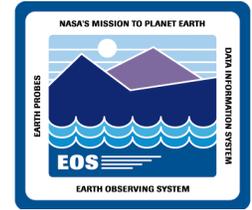


# **ECS Bulk Metadata Generation Tool (BMGT)**

**Guy Swope**  
**[gswope@eos.hitc.com](mailto:gswope@eos.hitc.com)**  
**June 7, 2002**



# BMGT Presentation Overview

---

**Purpose**

**Upcoming Changes**

**ESDT and Product Specification**

**Operations Concept**

**Production Rules**

**Operations Impacts**

**BMGT Setup**

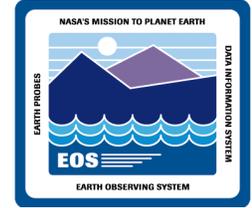
**Access to Metadata Products**

**Sizing Estimates**

**Issues**

**References and Documentation**

# Purpose of the BMGT



**Provide for an external representation of ECS metadata holdings**

- Collection, granule, browse and valids

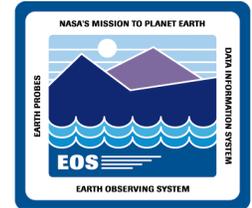
**Metadata obtainable via normal ECS distribution methods**

**Offer a capability for bulk distribution of browse data**

**Support for Value Added Providers**

- IIMS (ECHO), ESIPs, RESACs, and InfoMarts

# Upcoming Changes



## Generation of ECSBBR granule changed to limit size

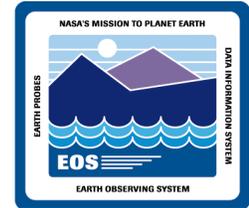
- Number of records
- ESDT Group
  - File naming convention of ECSBBR granules to match ECSMETG and ECSMETC convention

## Performance Enhancements

### Synergy III Changes

- Additional performance enhancements
- Export of Data Pool online URLs to support direct access to data stored in Data Pools
- Support for export of AIRS browse

## Looking at alternative methods for historical data population



# ESDT and Product Specification

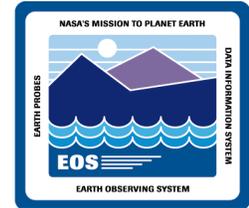
---

Four new ESDTs were created

Each stores an external representation of the ECS metadata

Format of each product controlled by an XML DTD

- **ECSMETC** – Stores products that contain an XML representation of ECS collection level metadata and the packaging options that may be used when ordering products from each collection
- **ECSMETG** – Stores products that contain an XML representation of ECS granule level metadata
- **ECSBBR** - Stores products that contain an XML representation of references to browse images
- **ECSMETV** – Stores products that contain an XML representation of ECS collection and granule valid values



# ESDT Groups

**ECSMETC and ECSMETG collections store products containing metadata for multiple collections and multiple granules**

**The metadata will be grouped by instrument and mission**

- **Exception is MODIS ESDTs, which are grouped by mission and major discipline**
  - **Ocean, atmosphere, land, and snow & ice**

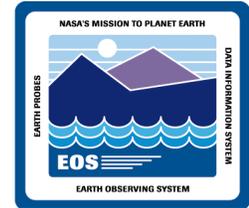
**Mapping between group and ESDTs maintained**

- **Being added to the ECS ESDT baseline found at the <http://ecsinfo.gsfc.nasa.gov/> website**
- **Mapping used by the BMGT in its Groups configuration file**
- **Same mapping used in the Data Pools**

**ECSMETC and ECSMETG granules have a group identifier Product Specific Attribute (PSA) called GroupId**

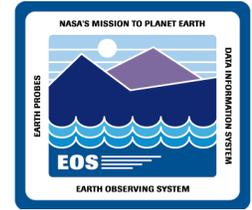
- **Allows for end users to search and find all metadata for a specific group of collections**

# ESDT Groups - Example



| <b>GroupId</b> | <b>Description</b>   | <b>DAAC Storing Metadata</b> |
|----------------|--|------------------------------|
| MOSA           | MODIS Snow and Ice collections/granules from the Aqua mission  | NDAAC                        |
| MOOT           | MODIS Oceans collections/granules from the Terra mission   | GDAAC                        |
| ASTT           | ASTER collections/granules from the Terra mission  | EDAAC                        |
| MSRT           | MISR collections/granules from the Terra mission   | LDAAC                        |
| MOPT           | MOPITT collections/granules from the Terra mission   | LDAAC                        |
| LSR7           | All collections/granules from the Landsat 7 mission  | EDAAC                        |
| SAG3           | All collections/granules from the SAGE III mission   | LDAAC                        |
| ACRM           | All collections/granules from the ACRIM mission  | LDAAC                        |
| OTHR           | Other Products. This group is used for ancillary products or other non-science products (e.g., ECSMETC, ECSMETG, ECSBBR, ECSMETV). | GDAAC, EDAAC, LDAAC, NDAAC   |

# Operations Concept



**Four new ECS data collections will be created at each site to store ECS metadata**

**The Bulk Metadata Generation Tool (BMGT) ran daily to populate these data collections**

**ECS ESDTs will be grouped according to discipline/instrument**

**One metadata product will be created per ESDT group per day**

**Each product will contain an external representation of the metadata**

- **For each collection and/or granule that was inserted, updated or deleted that day**

**One bulk browse product will be produced per day**

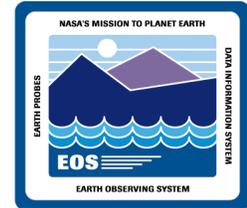
- **Contains references to each browse that was inserted, updated or deleted that day**

**The format used for the external representation of the metadata is XML**

**End users may use any of the standard ECS mechanisms to find and order this data**

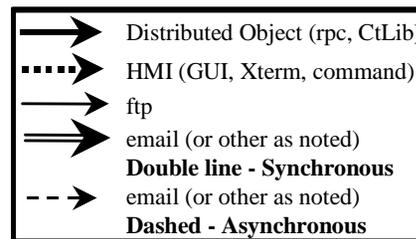
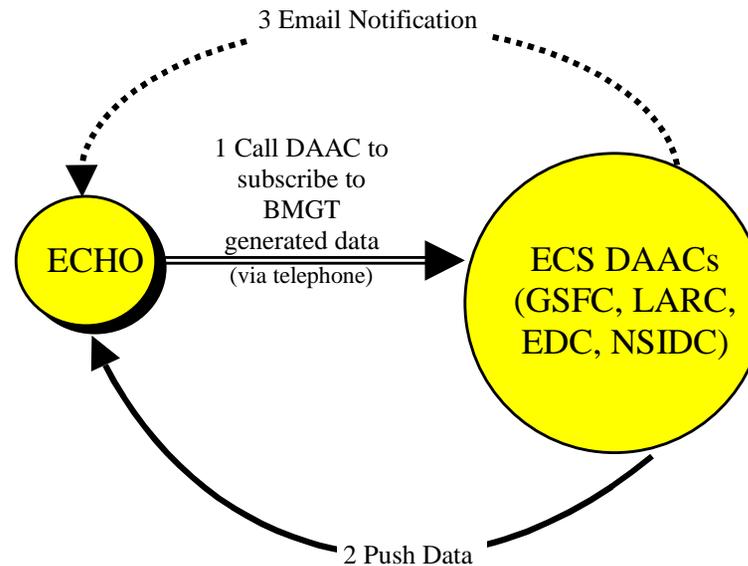
- **Mechanisms include EDG search, order, and subscription capabilities**

# Operations Concept - External User Perspective

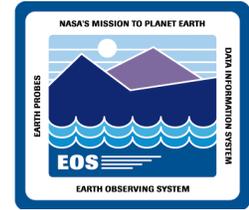


Method to obtain BMGT generated data is relatively simple

Uses standard ECS distribution mechanisms

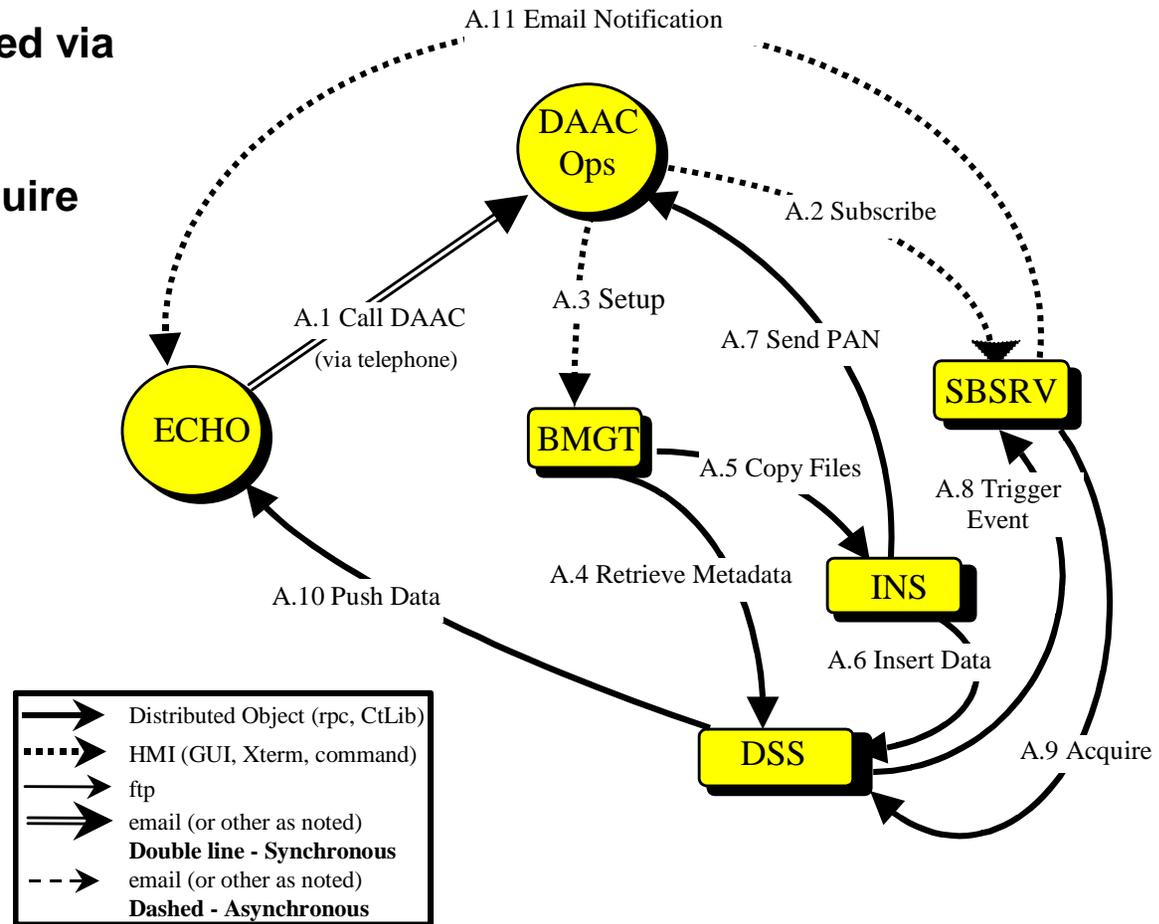


# Operations Concept - ECS Perspective



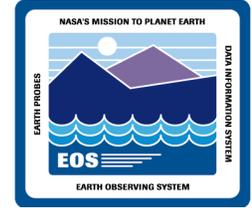
BMGT data inserts handled via SIPS interface

Subscription fires an acquire (distribution) request



# Bulk Browse Operations Concept

---



**Browse XML product contains references to browse images**

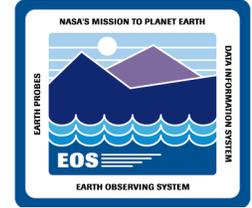
- Aptly named the Browse Reference File (BRF)

**Special code (DLL) adds the referenced browse images to the distribution**

- As if they were files in a multi-file granule
- No browse metadata files are created

**Browse images are distributed in the same manner as specified in the distribution request**

# Collection Metadata Production Rules



**For each target ESDT version that had collection level metadata inserted or updated, perform the following steps:**

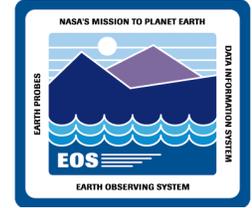
- 1. If this collection is the first collection in a group...**
  - Create a new file and append an XML representation of the packaging options to the file
  - Append an XML representation of the collection level metadata to the file.
- 2. Insert each file, as a product, into the ECSMETC data collection**
  - Set the value for the GroupId PSA
  - Set the starting date and ending date of the insert, update, and/or delete activity covered by this file.

**Extraction based primarily on two things**

- InsertTime and LastUpdate fields in main collection table in the ECS Science Data Server (SDSRV) database fall within generation time period
- ESDT is included in the BMGT Group's configuration file

**Assumes that no collections will be deleted**

# Granule Metadata Production Rules



For each target ESDT version that had granule level metadata inserted, update, and/or deleted, perform the following steps:

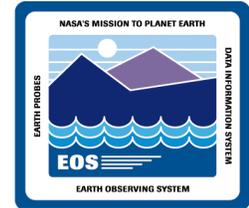
1. If this granule is the first granule in a group...
  - Create a new file and append an XML representation of the metadata for each active granule to the file
2. Insert each file, as a product, into the ECSMETG data collection
  - Set the value for the GroupId PSA
  - Set the starting date and ending date of the insert, update, and/or delete activity covered by this file

Extraction based primarily on two things

- InsertTime, LastUpdate and DeleteTime fields in main granule table in the ECS Science Data Server (SDSRV) database fall within generation time period
- ESDT is included in the BMGT Group's configuration file

For deleted granules, only the granule identifier (granuleUR) and delete time will be included in the file

# Browse Metadata Production Rules

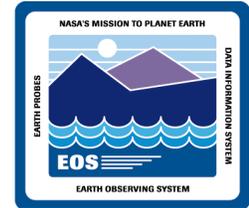


**For all browse images that were inserted or deleted within a specified time period**

- 1. Extract the browse identifiers and associated browse file names for each browse product**
- 2. Place the reference information into the browse XML file**
- 3. Insert the XML file, called the Browse Reference File (BRF) file, as a product into the ECSBBR data collection**
  - Set the value for start and end date of the insert, and/or delete activity covered by this file**

**Extraction based on InsertTime, deleteTime fields in main browse table in the ECS Science Data Server (SDSRV) database**

# Validates Metadata Production Rules



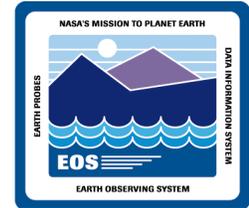
**If any collections were inserted, updated, and/or deleted during the period then**

- 1. Create a new file, append an XML representation of the validates information**
- 2. Insert the file, as a product, into the ECSMETV data collection**
  - Set the starting date and ending date of the insert, update, and/or delete activity covered by this file**

**Validates generation is dependent upon ESDT inserts or updates**

- If collection metadata is generated (i.e., ECSMETC products), validates metadata is generated (i.e., ECSMETV)**

# DAAC Operations Impact - Summary



## BMGT Installation

- Performed on the “acg” SGI hosts at each DAAC
  - Same hardware that hosts the ECS Science Data Server’s Sybase SQL servers
  - Chosen for performance reasons – 8 CPUs, same physical location as database

## Creation of subscriptions

- Part of typical DAAC activities
- Initially setup 4 subscriptions (for inserts of ECSMETC, ECSMETG, ECSMETV, ECSBBR data) for ECHO

## BMGT daily cron setup

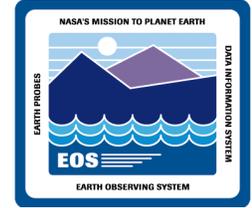
- Configuration options need to be set – probably just once
- Groups configuration file defines which ESDTs are exported needs to be modified periodically

## BMGT manual run to capture existing (historical) metadata

- Configuration options need to be set – probably just once
- Most manual activity in running the BMGT
  - Need to constantly modify the time period to generate metadata
    - week 1, 2000; week 2 2000 etc.
- Once done shouldn’t need to be done again

# Installation of BMGT Prerequisite Components

---

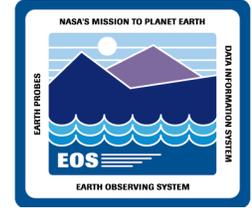


## DAAC personnel install new Earth Science Data Types (ESDTs)

- ECSMETC, ECSMETG, ECSBBR and ECSMETV
- Each stores an external representation of the ECS Science Data Server's (SDSRV) inventory database

## Associated Ingest and Science Data Server Subsystem changes should already be installed automatically

- Ingest database changes to support new SIPS data provider (i.e., BMGT)
- Science Data Server changes to distribute browse images during acquire of ECSBBR ESDT granules



# BMGT Installation & Setup

**Installation is performed as typical ECS subsystem installation**

- Via ECS Assistant Tool
- BMGT installable in one package – under the “OSS” (Operational Support Software) subsystem

**DAAC personnel install the ECS Bulk Metadata Generator Tool (BMGT) in OPS mode (or TS1 or TS2)**

**DAAC personnel configure the target ESDT versions to be exported**

- Done by editing the BMGT’s group configuration file
- Set of ESDTs should be the same set of ESDTs currently exported to EDG

**DAAC personnel configure packaging (distribution) options**

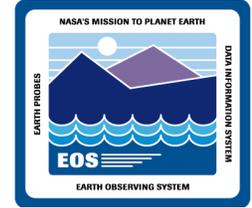
- Done by editing the BMGT’s general package file
- Set of distribution options should be the same used for the ECS V0 Gateway

**DAAC personnel configure contact information**

- Done by editing the BMGT’s contact info file
- DAAC contact information should be the same used for the ECS V0 Gateway
  - Typically DAAC User Services information

**“BMGT” SIPS interface needs to be setup to ingest BMGT generated data**

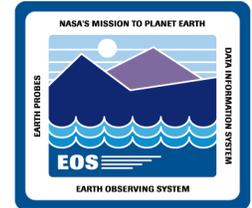
# Subscription Setup



## Subscriptions are setup as typical ECS subscriptions

- First set of subscriptions setup for ECHO interface testing
- Each DAAC will setup 4 subscriptions against the new BMGT-related ESDTs
- Subscription will be setup for FTP Push to the ECHO consumer
  - Parameters required for FTP Push (host, login, directory) obtained from ECHO team

# BMGT Standard Setup For Daily Ingest



DAAC personnel configure the BMGT to run daily as a cron job

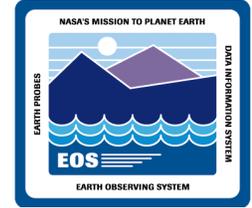
Recommended setup to generate bulk metadata products for previous day

- Typical generation would include
  - Each specified ESDT version that had granules inserted, updated or deleted
  - Associated browse reference file

Detailed description of full set of configuration options can be found in ECS document 170-WP-023-003

- *Bulk Metadata and Browse Export Capability for the ECS Project*

# BMGT Setup For Generation for Existing Data



**After the BMGT has been setup for normal operation**

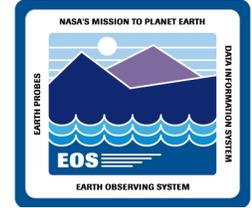
**DAAC personnel setup BMGT to generate products for existing (historical) metadata**

**DAAC personnel manually invoke BMGT**

- **Date range is defined by editing the BMGT's Parameters file**
- **Date range should not be too large**
  - **Because DAACs will have over a year's worth of operational data, all existing granules should not be processed at once**
- **DAACs will need to estimate best time range based on inventory**
  - **Limited by BMGT performance**

**Generation of existing metadata repeated by DAAC personnel**

- **Performed manually since date range needs to be changed each time**
- **Until all existing granules in the SDSRV inventory database have been converted**

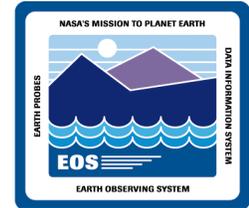


# Access to Metadata Products

Accessed using any of the standard ECS search and order methods

- 1. Contact DAAC User Services and request an FTP Push subscription**
  - Placed on the ECSMETC, ECSMETG, ECSBBR and ECSMETV data collections
  - Enable metadata products to be pushed as they are generated
  - A subset of the metadata can be obtained by qualifying the subscriptions via GroupIds
  - Subscription interface and distribution notifications can be found in ECS document number 423-41-57
    - Interface Control Document Between the EOSDIS Core System (ECS) and the Science Investigator-Led Processing System (SIPS), Volume 0, Interface Mechanisms, Section 4.3
- 2. Use EOS Data Gateway (EDG) web interface to search and order**
  - ECSMETC, ECSMETG, ECSBBR and ECSMETV products
  - Details on how to use the EDG may be found at <http://edcimswww.cr.usgs.gov/pub/imswelcome/>

# Access to Metadata Products cont'd



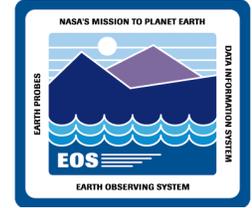
## 3. Use the ECS V0-to-ECS Gateway to search and order

- ECSMETC, ECSMETG, ECSBBR and ECSMETV products using the V0 Protocol.
- Interface specification for the V0-to-ECS Gateway found in ESDIS document number 505-41-30
  - ICD Between ECS and Version-0 System for Interoperability

## 4. Use the ECS Machine-to-Machine Gateway

- Effective with ECS Release 6A
- Search and order ECSMETC, ECSMETG, ECSBBR and ECSMETV products
- Interface specification for the ECS Machine-to-Machine Gateway can be found in ECS document number 209-CD-035-001
  - Interface Control Document for the Machine-to-Machine Search and Order Gateway for the ECS Project

# Sizing Estimates



**Products produced per day will depend on the frequency of inserts, updates, and deletes**

**ECSMETC products produced depend on the frequency of ESDT inserts, updates**

- Expected to happen infrequently

**ECSMETV products depend on ECSMETC generation**

- Expected to happen infrequently

**Most daily BMGT runs will not produce ECSMETC or ECSMETV products**

**ECSBBR products depend on the frequency of browse file inserts and deletes**

**ECSMETG products depend upon the frequency of granule inserts, updates and deletes**

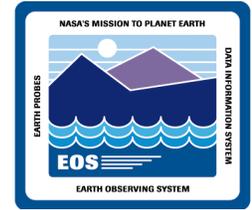
- Number of ECSMETG products that are generated per day will drive metadata volume

**Example: GSFC MODIS ECSMETG products would produce 3.65MB XML**

- Based on June 7, 2001 GSFC MODIS ingest estimates of 1300 granules

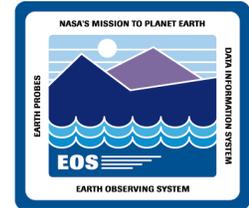
# Known Issues

---



## Granule Deletion Issues

1. Time delay between when a granule is deleted in ECS versus
  - Opportunity exists for an end user's order to be rejected
2. Some manual deletes currently done at DAACs will not register in XML product
  - DAACs need to set deleteTime field in ECS Science Data Service database to date granule was deleted



# References and Documentation

## Bulk Metadata Generation Tool

- Most information included in this presentation is in ECS document 170-WP-023-002
- *Bulk Metadata and Browse Export Capability for the ECS Project*
- Current posting on ECS Data Handling System (EDHS), <http://edhs1.gsfc.nasa.gov/> is out of date, expect update in next 2 weeks

Subscription interface and distribution notifications can be found in ECS document number 423-41-57

- *Interface Control Document Between the EOSDIS Core System (ECS) and the Science Investigator-Led Processing System (SIPS), Volume 0, Interface Mechanisms, Section 4.3*

Interface specification for the ECS Machine-to-Machine Gateway can be found via EDHS in ECS document number 209-CD-035-001

- *Interface Control Document for the Machine-to-Machine Search and Order Gateway for the ECS Project*

Interface specification for the V0-to-ECS Gateway found via EDHS in ESDIS document number 505-41-30

- *ICD Between ECS and Version-0 System for Interoperability*