of data will be imported from the flat files into the EDGRS database.

The flat files will provide metrics on data ingest, data processing, data archiving, and data distribution.

3.1 Data Ingest

The data ingest function at a metrics provider receives Earth science data from various sources. The data is transferred into the system, preprocessed, and transferred to a repository. A granule is the smallest aggregation of data that is independently managed (i.e., described, inventoried, or retrieved).

The data ingest process is broken down into ingest request records and requested granule records. An ingest request is a request to the system to import a data set (i.e. related set of granules) into the computer. It can have one or more granules associated with the request. A common request identifier links these records.

Formatted: Bullets and Numbering

3.1.1 Ingest Requests - Summary Level

This file will contain summary data for completed ingest requests. Each row corresponds to one ingest request and has a one-to-many correspondence with the rows in the Ingest Request - Granule Level file. Following is the description of each field in the group.

Sample file name: g0_OPSInReqHdr.20000215110239.flt

Note that the processing of multiple granules in a request may occur in parallel.

Script Column Name	Column Description	Column Width	Start Position	End Position
RequestID	Key that uniquely identifies each ingest request. This value is alphanumeric and must be distinct within this table. This key is used to link with records in the Ingest Request - Granule Level file. This field cannot be empty. (Alphanumeric, Sequential, no leading spaces)	11	2	12
ExternalDataProvider	Keyword name of the system providing the data to be ingested (Alphanumeric, no embedded spaces) (e.g. MODAPS, EDOS, GSFC-V0)	20	14	33
Mission	Keyword name of the mission that the data came from (Alphanumeric, no embedded spaces) (e.g. AM-1)	60	35	94
ProcessingStartDateTime	Timestamp for start of ingest for the first data granule in the request, accurate to at least minutes. (Local time: Format 1)	26	96	121
ProcessingEndDateTime	Timestamp that ingest processing ended for the last granule processed in the request, accurate to at least minutes. (Local time: Format 1)	26	123	148
TimeToXfer	Time to transfer the data into the system in seconds. This is the sum of the corresponding values in the Ingest Request - Granule Level file. (Integer)	11	150	160
TimeToPreprocess	Time from start of preprocessing of request to time of completion (success or failure) of preprocessing in seconds. This is the sum of the corresponding values in the Ingest Request - Granule Level file. Preprocessing converts data to binary format, and may yield ancillary data needed for further processing. This is done at the same time as metadata extraction. (Integer)	16	162	177
TimeToArchive	Time to archive the data in seconds. This is the sum of the corresponding values in the Ingest Request - Granule Level file. (Integer)	13	179	191

Script Column Name	Column Description	Column Width	Start Position	End Position
TotalDataVolume	Total data volume of the ingest in bytes. This is a sum of the data volumes of the composite granules.	20	193	212
TotalFileCount	Total number of files generated to fulfill the request (Integer)	14	214	227
TotalGranuleCount	Total number of granules for the request. (Integer)	17	229	245
TotalSuccessfulGranules	Total number of data granules successfully ingested (Integer)	23	247	269

Formatted: Bullets and Numbering

3.1.2 Ingest Requests - Granule Level

This file will contain metrics for the granules ingested. Each row corresponds to one granule in an ingest request and has a many to one correspondence with the rows in the Ingest Request - Summary Data file. Note that since the granules may be concurrently ingested, the start/stop times of multiple granules in a request may overlap. Following is the description of each field in the record:

Sample file name: g0_OPSInReqData.20000216004503.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
RequestID	Key that uniquely identifies each ingest request. This must match corresponding records in the Ingest Request - Summary Data file to link associated data. (Alphanumeric, Sequential)	11	2	12
DataGranuleID	Key that uniquely identifies each granule within a request ID. (Sequential integers)	13	14	26
DataType	The short name / descriptive acronym for a data type. This is the official reference name for the contents of a data collection. (alphanumeric, no embedded spaces) (e.g. AST0TS, MOD000, AM1DIAG1, AM1DIAG1.)	32	28	59
DataGranuleVolume	Total data volume of data granule in bytes This is a sum of the data volumes of the files comprising the data granule	20	61	80

Script Column Name	Column Description	Column Width	Start Position	End Position
DataGranuleState	State of the data granule. (Valid values are 'ArchErr', 'Archived', 'Cancelled', 'New', 'PreprocErr', 'Preprocessed', 'Terminated', 'Transferred', 'XferErr')	20	82	101
ProcessingStartDateTime	Processing start date and time for the ingest of a data granule, accurate to the minute. (Local time: Format 1*)*	26	103	128
ProcessingEndDateTime	Processing end date and time for an ingest of a data granule, accurate to the minute. Processing includes all activities required to import, preprocess, and archive the granule. (Local time: Format 1*)*	26	130	155
TimeToArchive	Time to archive this data granule in seconds. (Integer)	13	157	169
TimeToPreprocess	Time in seconds from start of preprocessing of data granule to time of completion (success or fail) of preprocessing (Integer)	16	171	186
TimeToXfer	Time in seconds from the start of transfer for first file in granule to time of receipt of status (success or fail) for last file in granule. (Integer)	11	188	198

3.2 Data Processing

At the DAACs, science data is processed or reprocessed either automatically, as data is received, or upon demand. The process begins with a production request (PR) being entered into the system. Data availability and system resources are considered. The PRs are resolved into data processing requests (DPRs) which execute program generation executives (PGEs). PGEs are the lowest level modules and contain the science algorithms used to produce data products. To keep track of this, EDGRS needs to know the definitions of the data (Data Type Master), the granules that are produced (Data Granule Metrics), and the processing requests that were executed (Data Processing Requests Completed, Data Processing Requests, and Data Processing Request Data).

Formatted: Bullets and Numbering

3.2.1 Data Type Master File

This file will contain a master listing of the Earth science data types and their mapping to instruments and platform.

Sample file name: g0_OPSPlDataTyp.20000215110239.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
DataTypeId	Identifier for the data type. It should include the data type name and version in the following format: (alphanumeric, no spaces) <datatype-name>#<version> (eg. AM1ANC#001, DFLAXENG#001)	20	2	21
Data Type Name	The short name / descriptive acronym for a data type. This is the official reference name for the contents of a data collection. (alphanumeric, no spaces) (eg. AST0TS, MOD000, AM1DIAG1, AM1DIAG1)	20	23	42
DataTypeDescription	A textual description of the data type. (alphanumeric string, spaces allowed)	60	44	103
Instrument	Short Name or acronym by which an instrument is known. This instrument generates this data type. (alphanumeric, no spaces) (eg. ASTER, MODIS)	20	105	124
PlatformName	The spacecraft/satellite or other vehicle that contains this instrument that generated this data type. (alpahnumeric) (eg. AM-1)	25	126	150
Version	Version of the data type. A change in the software algorithm that generates or processes this data represents a change in version. (Numeric, 3 digits, zero padded)	20	152	171
ArchiveCenter	Acronym defining the location where the data type is archived (Alpha) (Valid values are 'ASF', 'EDC', 'GSF', 'JPL', 'LAR', 'NSC','ORN', 'CSN')	20	173	192
ProcessingLevel	Processing level of the data (Alphanumeric) (Valid entries are 'L0', 'L1', 'L1A', 'L1B', 'L2', 'L3', 'L4')	15	194	208

3.2.2 Data Granule Metrics

Formatted: Bullets and Numbering

This file will contain metrics regarding the data granules involved in data processing.

Sample file name: g0_OPSPlDgShort.20000216004503.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
GranuleId	Key that uniquely identifies each data granule. The first portion of the granule ID should include the data type ID (Alphanumeric, no spaces). (eg. AM1ANC#00103172000000000000)	130	2	131
DataTypeId	Identifier for the data type. It should include the data type name and version in the following format: (alphanumeric, no spaces). This field matches the same name field in the Data Type Master file defined in the previous section. <datatype-name>#<version> (eg. AM1ANC#001, DFLAXENG#001)	20	133	152
UniversalReference	Universal reference associated with this data granule. (Alphanumeric, no spaces) The final format is TBD.	254	154	407
StartTime	Start date and time for the data granule (GMT: Format 1)	26	409	434
StopTime	Stop date and time for the data granule(GMT: Format 1)	26	436	461
Version	Used to distinguish between two granules having same data type and start time as the result of reprocessing.	11	463	473
InsertTime	Time of original insertion (Local time: Format 1)	26	475	500
TimeStamp	Date and time this record was last modified. (Local time: Format 1)	26	502	527

3.2.3 Data Processing Requests Completed

Formatted: Bullets and Numbering

This file will contain metrics on completed DPRs, with one record per DPR. ECS provides detailed information on system resources expended to produce data products.

Sample file name: g0_OPSDprs.20000216004503.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
DprId	Key that uniquely identifies the DPR (Alphanumeric, no spaces) (eg. MoPGE01#2017191500OPS)	29	2	30
CompletionDate	Date that the DPR was completed (Local time, Format 1)	26	32	57
DprElapsedTime	Elapsed time of DPR execution in seconds. Fractional seconds are reported to up to 8 digits beyond the decimal.	20	59	78
PgeElapsedTime	Elapsed time of PGE execution in seconds. Fractional seconds are reported to up to 8 digits beyond the decimal.	20	80	99
PgeBlockInputOperations	Number of block input operations (Integer)	23	101	123
PgeBlockOutputOperations	Number of block output operations (Integer)	24	125	148
PgeSwaps	Number of swaps during PGE execution (Integer)	11	150	160
PgeMaxMemoryUse	Maximum memory used by PGE (Accurate to one decimal place, units TBD)	20	162	181
PgeSharedMemoryUse	Amount of shared memory used by the PGE (Accurate to one decimal place, units TBD)	20	183	202
PgeCPUTime	Amount of CPU time used by the PGE in seconds. Fractional seconds are reported to up to 8 digits beyond the decimal.	20	204	223
SystemCPUTime	Amount of system time spent during execution in seconds. Fractional seconds are reported to up to 8 digits beyond the decimal.	20	225	244
PgePageFaults	Number of page faults during PGE execution (Integer)	13	246	258
Platform	Name of the processor used for this execution (Alphanumeric) (eg. g0spg01)	60	260	319
ExitCode	Final exit condition returned by PGE to the processing system. This should be an integer value where 0 is the success return value. Other error codes are specific to the PGE developer.	20	321	340

3.2.4 *Data Processing Requests*

This file will contain additional information about data processing requests.

Sample file name: g0_OPSPJDataProReq.20000824004504.flt

Formatted
Formatted
Formatted
Formatted
Formatted

Script Column Name	Column Description	Column Width	Start Position	End Position
DprId	Key that uniquely identifies the DPR (Alphanumeric, no spaces) (eg. MoPGE01#2017191500OPS) This key must match the key used in the data processing requests completed table. A record here should align with a record in that table.	29	2	30
PgeId	Identifier for the production program. No assumptions are made on the format other than uniqueness.	17	32	48
DataStartTime	Start date and time for the data granule (GMT: Format 1)	30	50	79
DataStopTime	Stop date and time for the data granule(GMT: Format 1)	30	81	110
CompletionState	A status indicator describing the current status of the DPR. Valid values include the following (with * by completion states). Other values are negotiable. CANCELED * CQ_HOLD CQ_RELEASE FAILEDPGE * FAILPGEDEL * NULL (Unknown state) ON_QUEUE STARTED SUCCESS * SUCC_DEL *	10	112	121

3.2.5 *Data Processing Request Data*

This file will contain additional information about data processing requests. It is primarily used to obtain the granule identifer. Many fields in this table are not currently used by EDGRS, as indicated with TBD*.

Sample file name: g0_opspldprdata.20000824004504.flt

Formatted

Script Column Name	Column Description	Column Width	Start Position	End Position
DprId	Key that uniquely identifies the DPR (Alphanumeric, no spaces) (eg. MoPGE01#2017191500OPS) This key must match the key used in the data processing requests completed table. A record here should align with a record in that table.	29	2	30
completionDate	Date that the DPR was completed (Local time, Format 1). This should match the associated completion date/time in the data processing requests completed table.	29	32	60
granuleId	Key that uniquely identifies each data granule. The first portion of the granule ID should include the data type ID (Alphanumeric, no spaces). This is the same format as described in Section 3.2.2. (eg. AM1ANC#00103172000000000000)	130	62	191
logicalId	Numeric identifier to distinguish among distinct PGE/dataType combinations. Format TBD*.	10	193	202
primaryType	ID of the primary data type for this alternate input. Format TBD*.	10	204	213
accepted	Indicates whether the data granule, when available, has passed metadata checks. Format TBD*.	10	215	224
theOrder	Order number associated with production request. Format TBD*.	10	226	235

Formatted

Script Column Name	Column Description	Column Width	Start Position	End Position
type	The type (int, string, etc.) of the metadata parameter to be compared. Format TBD*.	10	237	246
temporalFlag	A flag indicating whether the most recent alternate input can be used if a primary granule of the same data type cannot be found for the production request time frame. Format TBD*.	10	248	259
timeWait	Length of time to wait for this data type. Format TBD*.	10	261	270
ioFlag	Indication as to whether metatdata check should be performed on input or output. Format TBD*.	10	272	281
timerExp	Indicates whether the timer has expired. Format TBD*.	10	283	292
timerStart	Indicates whether the timer has started. Format TBD*.	10	294	303
numNeeded	The number of inputs required by the PGE. Format TBD*.	10	305	314
waitForFlag	Indicates whether DPR should be released after last timer, even if the chain is incomplete. Format TBD*.	10	316	325
linkId	Reserved for future use. Format TBD*.	10	327	336
minGranReq	Minimum number of granules of this input type required by the PGER when forming a DPR. Format TBD*.	10	338	347
* TBD means that insufficient ECS document is available to further define this item				

3.3 Data Archiving

Data received from ingest or processing are archived for permanent storage and distribution. It is primarily an automated process. Granules placed into the data archive are identified in a table

called Archived Granules. This file contains information on archived data granules. It will contain one record for each granule archived.

Sample file name: g0_OPSDsMdGran.20000215110239.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
DbID	Key that uniquely identifies a granule description record. (Alphanumeric)	19	2	20
ShortName	The short name / descriptive acronym for a data type. This is the official reference name for the contents of a data collection. (alphanumeric, no spaces). This should be the same form as data type in Sections 3.1.2 and 3.2.1. (eg. AST0TS, MOD000, AM1DIAG1, AM1DIAG1.)	9	22	30
SizeMBECSDataGranule	Volume of data contained in the granule in megabytes. This should be accurate to 6 digits to the right of the decimal (bytes).	25	32	56
BeginningDateTime	Time archiving began (Local time: Format 1)	26	58	83
EndingDateTime	Time archiving ended (Local time: Format 1)	26	85	110
InsertTime	Time of original insertion (Local time: Format 1)	26	112	137
ProductionDateTime	Date and time specific granule was produced by a PGE (Local time: Format 1)	26	139	164
LastUpdate	Time of last update (Local time: Format 1)	26	166	191

3.4 Data Distribution

After data has been archived, it can be distributed. Distribution from the archive is done for multiple reasons. This includes in response to user orders, as triggered by previously submitted subscriptions, or in response to requests by PGEs to process data. Data can be retrieved or transmitted electronically or issued on physical media.

Formatted: Bullets and Numbering

3.4.1 Distribution Requests

This file will contain one record per distribution request. There is a one-to-many relationship between the rows in this file and the rows in the Distribution Granules file.

Sample file name: g0_OPSDsDdRqst.20000215110239.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
RequestId	Key that uniquely identifies each distribution request. This key links this table to the Distribution Granules table defined in the next section. It also links this table to the request ID in the User Request Table (c.f. Section 3.4.4) (Alphanumeric)	50	2	51
OrderId	Key that uniquely identifies each user order This must match corresponding records in the User Orders file to link associated data. (Alphanumeric)	50	53	102
State	Queue state of the distribution request object (Valid entries are 'Pending', 'Active', 'Shipped', 'Failed', 'Suspended','Suspended With Errors', 'Cancelled', 'Transferring'.)	50	104	153
UserID	User ID of the user initiating the request. This ID should match the User ID in the User Profile Table (c.f. Section 3.4.5). It can also be a "guest" account that should have a generic "guest" identifier. Other values are allowed if they have a defined purpose and are used consistently, such as use by a system level staging program for use by production processing. These values need to be defined in advance and distinguished as a special case. (Alphanumeric)	50	155	204
SizeInMB	Total bytes in distribution request	20	206	225
StartTime	Time the distribution started (Local time: Format 2)	25	227	251
EndTime	Time distribution ended (Local time: Format 2)	25	253	277
NRGranules	Number of granules per request (Integer)	11	279	289
NRReqfiles	Number of files generated to fill request (Integer)	11	291	301
NRMedia	Number of media of type Media Type used to fill request. (Integer)	11	303	313

Script Column Name	Column Description	Column Width	Start Position	End Position
MediaType	Type of media used for a request (i.e. 8MM tape, CDRom). For FTP transfers, use the values *FtpPush* for system or subscription based transfers and *FtpPull* for requests from the User Request area. (Alphanumeric)	50	315	364
UserProfile	Profile ID. This is currently the same value as the UserID field.	50	366	415
FTPHost	Hostname to connect to for FTP push (Alphanumeric)	255	417	671
FTPPushDest	Target system directory (Alphanumeric)	255	673	927
FTPPullHost	Host user will ftppull from (Alphanumeric)	255	929	1183
ESDTType	The short name / descriptive acronym for a data type *and version*. This is the official reference name for the contents of a data collection. (alphanumeric, no spaces) It should be in the following format: <datatype-name>.<version> (e.g. MOD02OBC.001)	50	1185	1234

3.4.2 Distribution Granules

This file will contain one record for each granule distributed. There is a many-to-one relationship between the rows in this file and the rows in the Distribution Requests file.

Sample file name: g0_OPSDsDdGran.20000216004503.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
RequestID	Key that uniquely identifies each distribution request. This key links these granule records to the Distribution Requests Table.	50	2	51
Status	Status of request (valid values are 'Pending', 'Staging', 'Shipped', 'Failed', 'Suspended With Errors', 'Suspended', 'Transferring').	150	53	202
StartTime	Time the distribution started (Local time: Format 2)	25	204	228

Script Column Name	Column Description	Column Width	Start Position	End Position
GranuleSize	Sum of sizes of files in granule in bytes	20	230	249
GranuleID	Key that uniquely identifies each data granule. This is the same format as the Universal Reference field in Section 3.2.2	150	251	400
ESDType	Earth science data type (ESDT). All granules in a request pertain to the same ESDT. It should include the data type name and version in the following format: <datatype-name>.<version>	50	402	451

3.4.3 User Orders

This file will contain one record for each user order. There is a one-to-many relationship between the rows in this file and the rows in the User Requests file.

Sample file name: g0_OPSEcAcOrd.20000215110239.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
OrderId	Key that uniquely identifies each user order This must match corresponding records in the User Requests file to link associated data. (Alphanumeric, Sequential)	10	2	11
HomeDAAC	Where the data request was placed (Valid values are 'ASF',' EDC', 'GSF', 'JPL', 'LAR', 'NSC', 'ORN', 'CSN')	13	13	25
UserId	UserID of user placing order. This identifier must match the one in the User Profile Table defined below. (Alphanumeric)	12	27	38
FirstName	User's first name (Alphanumeric)	20	40	59
MiddleInit	User's middle initial (Alphanumeric)	10	61	70
LastName	User's last name (Alphanumeric)	20	72	91
EmailAddr	User's email address (Alphanumeric)	255	93	347
OrderStatus	Status of user order. (Valid entries are 'Abort', 'Operator Intervention',' Pending',' Shipped', 'Terminated')	22	349	370
OrderMedia	Media type of the user's order (Valid Entries are: 8mm tape, CD-ROM, FTP)	20	372	391

Script Column Name	Column Description	Column Width	Start Position	End Position
OrderGranule	The number of granules that fill the user's order (Integer)	12	393	404
ReceiveDateTime	Date order received (Local Time: Format 1)	26	406	431
StartDateTime	Time DDIST first started processing request (Local Time: Format 1)	26	433	458
FinishDateTime	Time DDIST finished processing request (Local Time: Format 1)	26	460	485
TimeOfLastUpdate	Time of last update (Local Time: Format 1)	26	487	512
ShipDateTime	Date product shipped (Local Time: Format 1)	26	514	539

3.4.4 User Requests

This file will contain one record for each user request. There is a many-to-one relationship between the rows in this file and the rows in the User Orders file.

Sample file name: g0_OPSEcAcRqst.20000215110239.flt

Script Column Name	Column Description	Column Width	Start Position	End Postion
OrderId	Key that uniquely identifies each order. This key must match the orderid in the User Order Table. (Alphanumeric)	10	2	11
OrderHomeDaac	Where the request was ordered	10	13	22
RequestId	Key that uniquely identifies each user request. This key must match the one in the Distribution Request Table (c.f. Section 3.4.1). (Alphanumeric)	10	24	33
RequestProcessingDAAC	Where the request was processed (valid values are 'ASF', 'EDC', 'GSF', 'JPL', 'LAR', 'NSC', 'ORN', 'CSN')	22	35	56
FirstName	User's first name (Alphanumeric)	20	58	77
MiddleInit	User's middle initial (Alphanumeric)	10	79	88
LastName	User's last name (Alphanumeric)	20	90	109
EMailAddr	User's email address (Alphanumeric)	255	111	365

Script Column Name	Column Description	Column Width	Start Position	End Postion
RequestStatus	Status of request (Valid entries are 'Pending', 'Staging', 'Shipped', 'Failed', 'Suspended With Errors', 'Suspended', 'Transferring').	22	367	388
NumFiles	Number of files that fills request (Integer)	12	390	401
NumBytes	Number of bytes that fill request (Accurate to one decimal place)	20	403	422
NumGranule	Number of granules that fill request (Integer)	12	424	435
MediaType	Media type (FtpPull, FtpPush, 8MM, etc.)	20	437	456
ShipAddrStreet1	Street shipping address #1 (Alphanumeric)	32	458	489
ShipAddrStreet2	Street shipping address #2 (Alphanumeric)	32	491	522
ShipAddrStreet3	Street shipping address #3 (Alphanumeric)	32	524	555
ShipAddrCity	City shipping address (Alphanumeric)	35	557	591
ShipAddrState	State shipping address (Alphanumeric)	20	593	612
ShipAddrZip	Zip code of shipping address (Alphanumeric)	15	614	628
ShipAddrCountry	Country of shipping address (Alphanumeric)	30	630	659
ShipAddrPhone	Phone # at shipping address (format XXX-XXX-XXXX)	22	661	682
ReceiveDateTime	The user request was created (Local time: Format 1)	26	684	709
StartDateTime	Time that the system began processing the request (Local time: Format 1)	26	711	736
FinishDateTime	Time that the system finished processing request (Local Time: Format 1)	26	738	763
TimeOfLastUpdate	Time of last update (Local time: Format 1)	26	765	790
ShipDateTime	Date product shipped (Local time: Format 1)	26	792	817
FTPAddress	Requestor's FTP staging address (Alphanumeric)	50	819	868
DestinationNode	Destination node for FTP acquires (Alphanumeric)	20	870	889
DestinationDirectory	Destination directory for FTP acquires (Alphanumeric)	20	891	910

3.4.5 User Profile

This file will contain one record per registered user.

Sample file name: g0_OPSUsrProfile.20000215110239.flt

Script Column Name	Column Description	Column Width	Start Position	End Position
UserId	Uniquely identifies a registered user. (Alphanumeric)	12	2	13
HomeDAAC	Name of a location where the user is registered. (valid values are 'ASF', 'EDC', 'GSF', 'JPL', 'LAR', 'NSC', 'ORN', 'CSN')	10	15	24
FirstName	User's first name (Alphanumeric)	20	26	45
MiddleInit	User's middle initial (Alphanumeric)	10	47	56
LastName	User's last name (Alphanumeric)	20	58	77
UsrType	User's email address (Alphanumeric)	20	79	98
EmailAddr	User's email address (Alphanumeric)	255	100	354
InternetAffiliation	User's Internet Affiliation. Valid values are: Government, K-12, University, Commercial, Other	19	356	374
Organization	User's organization (Alphanumeric)	31	376	406
ProjectName	User's project name (Alphanumeric)	30	408	437
Affiliation	User's affiliation. (Valid Values are 'Gov. Research', 'Government', 'Other', 'Commercial', 'University', 'Null')	16	439	454
ResearchField	Research field available in the system. Optional free text entry by user. (eg. Global Biosphere GSFC, Land Processes EDC, Atmospheric Aerosols LaRC, Biological Oceanography JPL	64	456	519
CreationDate	Date/time this user id/profile was created. (Local time: format 1)	26	539	564
ShipAddrStreet1	Street shipping address #1 (Alphanumeric)	32	566	597
ShipAddrStreet2	Street shipping address #2 (Alphanumeric)	32	599	630
ShipAddrStreet3	Street shipping address #3 (Alphanumeric)	32	632	663
ShipAddrCity	City shipping address (Alphanumeric)	35	665	699
ShipAddrState	State shipping address (Alphanumeric)	20	701	720

Script Column Name	Column Description	Column Width	Start Position	End Position
ShipAddrZip	Zip code of shipping address	15	722	736
ShipAddrCountry	Country of shipping address (Alphanumeric)	30	738	767
ShipAddrPhone	Phone# at billing address	22	769	790
MailAddrStreet1	Street Mailing address #1 (Alphanumeric)	32	792	823
MailAddrStreet2	Street Mailing address #2 (Alphanumeric)	32	825	856
MailAddrStreet3	Street Mailing address #3 (Alphanumeric)	32	858	889
MailAddrCity	City Mailing address (Alphanumeric)	35	891	925
MailAddrState	State Mailing address (Alphanumeric)	20	927	946
MailAddrZip	Zip code of Mailing address	15	948	962
MailAddrCountry	Country of Mailing address (Alphanumeric)	30	964	993
MailAddrPhone	Phone# at Mailing address	22	995	1016
NasaUser	Does the user work for NASA (Y) or not (N)	8	1018	1025

3.5 Log File Format

An action log file should be generated that tracks a single execution of the set of scripts covering a single event period (i.e., one execution of all of the scripts). This log file should indicate what scripts were executed, and the date/time that it started/stopped. The suggested format for this file is as follows. The sequencing and format of the lines will be used for text extraction.

Line	Item Description	Example
1-5	File Header describing the version of the control script, the time of execution of the script on the system in datestamp format and full date format using local time.	EDGRS Development Version 0.1.0 --> Running Now The datestamp for this execution is: 20000824004504 Thu Aug 24 00:45:04 EDT 2000 < 2 blank lines>
6	DAAC/Data Provider <g> where g is g for gsfc, l for larc, e for edc, and n for nsidc for ECS systems. TBD for other systems.	EDGRS is running on <g> DAAC.
7	Earliest date for which metrics are pulled.	Initial Date of Pull Window is: 08/20/2000
8	Latest date for which metrics are pulled.	Terminal Date of Pull Window is: 08/24/2000

Formatted
Formatted
Formatted

Line	Item Description	Example
9	The mode of the system being polled. 'OPS' is used for the production system. 'TS1' and 'TS2' are used for test systems.	The program is running in Mode: OPS
10-12	Statement surrounded by blank lines	Output file for the scripts executed above may be viewed in the corresponding directory.
13-n	For each pull script run, the log file should contain the following sublines: 1. Date/time execution started 2. Executing Script: <script name> At the end of all scripts, the time of the completion of the last script is also recorded.	Thu Aug 24 00:45:05 EDT 2000 Executing Script: Dprs.ksh Thu Aug 24 00:45:08 EDT 2000 Executing Script: PlDprData.ksh Thu Aug 24 00:45:41 EDT 2000 Executing Script: DsDdGran.ksh Thu Aug 24 00:45:42 EDT 2000 Executing Script: DsDdGranArch.ksh Thu Aug 24 00:56:18 EDT 2000 Executing Script: DsMdGran.ksh Thu Aug 24 00:57:04 EDT 2000 Executing Script: EcAcOrd.ksh Thu Aug 24 00:57:05 EDT 2000 Executing Script: EcAcRqst.ksh Thu Aug 24 00:57:06 EDT 2000 Executing Script: InReqData.ksh Thu Aug 24 00:57:21 EDT 2000 Executing Script: InReqHdr.ksh Thu Aug 24 00:57:24 EDT 2000 Executing Script: PlDataTyp.ksh Thu Aug 24 00:57:24 EDT 2000 Executing Script: PlDataProReq.ksh Thu Aug 24 00:57:30 EDT 2000 Executing Script: PlDgShort.ksh Thu Aug 24 00:58:54 EDT 2000 Executing Script: UsrProfile.ksh Thu Aug 24 00:58:54 EDT 2000 Executing Script: DsDdRqst.ksh Thu Aug 24 00:58:55 EDT 2000 Executing Script: DsDdRqstArch.ksh Thu Aug 24 01:14:03 EDT 2000
Last	Concluding statement indicating everything completed.	EDGRS Development Version 0.1.0 --> Run Completed

Formatted

Formatted

Formatted

Formatted: Bullets and Numbering

Section 4 - Data Delivery

This section addresses the methodology for file delivery from the metrics providers to the EDGRS environment.

Formatted: Bullets and Numbering

4.1 Directories and Folders

EDGRS interfaces with the metrics providers via ASCII text files. These files are FTP 'd from the metrics providers into assigned folders on the EDGRS system.

Each ECS DAAC (i.e. GSFC, NSIDC, LaRC, and EDC) will execute EDGRS provided scripts to create the ASCII text files. The scripts will automatically connect the DAAC to the EDGRS system at IP address 192.86.19.170 and push the flatfiles to reserved directories on the EDGRS system as follows:

Deleted: execute EDGRS

D:\ Edgrs\Flatfiles\Gsfc - This is where all GSFC flatfiles are currently being FTPed.

D:\ Edgrs\Flatfiles\Nsidc - This is where all NSIDC flatfiles are FTPed.

Deleted: will be

D:\ Edgrs\Flatfiles\Larc - This is where all LaRC flatfiles are FTPed.

Deleted: will be

D:\ Edgrs\Flatfiles\Edc - This is where all EDC flatfiles are FTPed.

Deleted: will be

D:\ Edgrs\Flatfiles\LaTIS - This is where all LaTIS flatfiles are FTPed.

To avoid the simultaneous transfer of data from multiple DAACs , a staggered FTP windowing scheme will be employed. Each DAAC's schedule will be resolved upon installation of the EDGRS software.

Deleted: ¶

Metric providers not using ECS will independently generate ASCII text files formatted in accordance with this document. The process for transferring those files can be similar to that used by ECS. The specific folders and other details will be specified in separate operations agreements between EDGRS operations and the provider organization.

Formatted: Bullets and Numbering

4.2 Login Procedures/Accounts

For the EDGRS scripts to extract data from the ECS tables and send it to the centralized location, the following must occur:

- An account must be set up with the authorization to query (via Sybase SQL embedded into UNIX scripts) the ECS tables and create files in a place from which they can be FTP'd.
- The scripts that pull the data must be installed in this account.

- A UNIX script that initiates the queries must be installed. This script should be set up to initiate automatically on a predefined interval via the 'cron' function or with a script that sets up the next execution automatically.
- A UNIX script that initiates the FTP of the data retrieved in the prior step should be installed to run in a periodic manner as above.

System accounts and processes for non-ECS organizations will probably be similar as determined by the organization sending the data. The EDGRS ID and password to use when FTP'ing data to EDGRS will be covered by individual operations agreements between the metrics provider and EDGRS.

Formatted: Bullets and Numbering

4.3 Time span and overlap (data content)

The number of days of data will be provided in each flat file should be determined by each metrics provider in consultation with EDGRS personnel. This will ensure that data is available in the event of system down time or missed file transfers. The trade off between data volumes to be transmitted and the need for recovery in the event of transmission problems or system down time can be determined on a site by site basis. For the EDGRS-provided scripts, this is a configurable variable and can be adjusted if necessary.

Formatted: Bullets and Numbering

4.4 Periodicity of file delivery

Files will be transmitted on a daily basis. Some scripts obtaining data from the ECS environments may execute hourly in order to ensure retrieval of data before its deletion. Metrics providers querying their own databases to generate their own ASCII text files should plan to do so on a basis that assures that metrics are collected before the data is purged from the providing databases. Transmission frequency can be negotiated on a case-by-case basis but is preferred to be daily.

Formatted: Bullets and Numbering

4.5 Points of Contact

To ensure efficient data acquisition, communication between EDGRS and personnel at the metrics providers may be necessary, and the following positions are defined.

- **EDGRS Point of Contact (POC)** - representative designated by the EDGRS operational/development team for interfacing with the DAACs and other metrics providers on any required issues.
- **DAAC POC Coordinator** - individual responsible for coordinating between the EDGRS POC, the DAAC POCs, and the ECS development/maintenance team. The level of involvement may change depending on the current activities.

Deleted: and

Deleted: That individual will also coordinate between the other DAAC POCs and the ECS development/maintenance team.

- **DAAC POC(s)** - individual at each DAAC or Metrics Provider responsible for their environment. If there is no DAAC POC Coordinator assigned, the EDGRS POC will communicate directly with the DAAC POCs.
- **Other POCs** – individual at another metrics provider responsible for coordination between that system and the EDGRS development/maintenance team.

The list of POCs will be maintained on a Web page available via http://edgrs.gsfc.nasa.gov:8000/. As the need to change POCs arises, updates should be provided to the EDGRS POC.

Field Code Changed

Deleted: http://edgrs.gsfc.nasa.gov