

The ESDIS Metric System
Custom Log Implementation Guide

Version 1.01

Table of Contents

1	REFERENCE MATERIALS	1
1.1	Document Version Control	1
1.2	Master Documents	1
2	Introduction	2
3	Custom Log File Format	3
3.1	Requirements for Custom Logs	3
4	Custom Log Files Naming Convention	7
5	Data File Manifest File Format	9
5.1	Data File Manifest File Naming Convention	10
5.2	Data File Manifest File Updates	10
5.3	Data Files Interface Mechanism	10

Table 1	Custom Log File Format	5
---------	------------------------------	---

Table 3	Data File Manifest File Format.....	9
---------	-------------------------------------	---

Figure 1	Custom Logs - Media Distribution Example	5
----------	--	---

Figure 2	Custom Logs – Fields Appended to NCSA Combined Logs Example	6
----------	--	---

1 REFERENCE MATERIALS

1.1 Document Version Control

Once delivered, please record all changes made to this document.

Date	Version	Author	Section	Amendment
2006				Beta version of document made available
10/23/07	1.0	M.Eaton		Official Version 1.0 released
12/14/07	1.01	D. Bruton		Editing and Minor Formatting

1.2 Master Documents

Document ID	Source	Title
	Unica	NetInsight User's Guide

2 Introduction

Custom logs allow Data Providers to send EMS information that was created from customized servers/log writers, offline activities or almost anything that adheres to the custom log file requirements. For example, offline orders or media distribution could be tracked in a text file, or additional information could be appended to each transaction captured in NCSA Common/Combined logs. See **Figure 1 Custom Logs - Media Distribution Example** and **Figure 2 Custom Logs – Fields Appended to NCSA Combined Logs Example**. Regardless of the type of information contained in the custom log, the EMS can process it as long as it adheres to the custom log requirements.

3 Custom Log File Format

Custom logs offer Data Providers the flexibility to supply the EMS with information not available in standard server logs or to supply the information from interfaces that do not produce logs automatically. Error! Reference source not found. identifies required, optional and custom fields. While only five fields are required for the EMS to ingest custom logs, using additional fields will add significantly to the metrics reporting capabilities. A general rule of thumb is: if you might want a report on specific information, include it as a field in the log file.

3.1 Requirements for Custom Logs

1. Only one record is allowed per line.
2. Each line represents a single transaction.
3. Fields on each line must be separated by delimiters. The preferred field delimiter is '|'; however, any delimiter may be used. A second delimiter can be added where necessary if the first delimiter is contained within a field such as [], "", or any other pair used to enclose a field.
- 4. There are no restrictions on field length.**
5. Each line has a standard format that is consistent throughout the log file. If a line does not conform to the standard, it will not be loaded into the EMS.
6. Each line must contain fields specifying date-time, request, requestor, volume, and status code (see Error! Reference source not found. below for details).
7. Data Providers must specify the log format in the Data Manifest File.

Column Name	Description	EMS Req'd	Example
Date Time	Date Time formats are flexible. Dates must contain Day, Month and Year. Time must contain hour, minute, second and day/night indicator (24-hour clock or AM/PM). A date-time format is only considered valid if it contains the following fields: Day, Month, Year, Hour, Minute, Second.	YES	<ul style="list-style-type: none"> • Date and time in Common Log Format (for example, "01/Jan/2004:01:01:01 -0500") • Date and time in ECS Firewall Log Format (for example, "Feb 1 01:59:34")
Request	The request for a file, product or service	YES	URL, GET, Page, Media Product Name, Special Service, Filename (including or excluding path)
Requestor	The person, machine or service making a request. An additional username field is also provided. If this is a physical distribution media log, the name of the requestor is sufficient.	YES	IP Address, Hostname, Username
Volume	Volume in bytes	YES	54123

Column Name	Description	EMS Req'd	Example
Status Code	FTP or HTTP status or custom codes	YES	<ul style="list-style-type: none"> • HTTP example: 200=successful) • FTP example: c=complete) • If custom codes are used, definitions must be identified in the Operations Agreement (OA).
Time Zone	Time zone in which the server resides, in the following format (+/-0000)	Recommended	-0500 ... +0100
Query String	URL-encoded query string	Optional	data.cgi?BBOX=-100.0,10.0,-70.0,35.0&TIME=2004-12-31/2004-12-31
Useragent	User browser and machine configuration information	Optional	(compatible; MSIE 6.0; Windows NT 5.0; .NET CLR 1.1.4322)
Username	Username that can be linked to individuals or processes	Optional	Guest User
Data Provider Defined Fields	Data Providers can include an unlimited number of additional fields but need to specify these fields in the appropriate location in the Data Manifest File. See section 4.2.2.1.1 for details.	Optional	Affiliation, State, Media Type, Country, User Type, Product, Mission, Instrument
timestamp OrZone	char(2) NOT NULL	YES	
requestDateTime	VARCHAR2(125)	YES	20-JUL-04 12.15.00.000000 PM
requestPath	VARCHAR2(3500)	YES	/AE_Land.001/2005.01.22/AMSR_E_L2_Land_B01_200501221541_A.hdf
referrer	VARCHAR2(3500)	NO	UNKNOWN
userAgent	VARCHAR2(3500)	NO	UNKNOWN
cookie	VARCHAR2(3500)	NO	UNKNOWN
volumeBytes	VARCHAR2(60)	YES	9570384, 0
rawVolumeBytes	VARCHAR2(60)	NO	0
subsetVolumeBytes	VARCHAR2(60)	NO	0
ipHost	VARCHAR2(100)	YES	163.com, nsidc.org
ipServer	VARCHAR2(100)	NO	
tranStatus	VARCHAR2(100)	YES	canceled, skipped
tranRate	VARCHAR2(60)	NO	
transferDuration	VARCHAR2(100)	NO	
childProcesses	VARCHAR2(100)	NO	

Column Name	Description	EMS Req'd	Example
accessMode	VARCHAR2(100)	NO	
accessDirection	VARCHAR2(50)	NO	
specialActionFlag	VARCHAR2(50)	NO	
product	VARCHAR2(255)	NO	MOD10_L2, AE_RAIN
mission	VARCHAR2(100)	NO	TERRA, AQUA
instrument	VARCHAR2(100)	NO	AMSR-E, MODIS
email	VARCHAR2(255)	YES	ops@nsidc.org
userName	VARCHAR2(100)	YES	NSIDC5954, cmops
userType	VARCHAR2(15)	YES	5, 4
userAffiliation	VARCHAR2(100)	YES	UNIVERSITY, NON-PROFIT, Anonymous
userCompanyAgency	VARCHAR2(150)	NO	
userSearchTerm	VARCHAR2(500)	NO	
orderID	VARCHAR2(100)	NO	
numberOfOrders	VARCHAR2(100)	NO	
serviceType	VARCHAR2(100)	NO	
serviceMode	VARCHAR2(100)	NO	
country	VARCHAR2(75)	NO	

Table 1 Custom Log File Format

Log Line Format:
Requestor|&| [date|&|time]|&|product_id|&|request|&|bytes|&|media_type

Sample Log Line:
Joe.Smith|&| [04/Oct/2004:10:23:13 -0800] |&|107|&|File.hdf|&|12000|&|DVD

Note: The primary delimiter for this line is the EMS preferred delimiter |&; however, because the [] surrounding Date Time are included, this is considered a single field because of the second delimiter.

Figure 1 Custom Logs - Media Distribution Example

Log Line Format:

*host, rfc931, username, [date::time], "request", statuscode, bytes,"referrer",
user_agent", 'country', 'page_server'*

Sample Log Line:

*69-167-32-159.stmnca.adelphia.net - - [06/Feb/2006:00:01:25 -0500] "GET
http%3A//nsidc.org/sotc/glossary.html&rf=http%3A//nsidc.org/sotc/sea_level.html
&rs=1280x1024&cd=32&ln=en&tz=GMT%20-08%3A00&jv=1 HTTP/1.1" 200 85
"- "Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR
1.1.4322)" 'United States' 'nsidc.org'*

Note: In this example, white spaces are the primary delimiter. However, many fields also use "" or " as secondary delimiters. Also, certain additional fields have been added to the NCSA Combined format which makes this a custom log format.

Figure 2 Custom Logs – Fields Appended to NCSA Combined Logs Example

4 Custom Log Files Naming Convention

Filenames are composed of three parts:

Timestamp	Designates the year, month and day the content of the data file was created. If a revised file is being sent (see below), the timestamp represents the date on which the original file was created with the .rev<1-n> file extension used to identify the file as a revision.
Root File Name	Aggregation of the Provider, File Type, and Data Source components of a file name. The Root File Name must be unique for each provider.
Extension	Designates the type of file and the revision status

The name of the log file must be in the following format for all Data Providers:

$\underbrace{\langle\text{YYYYMMDD}\rangle}_{\text{Timestamp}} \text{_} \underbrace{\langle\text{Provider}\rangle \text{_} \langle\text{FileType}\rangle \text{_} \langle\text{DataSource}\rangle}_{\text{Root File Name}} \text{.flt.rev}\langle\text{1-n}\rangle$

where:

YYYY	Designates the four-digit year																								
MM	Designates the two-digit month, 01 through 12																								
DD_	Designates the two-digit day, 01 through 31, followed by an underscore																								
Provider_	Designates the provider of the data, using a mutually agreed-upon acronym defined in the Operations Agreement (OA)																								
FileType_	Designates the type of file sent, followed by an underscore, where Type is:																								
	<table> <tr> <td>“DistFTP”</td> <td>FTP distribution log</td> </tr> <tr> <td>“DistHTTP”</td> <td>HTTP distribution log</td> </tr> <tr> <td>“DistMedia”</td> <td>Media distribution log</td> </tr> <tr> <td>“DistSCP”</td> <td>SCP log</td> </tr> <tr> <td>“DistSpcSrv”</td> <td>Special services log</td> </tr> <tr> <td>“DistFireWall”</td> <td>Firewall log</td> </tr> <tr> <td>“DistCustom”</td> <td>Custom log</td> </tr> <tr> <td>“DistFTPSUBSCRIPT”</td> <td>Subscriptions accessed by FTP</td> </tr> <tr> <td>“DistSCPSUBSCRIPT”</td> <td>Subscriptions accessed by SFTP/SCP</td> </tr> <tr> <td>“ECHOsearches”</td> <td>Custom for ECHO search results</td> </tr> <tr> <td>“ECHOOrders”</td> <td>Custom for ECHO orders</td> </tr> <tr> <td>“ECHOCustom”</td> <td>Future use</td> </tr> </table>	“DistFTP”	FTP distribution log	“DistHTTP”	HTTP distribution log	“DistMedia”	Media distribution log	“DistSCP”	SCP log	“DistSpcSrv”	Special services log	“DistFireWall”	Firewall log	“DistCustom”	Custom log	“DistFTPSUBSCRIPT”	Subscriptions accessed by FTP	“DistSCPSUBSCRIPT”	Subscriptions accessed by SFTP/SCP	“ECHOsearches”	Custom for ECHO search results	“ECHOOrders”	Custom for ECHO orders	“ECHOCustom”	Future use
“DistFTP”	FTP distribution log																								
“DistHTTP”	HTTP distribution log																								
“DistMedia”	Media distribution log																								
“DistSCP”	SCP log																								
“DistSpcSrv”	Special services log																								
“DistFireWall”	Firewall log																								
“DistCustom”	Custom log																								
“DistFTPSUBSCRIPT”	Subscriptions accessed by FTP																								
“DistSCPSUBSCRIPT”	Subscriptions accessed by SFTP/SCP																								
“ECHOsearches”	Custom for ECHO search results																								
“ECHOOrders”	Custom for ECHO orders																								
“ECHOCustom”	Future use																								

DataSource	Designates the source or type of system generating the log file, where Type is:	
	“ECSDataPool”	Data Pool log
	“ECS”	Specifies log is from an ECS system other than Data Pool.
	<Custom>	Data Providers may populate with information needed to uniquely identify the source of a log file, including virtual server names, subsystem names or other unique information.
.flt	Indicates that the file is a plain text file	
.rev<1-n>	Indicates that the file has been resent because of errors, with the number incremented for each update starting with .rev1 (for example, rev1, rev2, rev3... revN)	

The EMS data files are automatically or manually created by Data Providers and then securely transmitted to the EMS. All data files should be delivered as specified in Table 3.2-1 of the ICD. Timing of the delivery will be specified in the Data File Manifest File. The Data File Manifest File, as presented in **Table 2 Data File Manifest File Format**, is used by Data Providers and the EMS Team to identify all data sources delivered to the EMS. If the EMS does not receive expected files, as outlined in the Data File Manifest File, or encounters errors during processing of received files, an email will be sent to the appropriate Data Provider identifying missing files and/or processing errors.

5 Data File Manifest File Format

The Data Providers must complete a Data File Manifest File, a worksheet that identifies all data files being delivered to the EMS. A template is available for download from <http://ems.eos.nasa.gov/templates/DataFileManifestTemplate.xls>. Each column must represent a single data file and follow the format described in the table below, where 'Y' indicates mandatory data and 'A' indicates that the information is required if available.

Category	Description (examples only)	Req'd	Data File 1	Data File 2	Data File N
Provider	Name of the Data Provider (for example, GES, GSFCV0, GSFC S4PA, EDC, LATIS, MODAPS). Mutually agreed-upon list is in the OA.	Y			
Provider Category	Classification of the Data Provider (for example, SIPS, ECS, NON-ECS). Mutually agreed-upon list is in the OA.	Y			
Reference URL	The base address/url of where the data file resides (for example, http://podaac.jpl.nasa.gov). If none exists, enter NA.	Y			
Source IP	The IP address of the host(s) that sends flat and log files to EMS. This information is required by EMS to allow sftp/scp/rsync transfers. Each data file should have an entry (even if the source IP remains the same) to ensure firewall rules allow access to the EMS.	Y			
Root File Name	The root name of the file, using the EMS ICD file naming convention. The date does not need to be included; only the root of the file name is required (for example, podaac_log_http.ftl).	Y			
Data File Format	The format of the data file and whether it is standard or custom. If standard, then specify the format type. See Figure 4.2.3-1 in the ICD– for example, Cisco PIX, Gauntlet, IBS Firewall, IIS Standard/Extended, Microsoft Proxy).	Y			
Push Frequency	How often the files are sent to the EMS. The minimum push frequency is once per day for all data files.	Y			
Push Time	The time of push events based on Greenwich Mean Time (GMT).	Y			
Comments	Comments that the Data Provider may provide to assist the EMS staff in understanding the data and the process	A			
Data File Format Details	If the data file format is custom, refer to <i>The EMS Custom Log Implementation Guide</i> at http://ems.eos.nasa.gov/documents/TheEMSCustomLogImplementationGuide.doc for the requirements.	Y (custom only)			

Table 2 Data File Manifest File Format

5.1 Data File Manifest File Naming Convention

The name of the Data File Manifest File must be in the following format:

<YYYYMMDD>_<Provider>_ DataFileManifest.xls

where:

YYYY	Designates the four-digit year for the time the Data File Manifest File was created
MM	Designates the two-digit month, 01 through 12
DD_	Designates the two-digit day, 01 through 31, followed by an underscore
Provider_	Designates the provider of the data, using a mutually agreed-upon acronym defined in the OA, followed by an underscore
DataFileManifest.xls	Indicates that the file is the Data File Manifest worksheet

5.2 Data File Manifest File Updates

Each time a Data Provider adds or removes a data file or makes any changes that impact fields within the Data File Manifest, a revised Data File Manifest must be sent to the EMS. The revised Data File Manifest must contain all valid entries for data files that will be sent to the EMS. If the Data File Manifest contains invalid entries, error notification emails will be sent.

5.3 Data Files Interface Mechanism

Data Providers are responsible for pushing data files to the EMS data server located at ftp://ws1.ems.eosdis.nasa.gov. Each file that a Data Provider delivers to the EMS should contain one day's worth of data. A slight overlap from one day to the next during day boundaries is acceptable.

The EMS Team will provide each Data Provider with an account for transferring data to the EMS. Data Providers will be able to access their EMS accounts via password-free, key-based authentication. Access to the EMS account will be restricted to a Data Provider-defined IP address as specified in the OA. The preferred method for Data Providers to transfer data files to the EMS is via an rsync client tunneled through ssh. Alternatively, Data Providers can use sftp if the rsync client is not available. The method that a Data Provider will use to transfer data to EMS will be identified in the OA.

In the event a Data Provider needs to resend a data file because of corruption, errors or any other reason, the original filename for that day's worth of data must be reused with the addition of the suffix .rev1 to show it is an updated (revised) file. If additional updates are necessary for the same day's worth of data, then .rev1 will be incremented for each update (that is, .rev1, .rev2, .rev3... .revN). Thus, if a Data Provider creates a flat file describing ingest on May 14, 2007, the

filename format is 20070514_<Provider>_<FileType>_<DataSource>.flt. If the Data Provider later determines the contents of this file to be incorrect, the Data Provider will resend the ingest data for that day but change the filename to 20070514_<Provider>_<FileType>_<DataSource>.flt.rev1.