THEMIS
Mission Archive Plan (MAP)

thm_soc_123_map_themis
March 8, 2013

1.0   Purpose

This document provides a plan that defines the current and permanent states of the data products, data handling, documentation and data analysis tools and defines a reasonable plan to achieve the permanent state of these products and processes for the end of mission (EOM).

2.0   Current and Permanent Data Processing and Data Inventory
2.1   Data processing and Handling
2.1.1   Science Operations Center
2.1.1.1 Overview

The THEMIS Science Operations Center (SOC) retrieves, processes, and archives all data files from the space-based probes and Ground-Based Observatories (GBOs). All data and data products are freely available with no passwords or security logons via the THEMIS project website and the THEMIS Data Analysis Software (TDAS). Numerous external institutions "mirror" the data and data products.

2.1.1.2 Space-Based Instrument Data Collection and Processing

Autonomous SOC scripts obtain scheduling information from the Mission Operations Center (MOC) and use this information to retrieve and validate space-based instrument raw data files produced during ground station probe contacts. Statistics produced during this process are stored in a MySQL database for operations personnel to access for review. The raw data files are archived on a Redundant Array of Independent Disks (RAID) system for subsequent processing and are backed up onto CD-R media. Retrieval of raw data files triggers production of numerous data products. Initially, the raw data files are converted into *Level 0* (raw packet) data files   The Level 0 files are then converted into L1 data files in Common Data Format (CDF). L1 files contain raw and un-calibrated data, but the CDF format allows wider access and is platform independent. Following creation of L1 data files, *L2* data files in CDF format are created. They include calibrated data in physical units. Both L1 and L2 files are archived on the RAID system. The Level 0-2 data products are used to produce Summary Data, mainly in the form of data plots which are available via the THEMIS website to the on-duty scientist (Tohban) to assess the quality and state of instrument data collection. Spacecraft ephemeris data are routinely updated and accessed to produce probe state files that are folded into Level 0-2 and summary data processing. These state files are used to produce orbit plots and ground tracks of the probes and are also available via the web site. Following reception of the raw ground station data files after a contact, the Level 0-2 processing is completed within 1 hour. L0, L1, and L2 products are then available from the THEMIS website under the URLs at http://themis.ssl.berkeley.edu/data/themis .

2.1.1.3 Instrument Information Available on Web Pages

This link (ftp://apollo.ssl.berkeley.edu/pub/THEMIS/3 Ground Systems/3.2 Science Operations/Science Operations Documents/THEMIS - ARTEMIS Instrument Information Available on Web Pages) provides a document with links to the following types of information concerning the THEMIS Instruments and their data: Overviews, Table of Parameters, Modes and Data Products, How to Select an Appropriate Data Type, Sources of Non Ideal Instrument Performance, FAQ's, Caveats as well other helpful documents.

2.1.1.4 Ground-Based Instrument Data Collection and Processing

Ground-based instruments include All Sky Imagers (ASI) and Ground Magnetometers (GMAG). The THEMIS project, in collaboration with the University of Calgary, has deployed 20 GBOs across Canada and Alaska. Hard drive retrieval from the sites takes about 3 months with an additional three months for the data to be transferred to UC Berkeley. Each GBO includes an ASI. Eleven GBOs have a UCLA GMAG. The remaining GBOs take advantage of existing GMAG networks, using co-located magnetometers near the GBO facility: 7 from the University of Alberta CARISMA network, and 11 from the University of Alaska Geophysical Institute GIMA network. The SOC retrieves magnetometer data from all 11 GIMA stations and processes these data into the same products as the UCLA and CARISMA magnetometer data. Using TDAS software, the University of Alberta converts the remaining CARISMA network magnetometer data (total of 23 stations) into the same L2 CDF data products, which are accessible to external users of the TDAS software. Additional GMAG networks have been recently added and include: AUTUMN (Athabasca), 5 stations; MACCS (9), USGS (14), NRCan (2), PENGUIn (Antarctica) (3), Greenland (34), AARI (Russia) (4), SGU (Sweden) (1), Leirvogur (Iceland) (1), McMAC (9).

2.1.1.5 ASI

There are 2 avenues for retrieving the ASI image data. The University of Calgary collects low resolution data daily via an internet connection to each GBO. An automated RSYNC process at the SOC also runs daily to "mirror" this data on the RAID system. Once the data are mirrored they are converted into L2 products in CDF format. These data are also accessible via the TDAS software. The second avenue of collection involves the retrieval of the GBO hard drives containing the high resolution images. Hard drives from stations in Canada are collected by the University of Calgary, while hard drives from stations in Alaska and Greenland are collected by SSL. After hard drive retrieval, the data are downloaded locally and validated. Alaska and Greenland drives are then shipped to the University of Calgary for final

archiving. In 2011 it was decided that data from Calgary should not be transferred to SSL via large transfer drives anymore, but instead high resolution data are copied to SSL about 4 times per year over the internet. Once the data are received at SSL they are converted into L2 products in CDF format. As of this writing, the low resolution data are complete up to February 10, 2013. High resolution data are complete for the whole system up to September 30, 2012 with some single station files ranging up to January 2013.

2.1.1.6 GMAG

Automated RSYNC and wget processes run daily to collect the GBO UCLA GMAG data from the University of Calgary, the GIMA data from the University of Alaska, the CARISMA data from the University of Alberta, and data from the other previously mentioned networks, and store these data on the RAID system. Once the data are retrieved they are converted into L2 products in CDF format. Daily stack plots are also made. SPDF runs a nightly file transfer, to ingest and republish the GMAG L2 CDFs from the THEMIS GBOs and other GMAG networks.

2.1.2 Probe Data Processing

After each probe-ground station contact, any telemetry files acquired by the ground station are transferred to the THEMIS SOC and automatically dispatched for processing to Level 0 (raw telemetry packets), L1 (time-tagged, uncalibrated data files in CDF format), and L2 (calibrated data CDFs in geophysically relevant coordinate systems) files. All data products are made available to the public via the THEMIS web site immediately after they are produced.

The VC (raw CCSDS telemetry frames) to L0 (decompressed CCSDS packets) to L1 (uncalibrated CDF) to L2 (calibrated CDF) data processing scripts, algorithms and software are under version control in the THEMIS Subversion (SVN) repository under the "tmtools" directory. The processing is documented in the THEMIS L1 File Definition, and in the Developer's Guide.

2.1.3 Ephemeris Data

During normal operations THEMIS orbit solutions are generated on Mondays, Wednesdays, and Fridays of each week. 30-day predictive ephemerides, including upcoming maneuver operations, are generated from these orbit solutions for scheduling and operational use. The predictive ephemerides are available to end users via the V00 and V01 STATE CDFs. Definitive ephemerides are created from the orbit solution archive on a weekly basis, covering the previous calendar week, and are available via the V02 STATE CDFs. Additional spin phase and spin axis attitude corrections, derived from post-processed FGM data, are generated every few months and are available in the V03 STATE CDFs. (V03 STATE processing is no longer routinely performed for the two ARTEMIS probes, since the spin phase and spin axis correction algorithms do not work well given the weaker magnetic field in the lunar orbit environment).

2.2 Inventory of Data Products

2.2.1 L1 CDF Data Products

The L1 CDF files contain one file per data type, each covering approximately one UTC day. The L1 CDFs are available via the THEMIS UCB SSL Data Server or web site immediately after the L1 processing completes. L2 data are currently routinely archived at the VMO and SPDF. Based on an agreement with the SPDF, they maintain an archive of all L1 Probe Based data files and the information needed to construct more complicated parameters from them, such as particle pitch angle and phase space distributions. Note that these parameters are already routinely available to users of TDAS and the THEMIS data system at UCB. TDAS is easily configured to use the UCB data server, the SPDF archive, or any of several THEMIS mirror sites as a data source for downloading CDF data products. The SPDF data repository is updated nightly to keep the SPDF L1 and L2 THEMIS archives up to date with the state of the data products at the UCB server.

2.2.2 L2 CDF Data Products

L2 THEMIS CDF files contain THEMIS calibrated data quantities in physical units. These data can be used by scientists directly; the instrumental details have been accounted for in the calibration process. L2 data files are stored in the permanent archive at UCB and are distributed to mirror sites and the SPDF. L2 data files are created daily using L1 data and calibration data, and are updated and reprocessed when necessary. For example, updates in calibration data for a given instrument/time period will result in new L2 data files for that instrument and time period. Currently (as of 15-feb-2013), there are L2 files for all THEMIS instruments, including EFI, ESA, FBK, FFT, FIT, MOM, SCM and SST data.

All THEMIS L2 CDFs, with all their ancillary products, are archived and retrievable via SPDF's CDAWeb. The software and the master CDFs for the L2 files are under version control using Subversion (SVN).

2.2.3 L1 and L2 CDF Data Descriptions

Please select the following link to see the current L1 or L2 CDF descriptions: http://themis.ssl.berkeley.edu/esa_desc.shtml.
On each instrument or data type's L1 or L2 CDF Data Description web page, there is also a link to the Data Variable Description Document, where all scientific parameters and variables are listed by instrument, level (L1 or L2), variable name and description of the variable.
In addition the following table describes THEMIS science parameters: ftp://apollo.ssl.berkeley.edu/pub/THEMIS/3 Ground Systems/3.2 Science Operations/Science Operations Documents/THEMIS Science Parameters.

2.2.4 Summary, ASI Keogram and Tohban Plots

The THEMIS web site at http://themis.ssl.berkeley.edu/summary.shtml?autoload=1 provides pre-generated summary plots of THEMIS data online. The interface allows zoom-in and panning to a variety of time ranges. Plots are provided for THEMIS particle, field

and ground instruments, as well as for scientifically relevant derived and reference quantities. A list of the 26 plots we provide and their descriptions can be found at: ftp://apollo.ssl.berkeley.edu/pub/THEMIS/3 Ground Systems/3.2 Science Operations/Science Operations Documents/THEMIS Summary Plot Description Tables/ . The parameters displayed for single spacecraft are described in detail at http://themis.ssl.berkeley.edu/key.png.

2.3   Availability and Access

All data and data products are freely available via the THEMIS project website (see section 4.2) and the THEMIS Data Analysis Software (TDAS - see sections 4.3 and 4.4). In addition, numerous external institutions "mirror" some or all of the data and data products. The Space Physics Data Facility (SPDF) mirrors the L1 and L2 data created at the THEMIS SOC and makes the data and data products available to the wider space physics community via the CDAWeb and SSCWeb web sites. The Japanese Aerospace Exploration Agency (JAXA) downloads THEMIS data to provide magnetic field and particle data plots, spacecraft footprints and ground-based instruments, and a conjunction finder on their website. The Taiwan ASCII and IDL SAVE FTP System downloads CDF data from the UCB SOC and makes those data available in ASCII and IDL save format. Additional THEMIS mirror sites are maintained in France by the Plasma Physics Data Center (CDPP) at the Center for Space Radiation Studies (CERS), and by the Institute of Geophysics and Planetary Physics at UCLA. The TDAS software is easily configured (as described in the User's Guide) to use the SPDF, JAXA, CDPP, or UCLA mirror sites as alternatives to the THEMIS data server.

2.4   Evidence of Use

The table shows the usage of THEMIS Science Data Products and Web Pages served by the UCB SOC, for March 2012 through mid-February 2013. Note: these totals include requests for both THEMIS and ARTEMIS data, as well as data from the GBOs and other GMAG networks.

From January 2010 through February 2013, SPDF's CDAWeb has received over 858,258 requests for THEMIS datasets. There were also over 1.6 million CDAWeb FTP THEMIS file requests, representing 15.6 terabytes of data volume. (Note: Additional data requests fulfilled by the various mirror sites listed above are not included in these statistics.) In addition, during 2012, the SSCWeb had 11,818 requests to see THEMIS and 1081 requests for ARTEMIS Orbit information.

Since the last Senior Review, there have been over 300 refereed THEMIS and ARTEMIS publications in Space Science Reviews, GRL, JGR, Annales Geophysicae, Physics

Summary by Month										
Month	Daily Avg				Monthly Totals					
	Hits	Files	Pages	Visits	Sites	KBytes	Visits	Pages	Files	Hits
Feb 2013	119250	106377	2344	113	919	12771917076	1477	30480	1382906	1550260
Jan 2013	59642	51733	2530	141	4460	17721187137	4371	78438	1603747	1848932
Dec 2012	59165	52462	2920	122	6260	10283374936	3795	90546	1626343	1834126
Nov 2012	67086	62629	3599	182	6463	27716995086	5467	107979	1878876	2012603
Oct 2012	94098	73560	9000	215	6774	22732804737	6668	279028	2280373	2917043
Sep 2012	49171	43571	2526	166	6501	11084903550	4993	75809	1307147	1475137
Aug 2012	90747	76125	3205	167	6713	15740826667	5191	99364	2359900	2813169
Jul 2012	87031	74017	5304	154	5017	18409225259	4792	164438	2294541	2697984
Jun 2012	92852	75991	3314	149	1856	34584794795	4483	99442	2279749	2785578
May 2012	126862	110129	6389	144	2151	26417692035	4490	198076	3414005	3932745
Apr 2012	106229	73387	6024	136	2342	45646422686	4103	180721	2201613	3186891
Mar 2012	81746	62451	4371	161	2551	31426477073	4996	135531	1935987	2534126
Totals						274536621037	54826	1539852	24565187	29588594

of Plasmas, Physical Review Letters, and other journals, with over 130 articles in refereed journals in 2012 alone. THEMIS/ARTEMIS studies dominate the magnetospheric community, with more than 100 abstracts at the Fall 2012 AGU meeting.

2.5   VO's

The goal of the VMO is to facilitate query-based discovery and provide access to past, present, and future NASA high-altitude magnetospheric missions. The core function of the VMO data environment is to search and retrieve pointers to magnetospheric data while presenting the user with a common interface, either a web interface or an application programming interface (API). The THEMIS mission, launched in February 2007, is part of this effort. Dr. Vassilis Angelopoulos, a VMO Co-Investigator, supervises the scientific aspects of the effort, in terms of data quality and compatibility of the submitted data sets with the other data sets in the VMO database for maximum science return. His programming staff has described the THEMIS data sets in SPASE (Space Physics Archive Search and Extract) terms and set up a data service for VMO. They continue to facilitate interactions between the distributed THEMIS team (UCB, LASP, CETP, TUBS) and the VMO group. Dr. Angelopoulos is a member of the VMO SPASE definition group. The VMO effort began in 11/2007 and was completed in 5/2011.

2.6   Relationship of Metadata to SPASE Data Model

The THEMIS metadata is currently stored in the CDFs that serve as the primary means of THEMIS data distribution. Most of the attributes required by the SPASE model can be found in these CDFs. Any additional metadata that are required can be found in text-based mission documents. The CDF data do not, however, have the hierarchal structure of SPASE XML. SPASE metadata are generated by first using the online CDF to SPASE XML converter. This turns the flat CDF structure into a hierarchal SPASE XML structure. Then the metadata are inspected by hand and any missing data are added. Finally, the metadata XML files are reviewed for content and structure completeness by the VMO team before they are made available for the VxO's. VxO's SPASE XML files have been made available to the GSFC VMO and the UCLA VMO. In order to search THEMIS data products and possibly dynamically generate higher level data products, middleware will be made available as needed. The requirements of each VxO have been met by collaboration between Goddard Space Flight

Center (GSFC), University of Maryland, and University of California at Berkeley and Los Angeles. 'The THEMIS data are available through the VMO interfaces (http://vmo.nasa.gov) shortly after THEMIS SPASE metadata are reviewed and approved by the VMO team.

2.7   Ground-Based Observatories (GBOs)

The THEMIS GBOs create two major data products: all-sky images of the sky and measurements of the local magnetic field vector. The local magnetic field vector is transmitted in real time at a cadence of 2 measurements per second. The all-sky images are reduced in size for real-time transmission into 1024-pixel representations (thumbnails) of the sky brightness that cover approximately 15 degrees of magnetic longitude and 8 degrees of magnetic latitude and are mapped to 110 km altitude. Full resolution sky images of 256x256 pixels are stored locally at each site on hard drive and are retrieved approximately every 4 months. Thumbnails are used to create auroral activity overview plots very quickly (about 24 hours later). Full resolution images are made available approximately 6 months after data collection. However, after an extraordinary event it is possible to retrieve and process a limited amount of full resolution data (1-2 hours) within 4-6 days. Currently there are ASK, ASF and AST L1 CDFs.

2.8   Resident and Permanent Archives

All THEMIS data and data products are archived on the RAID system. Currently, this includes approximately 86 Terabytes of data. Autonomous scripts archive raw space-based data to CD-R media as soon as 650 Mbytes accumulate (~every 5 days). The full complement of GBO data (ASI and UCLA GMAG) are retrieved by the University of Calgary and archived on a RAID system. UCLA and the University of Alberta run daily RSYNC jobs to retrieve and archive all GMAG data resident on the THEMIS SOC RAID.

The archive of all on-orbit L0, L1, and L2 data products is available from the THEMIS website as described in section 2.1.1.2. Mirror sites at SPDF, JAXA, UCLA, and CESR are available, updated nightly as additional data is produced by the THEMIS SOC, and are usable as alternative data repositories as described in section 2.3.

2.9   Realistic schedule for Enhancements - Summary

The enhancements listed in this section are scheduled to be completed by 9/2015. The SST calibration software will be updated to include automated removal of sunlight contamination, and interpolation of SST efficiency parameters to revised energy bins (to account for diode degradation). The L2 SST CDFs will be updated to include SST moments. The SPEDAS (multi-mission data analysis and visualization framework; see section 4.3.2) components of TDAS will be further enhanced to support development of TDAS-compatible plugin software by other missions.

3.0   Curent and Permanent Documentation Inventory
3.1   Inventory, Availability and Access of Spacecraft, Instruments and Instrument Calibration Documents

The current and permanent inventory of THEMIS documentation can be accessed from the THEMIS Documentation web page http://themis.ssl.berkeley.edu/software_docs.shtml. A THEMIS Documents Status List is available at: ftp://apollo.ssl.berkeley.edu/pub/THEMIS/3 Ground Systems/3.2 Science Operations/Science Operations Documents/THEMIS Documentation Status Lists Access to all THEMIS documentation is fully open with no accounts, passwords or registration procedures required. Currently 23 of the 25 required documents have been completed and the remaining will be completed by 9/2015.

4.0   Current and Permanent Data Analysis Tools
4.1   Trainings

Training presentations on the use of THEMIS data sets and software available from the THEMIS Web Site are routinely conducted at the Summer GEM and Fall AGU conferences. Typical training agendas include the THEMIS Web Site, installation of the Software, downloading data and use of the Graphical User Interface (GUI) and availability of information in the THEMIS User's Guide. In total, over 350 people have attended training sessions. In 2012, MMS and the Van Allen Probes, among other projects, have expressed interest in using TDAS as a platform for their data analysis needs, and THEMIS developers have attended several of their meetings and given TDAS presentations in support of these collaborative software development efforts.

4.2   Web Site

A wealth of information is available via the THEMIS web site at http://themis.ssl.berkeley.edu/index.shtml. 'The Mission' drop down menu provides information on the spacecraft, instruments, orbits, status, as well as links to many of our domestic and foreign partners. The 'Data' drop down menu facilitates access to all THEMIS data variables and their descriptions. In addition a summary plot viewer allows users to peruse individual probe, multi-probe, ground, and orbit data. Summary plots are available within 1 day of data reception. The 'Software' drop down menu provides access to the TDAS Software, Software Users Guide, all THEMIS Documentation and the list of Monthly THEMIS Enhancements. The 'Publication' hotspot links to scientific nuggets and the THEMIS Publication FTP site. All of the THEMIS Data, Software and FTP sites are totally open and available without any logon or passwords required.

4.3   Science THEMIS Data Analysis Software (TDAS)
4.3.1   TDAS Description

The THEMIS Data Analysis Software (hereafter referred to as TDAS) is an IDL-based software package that can be used to download, open, analyze and plot L1 or L2 data and process L1 data into L2 data quantities. The TDAS software is accessed via the IDL

command line interface and a Graphical User Interface (GUI) developed at SSL. For users that do not have an IDL license, THEMIS provides a version of the GUI package that can be called from the IDL virtual machine, which does not require a license.

4.3.2   TDAS organization

The TDAS software tree is divided into three branches. The "themis" branch contains software that is used for THEMIS data extraction and analysis. The "ssl_general" branch contains general use and utility software that is designed for data analysis of space science data. It is used extensively for previous missions such as FAST, WIND and POLAR. The software in the first two branches has been developed and is currently maintained at SSL. The third branch, "external," is reserved for software that is developed elsewhere and distributed as part of the TDAS package. Currently this branch contains the IDL-GEOPACK package, which is an IDL application for N.A. Tsyganenko's magnetic field modeling package.

Several other missions have expressed interested in leveraging TDAS as a platform for their own data analysis needs. The TDAS developers have proposed a set of extensions to TDAS called SPEDAS (Space Physics Environment Data Analysis System), to support the community in developing additional "plugin" modules for TDAS. The ERG and IUGONET projects have already developed plugin code using the SPEDAS framework to load, analyze, and display their data, providing a proof-of-concept that the SPEDAS framework will be a useful tool for more robust multi-mission data analysis capabilities.   The Van Allen Probes-EFW and MMS projects have also expressed interest in using SPEDAS as a data analysis and visualization platform.

4.3.3   TDAS Version Control

TDAS software management and version control is accomplished using the Subversion version control system (http://subversion.tigris.org/). All versions of the software are contained in a password-protected repository. Software developers and scientists submit code and code revisions to this repository and the software is exported from the repository in a series of "releases". A major release is issued approximately once every year, following an extensive QA process. Occasionally, bugs are found in the released version; when this occurs the bugs are fixed and a minor release is issued. Also, upgrades and enhancements that are deemed to be useful to the community (such as improvements in calibration software) may be included in a minor release. The current released version of the software is v7.01. In addition to the releases, a snapshot of the subversion repository is created every night and is made available on the THEMIS web site, which is described below. This "bleeding edge" software package always contains the most up-to-date version of the TDAS package, but is only unit-tested. A new "bleeding edge" version is built and released every night. The user community has a choice of using the latest released version or the latest "bleeding edge" version. Both can be freely downloaded from the THEMIS Software web page.

4.4   TDAS Command Line and GUI

The IDL command line environment gives users full access to the TDAS package, but a familiarity with IDL and the IDL-based "tplot" package used for data analysis is necessary. There are many examples in the form of crib sheets distributed with the software. For users unfamiliar with IDL, the THEMIS software development team has created a Graphical User Interface (GUI). The THEMIS GUI is built on top of the command line software. It provides full capability for loading and plotting THEMIS data, and support for some data analysis tasks, such as coordinate transformations, time-series analysis and power spectra. The on-line overview plots are also available as a GUI option, in case the user wants to specify the exact plot time range. The THEMIS GUI now loads GOES, Wind and ACE data as well. It is the only means to access high time resolution (0.5s) GOES magnetometer data on-line. The command line and GUI interface is extensively documented in the THEMIS User's Guide which can be downloaded using the following link: ftp://apollo.ssl.berkeley.edu/pub/THEMIS/3 Ground Systems/3.2 Science Operations/Science Operations Documents/Software Users Guides

4.5   Permanent Enhancements

The enhancements listed in this section are scheduled to be completed by 9/2015. The SPEDAS components of TDAS will be further enhanced to support development of TDAS-compatible plugin software by other missions. Plugins have already been developed for ERG, IUGONET, and OMNI; the Van Allen Probes-EFW and MMS missions have also expressed an interested in developing additional plugins.

4.6   THEMIS FTP Site

No logon or password is required to access the THEMIS FTP site. The site provides THEMIS Documents, Publications, Science Data Analysis Tools Enhancement Reports, Training Presentations and Meetings information. The THEMIS FTP site (ftp://apollo.ssl.berkeley.edu/pub/THEMIS/) is served by SPRG's main file server, apollo. Our backup server performs incremental saves of the THEMIS FTP site nightly using the EMC Networker backup software. A full backup of the THEMIS FTP site is performed on the 4th day of each month. Backups of the THEMIS FTP site are retained on tape for four months.

5.0   Summary

This THEMIS Mission Archive Plan (MAP) describes the current state of the THEMIS data products, data handling, documentation and data analysis tools. For all of these areas the MAP has delineated the mission's current status and its objectives for the possible end of the mission in 9/2018. The MAP details our plan to complete production of data products, data handling, documentation, and data analysis tools.