

NSIDC Special Report 12

An Introduction to EOS Data in the EOSDIS Core System (ECS) at NSIDC



Artist rendition of Terra satellite.
Image courtesy of NASA.

→

A screenshot of the ECS Order Manager GUI. The interface shows a table with columns for Order ID, Order Date, Order Status, Order Type, Order Amount, and Order Date. The table contains several rows of data, including order numbers and dates. The table is titled "ECS Order Manager" and has a "Distribution Requests" tab selected. The table has the following columns: Order ID, Order Date, Order Status, Order Type, Order Amount, and Order Date. The table contains several rows of data, including order numbers and dates.

An image of the ECS Order Manager GUI.

→



End users viewing and analyzing the data.

John Maurer and Amanda Leon
June 2009

**National Snow and Ice Data Center
Cooperative Institute for Research in Environmental Sciences
University of Colorado, Boulder, Colorado USA**

1. Citation

Maurer, J., and A. Leon. 2009. *An Introduction to EOS Data in the EOSDIS Core System (ECS) at NSIDC*. NSIDC Special Report 12. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

2. Introduction

The National Snow and Ice Data Center (NSIDC) and other NASA Distributed Active Archive Centers (DAACs) use ECS to ingest, archive, and distribute data. ECS stands for the **E**OSDIS **C**ore **S**ystem, and EOSDIS stands for the Earth Observing System Data and Information System. Raytheon develops and maintains this multi-million-dollar system for NASA using many of today's most advanced technologies.

The Earth Observing System (EOS) data sets that NSIDC collects come from the AMSR-E, GLAS, and MODIS instruments. NSIDC also collects many data sets that are not ingested, archived, or distributed through ECS. Examples include data from the AVHRR, SMMR, and SSM/I-SSMIS instruments. Procedures for handling non-ECS data sets are not covered in this document.

ECS is composed of multiple **servers** and commercial off-the-shelf (COTS) software products that run on several different machines networked together. The servers are individual C++ programs that handle specific tasks.

These machines, servers, and COTS work together to accomplish three main tasks related to ECS operations: **ingest**, **archive**, and **distribution**. **Ingest** refers to the acquisition of data from our external data providers. **Archive** refers to the transfer of ingested data onto a permanent storage device. **Distribution** refers to the transfer of archived data to users who request them. Each of these tasks is explained in greater detail below.

For a great introductory reference to NASA's Earth Science Enterprise, which includes EOSDIS and the DAACs, visit the NASA Earth Observatory: Overview of the Earth Science Enterprise (ESE) Web page .

Before moving any further, let us first take a quick look at what types of data are archived via ECS.

3. Principles of remote sensing

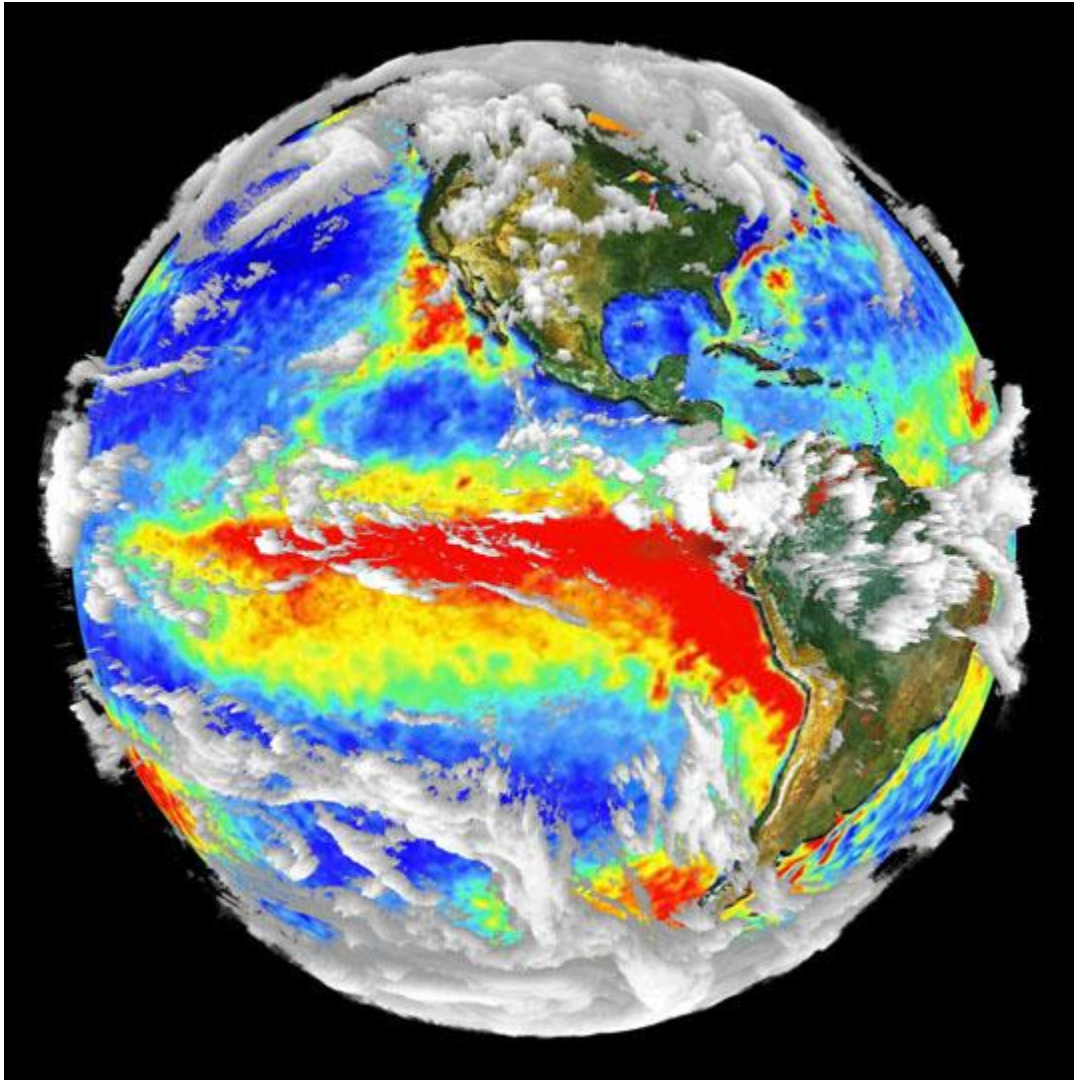


Figure 1. Global data from multiple sensors on Terra. Image courtesy of R.B. Husar, Washington University.

ECS ingests, archives, and distributes satellite **remote sensing** data products. Remote sensing involves obtaining information about an object without actually coming into contact with it. Photographs are a good example of remote sensing. Sensors obtain information not only in the visible portion of light, as with cameras, but they can also measure other portions of the **electromagnetic spectrum** (e.g., ultraviolet, infrared, and microwave radiation). These sensors may be housed in instruments used on the ground (in situ), on aircraft, or on satellites.

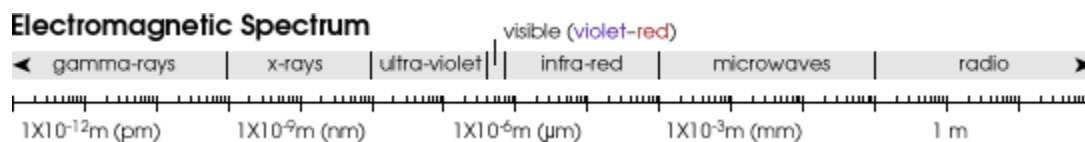


Figure 2. Image courtesy of NASA's Earth Observatory Web site.

Satellite instruments normally measure radiation at discrete wavelengths of the electromagnetic spectrum called **bands** or **channels**. An everyday camera, for comparison, measures all of the

light within the entire visible spectrum (400 nm to 700 nm wavelength) in a single band. The MODIS instrument, on the other hand, measures electromagnetic radiation at 36 individual bands between 400 nm and 14,500 nm, which spans from visible light to thermal infrared radiation. These bands range in width from 10 to 500 nm. Scientists use these bands to quantitatively assess properties of the Earth's land, oceans, and atmosphere that contribute to weather prediction, monitoring of natural disasters, global climate change assessment, and beyond.

Just as a camera views a certain amount of space through its lens (for example, you may have to back up in order to fit a person entirely in the field of view), satellite remote sensing instruments also have a limited field of view. A single MODIS data file, for example, covers a width of 2,300 km. By comparison, the Earth's diameter is 12,756 km. This gives MODIS the capability to view every part of the planet every 1-2 days.

There is also a limit to how small an area that a particular sensor views. For example, if you take a photograph of a person 1 km away, you cannot see the logo on his or her shirt, or even the color of his or her eyes. Similarly, remote sensing instruments have a specific **resolution**, which is the measure of the smallest object that one can "resolve," or view. In terms of a remotely sensed image, this resolution is also often referred to as its **pixel size**. Resolution and pixel size thus describe the smallest area on the Earth's surface that a remote sensing instrument can view. MODIS views objects as small as 250 m in certain bands, while other MODIS bands have pixel sizes of 500 m or 1 km. For comparison, the ASTER satellite remote sensing instrument (which NSIDC does *not* collect data from) can resolve objects as small as 15 m. Resolution of an instrument depends on many factors, including its altitude in space, the wavelength that it is measuring, its method of collecting data, and its design.

Lastly, remote sensing data are often viewed as digital images, which involve the same concepts as computer screens or televisions, where three photon beams—red, green, and blue—create all of the colors that we see. Any color can be generated by adding different relative amounts of these three primary colors (referred to as "RGB," for red, green, and blue).

The human eye cannot see beyond the visible portion of the electromagnetic spectrum. However, any combination of bands that measure radiance in ultraviolet, infrared, or microwave wavelengths can be assigned to the RGB bands to produce a color image. These images are referred to as **false-color composites** (see Figure 3) since they combine bands that are not in the visible portion of the spectrum.

True-color composites (see Figure 4) are created by displaying three bands that measure light in the red, green, and blue portions of the electromagnetic spectrum. A color photograph is a good example.

A simple greyscale image (see Figure 5) is generated by viewing one band at a time. Bright shades of grey correspond to places where the radiation is high in a given band, while dark shades of grey correspond to places where the radiation is low.

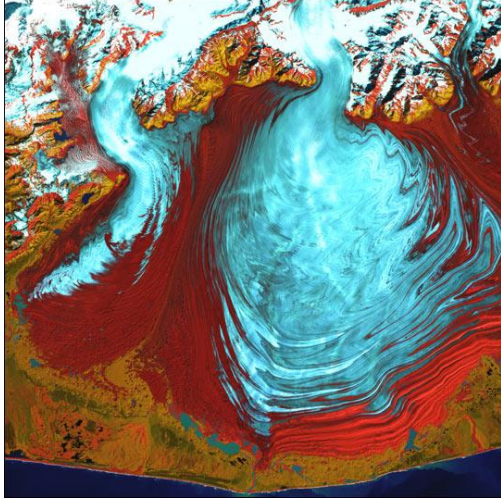


Figure 3. Malaspina Glacier, Alaska USA, 31 August 2000. Landsat ETM+ false-color composite. Image by the USGS EROS Data Center Satellite Systems Branch.

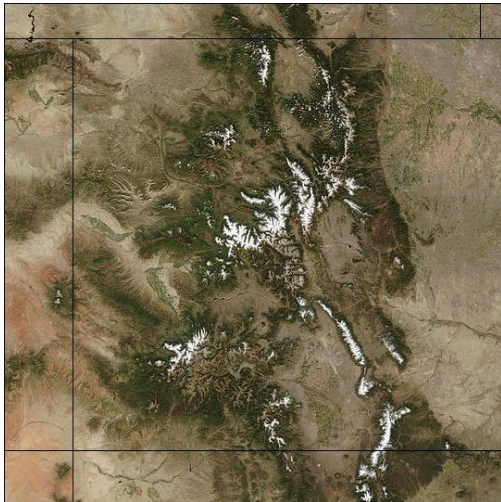


Figure 4. Colorado USA, 18 September 2002. MODIS true-color composite. Image courtesy of Jacques Desclotres, MODIS Land Rapid Response Team at NASA GSFC.



Figure 5. Larsen Ice Shelf, Antarctica, 21 February 2000. Landsat ETM+ greyscale image. Image courtesy of Ted Scambos at NSIDC.

Many data products are the result of scientific analyses of the original remotely sensed data. Snow extent and sea ice concentrations are examples. The images that result from these data usually include a legend that explains what the colors represent in the image. Figure 6 is an image processed at NSIDC that shows sea ice concentrations derived from SSM/I data over the South Pole in June 2009. The color bar on the right side of this image tells you what percentage of sea ice each color represents in the image. Figure 7 is an image showing global sea surface temperatures for 01 July 2009 derived from AMSR-E data. The color bar on the right-hand side of this image tells you what sea surface temperature (SST) in degrees Celsius each color represents in the image.

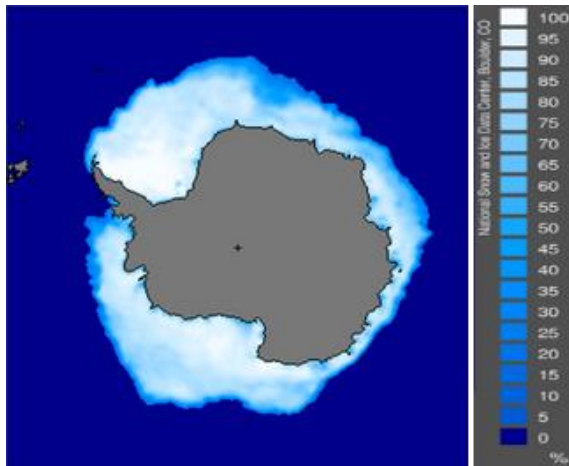


Figure 6. Monthly Antarctica sea ice concentration from SSM/I for June 2009. Image courtesy of the National Snow and Ice Data Center

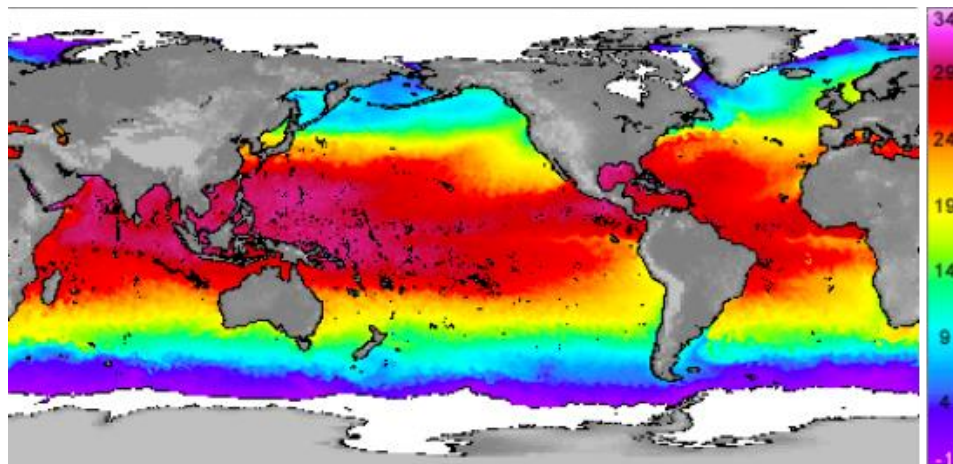


Figure 7. Global sea surface temperatures from AMSR-E. Average of three days ending on 01 July 2009.

A great place to learn about remote sensing and view images is NASA's Earth Observatory Image of the Day Web site. They post a new image and brief explanation every day. You can also view an archive of their images. View the Earth Observatory Remote Sensing Web page for a great introduction to remote sensing principles.

4. Components of ECS

Now that you have learned where remotely sensed data come from, let us discuss the three main components of ECS: ingest, archive, and distribution.

4.1 Ingest



Figure 7. Antenna at a ground-receiving station. Image courtesy of the Alaska SAR Facility (ASF).

NSIDC's EOS data come from the Aqua, ICESat, and Terra satellites. Data are transmitted from satellites to ground-receiving stations around the globe which transmit the data to a central location: the EOS Data and Operations System (EDOS) at the Goddard Space Flight Center in Greenbelt, Maryland. The raw data that EDOS collects are referred to as **Level-0** data. EDOS transmits Level-0 data via ECS to the various DAACs.

At this point, science computing facilities (SCFs) process the raw Level-0 data into products that are ultimately distributed to users. SCFs correct for various systematic errors introduced by the satellite before the raw data are distributed. These errors are corrected using precisely recorded position and attitude data from the satellite during the time of data acquisition (**ephemeris** data), or calibrated against other known measurements (**ancillary** data). After these errors are corrected and the data are referenced to time and geographic location, the data are considered **Level-1**. Time- and georeferencing involve recording the time of data acquisition and the latitude and longitude coverage in a **metadata** file to be distributed with the data. Level-1 data is the lowest level of data that is distributed to most users.

Higher-level data products are processed from the Level-1 data. **Level-2** data use scientific algorithms to calculate one or more geophysical parameters from the Level-1 data. Examples include snow cover, sea ice extent, sea surface temperature, land cover type, vegetation indices, aerosol and ozone distribution. Level-2 data gridded to a uniform map projection are called **Level-3** data. Lastly, **Level-4** data are model outputs or results from scientific analyses derived from multiple measurements of lower-level data, for example, climate change analyses. Higher levels of processing provide users with more value and information to the raw data collected by the satellite. Some users, however, do not need derived geophysical products for their intended application or may prefer to implement their own processing procedures based on their own scientific algorithms, map projection schemes, and analyses, and will therefore order the Level-1 data for these purposes.

NSIDC archives Level-0 EOS data received directly from EDOS and Level-1 through Level-3 EOS data products received from external SCFs.

NSIDC receives all levels of AMSR-E data from Aqua. The Level-0 AMSR-E data come to us from EDOS. Level-1 AMSR-E data are processed at the Japanese Aerospace Exploration Agency (JAXA) in Japan, sent to NASA's Jet Propulsion Laboratory (JPL) Physical Oceanography DAAC (PO.DAAC) in Pasadena, California USA, and then sent to NSIDC. Level-2 and Level-3 AMSR-E data products come to us from the AMSR-E Science Investigator-led Processing Systems (SIPS) at the Global Hydrology and Climate Center (GHCC) in Huntsville, Alabama USA.

We also receive all levels of GLAS data from ICESat. Again, the Level-0 data come from EDOS. Level-1, -2, and -3 data come from the ICESat SIPS at the Goddard Space Flight Center in Greenbelt, Maryland USA.

NSIDC receives MODIS Level-2 and Level-3 data from Terra and Aqua, which come from the MODIS Data Processing System (MODAPS) SCF at the Goddard Space Flight Center.

NSIDC uses ECS to ingest all of the data listed above from EDOS, the AMSR-E SIPS, the ICESat SIPS, and MODAPS via File Transfer Protocol (FTP). Each data file has an associated metadata file which stores information such as time of acquisition, size, geographic coordinates, and other information that is important for a user to know. Depending on the product, data files may also have associated low-resolution quick-look (browse) images, quality assurance files, and production history files.

Information describing the data files is stored in both Sybase databases and Extensible Markup Language (XML) files within ECS. A portion of this metadata is also sent to the the **EOS ClearingHOuse** (ECHO). ECHO is a registry of metadata describing data held in the NSIDC ECS archive as well as other DAACs and data centers.

4.2 Archive



Figure 8. Quantum Scaler i500 tape library. Image courtesy of Quantum.

The next step after ingesting data is to archive them. Archiving is the process of writing data to media for long-term storage. EOS data at NSIDC are archived in two locations. The primary storage location is a large, online Storage Area Network (SAN) disk array. The online disk archive enables immediate data distribution to users, since data do not need to be retrieved from tape. Secondary, or backup, storage is performed by an automated tape library. The SAN disk array consists of EMC CX Series Redundant Array of Inexpensive Disks (RAID) devices. The SAN has a capacity of approximately 94 terabytes and is managed using Quantum Corporation's StorNext File System (SNFS). ECS writes ingested data to the SAN SNFS online storage location as well as creates a secondary, temporary copy in a separate SNFS that acts as a staging area for the automated tape library.

The tape library used for the secondary archive is a Quantum Scaler i500 (see Figure 8), which

holds up to 220 Linear Tape-Open (LTO) tapes and eight tape drives. The library is currently configured with 200 LTO version four (LTO-4) tapes and six LTO-4 drives. LTO-4s have a data capacity of 800 gigabytes. The tape library is managed with Quantum's StorNext Storage Manager (SNSM). SNSM manages the transfer of data from the SNFS staging area to the LTOs and maintains a database of files and their location on the tapes. SNSM manages data archival through the configuration of a Policy Class, which is defined to store specific types of data and to correspond to directories on the SNFS. Each Policy Class is assigned its own tape or tapes.

4.3 Distribution



Figure 9. Image courtesy of Mack Trucks, Inc.

Despite the above picture, NSIDC does not actually use Mack trucks to distribute ECS data! ECS can distribute data electronically via either **FTP pull** or **FTP push**. FTP pull is when the data are staged locally to a machine at NSIDC, and the user initiates an FTP session to download (or "pull") the data to his own computer. An FTP push is when the data are automatically transferred (or "pushed") to a user-specified computer and directory path. The FTP push method requires that the user own a computer with a dedicated internet IP address or a host name where ECS can push the data. This option is not typical for most home personal computers. In the case of an FTP pull request, ECS sends the user an automatic e-mail specifying the information needed to login to an NSIDC server and collect the data. All ECS data from NSIDC are currently distributed free of charge.

Users initiate orders for ECS data from NSIDC through one of the following means, which are described in the following paragraphs:

- Warehouse Inventory Search Tool (WIST)
- Search 'N Order Web Interface (SNOWI)
- Subscriptions
- Data Pool

Warehouse Inventory Search Tool (WIST) is a search-and-order Web site which searches the metadata in ECHO. The WIST allows you to search for all publicly available ECS data by data type, location, time period, and various parameters within the data type(s) selected (for example, percent cloud cover). Low-resolution **quick-look** (browse) images are available for many data types and give you a sense of what certain data files, or **granules**, contain before you decide to order them. Spatial-coverage maps (see Figure 10) are produced in the WIST to show where one or more files are located on the Earth. Users submit data orders through WIST, and ECHO sends the order to the appropriate data center for fulfillment.

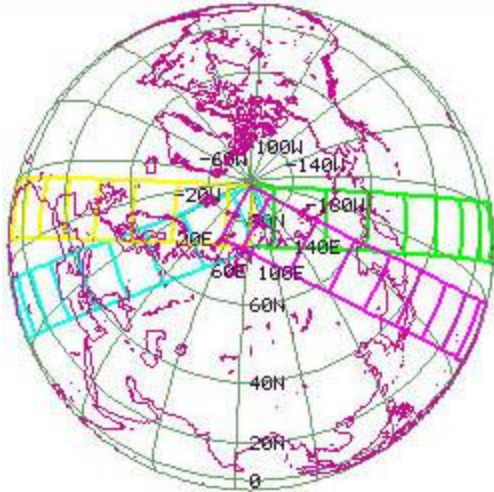


Figure 10. Spatial-coverage map for four AMSR-E Level-2A granules as displayed on the WIST. This map can be rotated online to view different parts of the Earth.

Users also have the option to **subset** certain data sets via the WIST using the HEW Subsetting Appliance (HSA), developed by the University of Alabama in Huntsville, Alabama USA. Data can be subsetted both geographically (i.e. using a specified latitude and longitude bounding box) and by desired parameters within the data. An advantage of subsetting is that it reduces the size of the data being distributed, thereby reducing the user's FTP transfer time and necessary storage space. Another advantage, of course, is that it allows you to receive data solely for the location in which you are interested. For example, a single MODIS Level-3 file covers an area of 2,300 km by 2,300 km, but you may only be interested in getting data for a smaller 25 km by 25 km region within that file. When ordering files on the WIST, subsetting will either be listed as available or unavailable, depending on the data type.

Due to the number of options and data sets possible with the WIST, some users may prefer to use NSIDC's less complicated Search 'N Order Web Interface (SNOWI). This simpler and perhaps more intuitive Web site has similar capabilities to the WIST, except that it does not provide spatial-coverage maps or subsetting. SNOWI also searches metadata in ECHO, but users can only order NSIDC data through this interface.

Rather than place orders via the WIST or SNOWI, users may also contact NSIDC User Services to set up **subscriptions** to have specific data automatically sent to or staged for them upon ingest at NSIDC. This distribution method is ideal for those users who wish to receive new data on a continual basis. Note that this option is only available for new data as they are ingested and not for data already archived at NSIDC. Though you can select a custom location, subscriptions currently do not allow subsetting.

Users may also directly download data via the NSIDC Data Pool, a large FTP server that holds all publicly available ECS data. The Data Pool is continually updated with new data as they are ingested at NSIDC. Users can browse the contents of the Data Pool by using the Web site or by initiating an anonymous FTP session to "ftp://n4ftlo1u.ecs.nasa.gov." These features allow users to directly download data rather than ordering data through the WIST or SNOWI and waiting for their orders to be completed. Subsetting of certain data types in the Data Pool is handled using the HDF-EOS to GeoTIFF converter (HEG), which also allows various data format and map projection conversions.

The Data Pool allows you to bookmark a particular search to find out what new granules that meet your search criteria have been ingested into the Data Pool since your last visit. Also, the Data Pool FTP site is structured in such a way that users can automate their own processes to download specific data, which are organized into predictable subdirectories by instrument, data type, and date.

Distribution is the key component to the ECS system. Ingest and archival of data have little purpose if there are no users who obtain these data. The ECS system can thus be likened to a library, and NSIDC Operators and User Services Representatives to its librarians. NSIDC plays an important role in "handing off" data that start with the satellite and end with students, scientists, and organizations who put those data to use.

5. Acronyms

The following acronyms are used in this document:

a.k.a. - also known as

AMSR-E - Advanced Microwave Scanning Radiometer - Earth Observing System

ASTER - Advanced Spaceborne Thermal Emission and Reflection Radiometer

AVHRR - Advanced Very High Resolution Radiometer

DAAC - Distributed Active Archive Center

ECHO - EOS Clearinghouse

ECS - EOSDIS Core System

EDOS - EOS Data and Operations System

EOS - Earth Observing System

EOSDIS - Earth Observing System Data and Information System

ETM+ - Enhanced Thematic Mapper Plus

FTP - File Transfer Protocol

GHCC - Global Hydrology and Climate Center

GLAS - Geoscience Laser Altimeter System

GUI - Graphical User Interface

HEG - HDF-EOS to GeoTIFF converter

HSA - HEW Subsetting Appliance

ICESat - Ice, Cloud, and land Elevation Satellite

IP - Internet Protocol

JAXA - Japanese Aerospace Exploration Agency

JPL - Jet Propulsion Laboratory

LP DAAC - Land Processes DAAC

MODAPS - MODIS Data Processing System

MODIS - Moderate Resolution Imaging Spectroradiometer

NASA - National Aeronautics and Space Administration

NSIDC - National Snow and Ice Data Center

PO.DAAC - Physical Oceanography DAAC

RGB - red, green, blue

SCF - Science Computing Facility

SIPS - Science Investigator-led Processing Systems

SMMR - Scanning Multichannel Microwave Radiometer

SNOWI - Search 'N Order Web Interface

ssh - secure shell
SSM/I - Special Sensor Microwave/Imager
SST - sea surface temperature
Vo - Version 0
WIST - Warehouse Inventory Search Tool

6. Glossary

The following terms are used in this document:

ancillary data - Measurements from other sources or sensors used to calibrate remote sensing data.

archive - The transfer of ingested data onto a permanent storage device.

band- A discrete range of wavelengths of the electromagnetic spectrum that a satellite remote sensing instrument measures (a.k.a. channel).

browse image - A low-resolution image that gives the user a sense of what a data file contains (a.k.a. quick-look image).

channel- A discrete range of wavelengths of the electromagnetic spectrum that a satellite remote sensing instrument measures (a.k.a. band).

distribution - The transfer of archived data to users who request them.

electromagnetic spectrum - The entire array of electromagnetic radiation (e.g., ultraviolet, visible, infrared, and microwave).

ephemeris data - Position and attitude data from the satellite during the time of remote sensing data acquisition.

false-color composite - An image that combines three bands that are not all in the visible portion of the spectrum.

FTP pull - An FTP session in which the user retrieves (or "pulls") data to his or her own computer.

FTP push - An FTP session in which data are automatically transferred (or "pushed") to a user-specified computer and directory path.

granule - The smallest data unit inventoried via ECS and distributed to users; typically, a granule is a single data file, though some granules may include multiple files.

ingest - The acquisition of data from external data providers.

Level-0 data - Unprocessed remote sensing data.

Level-1 data - Calibrated remote sensing data that have been referenced to time of acquisition and geographic location.

Level-2 data - Geophysical data derived from Level-1 data.

Level-3 data - Level-2 data gridded to a uniform map projection.

Level-4 data - Model outputs or results from scientific analyses derived from multiple measurements of lower-level remote sensing data.

metadata - Information about remote sensing data that may include such things as time of acquisition, size, geographic coordinates, quality assessment, and other information that is important for a user of the data to know.

pixel size - The measure of the smallest object that a remote sensing instrument can "resolve," or view (a.k.a. resolution). This is captured in a remote sensing image as one picture element or pixel.

quick-look image - A low-resolution image that gives the user a sense of what a data file contains (a.k.a. browse image).

remote sensing - Obtaining information about an object without actually coming into contact with it.

resolution - The measure of the smallest object that a remote sensing instrument can "resolve," or view (a.k.a. pixel size). This is captured in a remote sensing image as one picture element or pixel.

servers - Individual C++ programs that handle specific ECS tasks.

subset - To reduce a data file's contents to a specific geographic location and/or desired parameters within the data.

true-color composite - An image that combines three bands that measure light in the red, green, and blue portions of the electromagnetic spectrum.