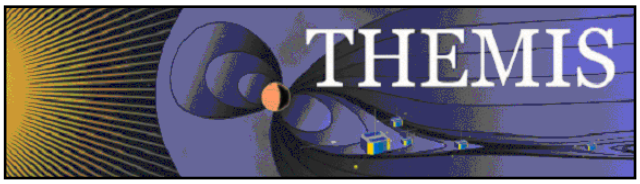


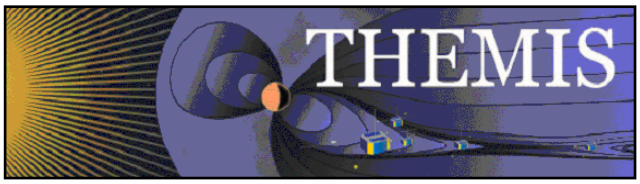
THEMIS
Science Data Analysis Software (TDAS)
Developers Guidelines

THM-SOC-128
March 2010

Pat Cruce, THEMIS Science Software Engineer

David King, THEMIS Science Software Manager

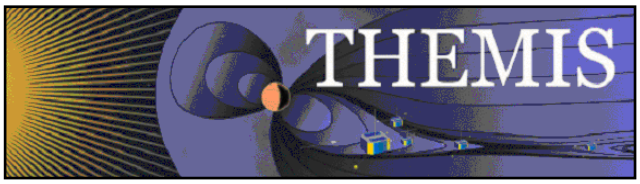
Vassilis Angelopoulos, THEMIS Principal Investigator



Document Revision Record

Rev.	Date	Description of Change	Approved By
-	3/25/2010	Initial release - Pat Cruce	D. King

Table of Contents



1. Purpose and Scope

As resources and priority dictate The THEMIS project will develop a comprehensive Software Developers Guide. In the interim this document relates the most important of the necessary guidelines for developers to develop software for TDAS.

2. TPLOT

2.1 Introduction

At their most basic, TPLOT variables, as implemented by “get_data” and “store_data”, are really just a map where a key-string is associated with three arbitrary data structures of any IDL type. To guarantee that TDAS routines can properly use these TPLOT variables we maintain a set of naming and structural conventions for each component of a tplot variable.
The following command shows each component of the tplot variable as used in the store_data command:

store_data,'varname',data=data,dlimits=dlimits,limits=limit

The following four sections, varname, data, dlimits, limits, will describe each of these components.

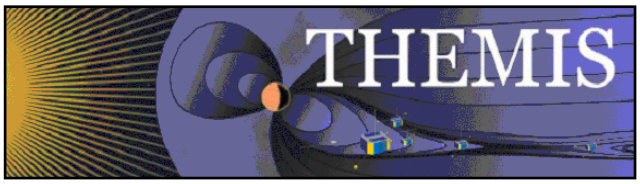
As a caveat, note that these are conventions we try to follow in writing new code, but they are not always consistently adhered to in older routines. As development continues we are working to make the whole code base more consistent and standards compliant.

2.2 Variable Names

This value can be any string that does not contain the characters: ' '(space),'*'(asterisk),'?'(question-mark),'['(left-bracket),']'(right-bracket),'\'(backslash). By convention, TPLOT names do not include other extraneous punctuation like parentheses and periods, but underscores are used to delineate separate fields in names. The generic naming convention is normally (source/site/spacecraft)+'_'+(instrument/category)+'_'+(datatype/specific type). For example: "tha_state_pos"

Detailed descriptions of the specific naming conventions used by THEMIS can be found in the document "thm_soc_105_FIELDS_VARNAMES_YYYYMMDD.pdf" This document is available on the THEMIS website(http://themis.ssl.berkeley.edu)

It is strongly recommended that all data being loaded by the TDAS software follow some sort of consistent and unique naming convention so that data can be found programmatically using globbing string searches.



2.3 Data Components

This value can be either a struct or a space-separated list of tplot names.

2.3.a Struct:

The struct-based data component is the more basic type for the data component. At a minimum it requires two tags, 'x' and 'y', it can optionally have a tag called 'v' or two tags called 'v1' & 'v2'.

2.3.a.1: x-component

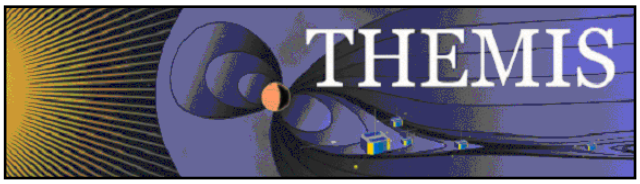
* The x-tag should be bound to the time-array associated with this variable.
* The time-array should be a one dimensional double precision array with N elements, where N is the number of time samples in the data quantity.
* Each element in the time-array should be a unix time(an offset in unleaped seconds since 1970-01-01/00:00:00).
* The time-array does not need to have a constant cadence(ie the difference between adjacent elements can vary from element to element).
* The time-array is required to be ascending (the i+1th element should always be greater than the ith element).

2.3.a.2 y-component

* The y-tag should be bound to the data values sampled at each time.
* It can be a one, two, or three dimensional array of any numeric type(complex type is not consistently supported throughout TDAS).
* The first dimension of the y-component should always have N elements where N is the number of elements in the time-array(X).
* The second dimension is optional and is used to group data with the same times. In the case of 3-vectors, the y-component should be an Nx3 array. In the case of spectral data, the y-component should be an NxM array, where M is the number of channels, frequencies, energies, etc... in the data.
* A third dimension can also be added with L elements. The third dimension is generally used when a two dimensional data point is collected for each time. For example, an MxL image may be collected at each time sample N, or a time-varying rotation matrix may be represented as an Nx3x3 array (N x Column x Row).
* TPLOT variables with 3-dimensional y-components are not always handled consistently by TDAS routines, and often have a special collection of routines to deal with that particular type of 3-d data.(for example time-varying rotation matrix routines are in ssl_general/cotrans/special)

2.3.a.3 v-component

* The v-tag should be bound to the scaling data for the M component of a variable with a two dimensional y-component.
* The v-component is optional. If not-present, it is assumed that data is spaced evenly.(ie it is equivalent to a v component equal to dindgen(M))



* The v-component is used to determine the spacing between the components along the non-temporal axis when transforming and plotting the data. This can be considered a second independent variable(after time)
* The v-component can be a one or two dimensional array of any numerical type(complex type is not consistently supported throughout TDAS).
* A one dimensional M-element v-component indicates that scaling is constant across time. For example, an FFT may have the same frequencies associated with frequency bins at each time.
* A two dimensional NxM element v-component indicates that scaling varies across time. It will contain scaling values for all M-channels at each time in N. For example, a particle instrument may be configured to have 16 energy bins at one set of times, and 32 energy bins at the times after a mode change. In such a case, the v-component would have Nx32 elements, and during the times when there are only 16 valid energy bins, the last 16 bins of the array would contain NaNs.

2.3.a.4 v1/'v2-components

* The v1/v2 tags should be bound to the scaling data for the M & L components of a variable with a three dimensional y-component
* The v1/v2 components are optional, but if one is present, the other should be present as well.
* Interpretation of the v1/v2 components and dimensional restrictions varies depending upon the routines used to process/plot the data.

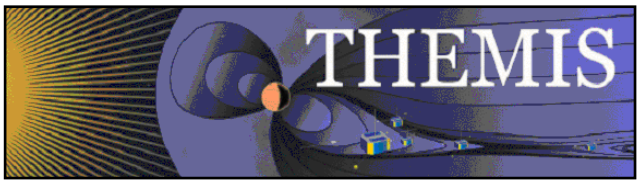
2.3.b List of TPLOT Names:

* A data component composed of a list of space separated variable names is called a pseudo-TPLOT-variable or a multi-TPLOT-variable.
* This convention is used to bind data quantities that have different dimensions together to indicate that they should all be plotted on the same plot.
* Data processing routines generally do not accept pseudo-variables as inputs.
* Settings that apply to all names listed in the pseudo-variable may be stored on the dlimits/limits of the pseudovariable, rather than on the dlimits of its individual component variables.

2.3.c Data examples.

Examples of all the data-forms using dummy data and names follow:

```
store_data,'varname1',data={x:dindgen(100),y:dindgen(100)}
store_data,'varname2', data={x:dindgen(100),y:dindgen(100,3)}
store_data,'varname3',data={x:dindgen(100),y:dindgen(100,15),v:dindgen(15)}
store_data,'varname4',data={x:dindgen(100),y:dindgen(100,15),v:dindgen(100,15)}
store_data,'varname5',data={x:dindgen(100),y:dindgen(100,200,300)}
store_data,'varname6',data={x:dindgen(100),y:dindgen(100,200,300),v1:dindgen(200),v2:dindgen(300)}
store_data,'varname_pseudo',data='varname1 varname2'
```



2.4. dlimits:

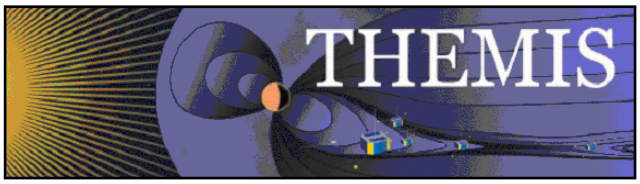
The dlimits struct is used to store meta-data(data that describes the data). The data in the dlimits is, by convention, set only by load routines and data processing routines. The dlimits has three components, the CDF substruct, the data_att substruct, and default plot settings.

2.4.a CDF substruct:

* The cdf substruct is a component generated automatically by cdf2tplot(and routines that use cdf2tplot).
* The cdf substruct should be considered read-only by all routines after load.,
* The cdf substruct has the following tag_name/signature: dl={cdf:cdf_substruct}, where cdf_substruct indicates a variable pointing the structure itself.
* The cdf substruct contains four elements: filename,gatt,vname,&vatt.
* filename is a string naming the cdf from which the variable was loaded
* vname is a string naming the variable that was loaded from the cdf.
*gatt is a sub-struct containing all the general attributes in the cdf.
*vatt is a sub-struct containing the variable attributes from the cdf that are specific to the loaded variable.

2.4.b data_att substruct:

*The data_att substruct contains non-plottable attributes of the data.
*These are used for informational purposes and for allowing various data processing routines to make various automatic decisions.
*These are normally set initially by load routine post-processing functions, but they may be modified by data processing routines and the user as necessary.
*data_att should contain, the following three elements, if applicable. units,coord_sys,st_type
units: A string containing the units for the variable. This string should not contain any additional formatting, but should include any SI prefixes that are relevant(ie list units as nT, not T). Units use the following symbols to indicate operators in compound units. '/'(slash:division),'('(left-parenthesis:open grouping),')'(right parenthesis: close grouping), '^'(up-carot: exponentiation),''(asterisk:multiplication). The strings 'unknown' or 'none' may also be used.
*coord_sys: string that contains a short identifier to identify the coordinate system of the data. Valid names include: 'spg','ssl','dsl','gei','gse','gsm','sm','geo','sse','fac','mva','rxy'. Definitions of many of these coordinate systems can be found in the document:
thm_soc_110_COORDINATES_YYYYMMDD.pdf
If one of these coordinate systems does not accurately describe the coordinate system of data, it is better to create a new identifier, rather than use one that does not fit. Alternatively, the strings ' unknown' or 'none' may be used.
*st_type: string that indicates whether a quantity is a position,velocity, or neither. This is required to properly transform some quantities into different coordinate systems. Valid values are 'pos','vel', or 'none'
*Other informational components may be included in the data_att like 'project','instrument','observatory','filename', but these values are not required for transformations.



2.4.c plotting options:

*Plotting options exist at the top level of the dlimits struct.
*They are commands and settings to the tplot & tplot_gui routines.
*They can be set using the 'options,/default' routine or by storing the dlimits struct directly.
*Common components include:

- **spec:1 Indicates that a quantity should be plotted as a spectrogram.
- **x/y/zrange:[0D,1D]: Indicates the data range.
- **x/y/zlog:1 Indicates the axis should be logarithmic.
- **labels['label1','label2','label3'] Indicates y-axis labels.
- **colors[0,1,2]: Indicates color indexes for the data.

*Additional values/examples can be found by consulting the TPLOT cribs in themis/examples/tplot_cribs/
*Plotting options can also be placed in the dlimits struct of a pseudo-TPLOT-variable

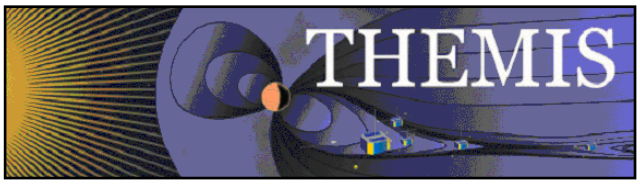
2.4.d dlimits examples:

Examples of hypothetical dlimits storage operations:

```
store_data,'varname1',dlimits={cdf:cdf_struct,data_att: $
 {coord_sys:'gse',st_type:'pos',units:'km'},colors:[2,4,6],labels:['X','Y','Z']}
store_data,'varname2',dlimits={cdf:cdf_struct,data_att:{coord_sys:'none',st_type: $
 'none',units:'eV/(cm^2*sec*sr*eV)'},spec:1,ylog:1,zlog:1,ysubtitle:'eV/cm!U2!N!C-s-sr-eV'}
```

2. 5. limits structure:

* The limits structure contains the same types of plot settings that are in the dlimits.
* The limits doesn't contain a data_att or cdf component, and can be 0 to indicate no limits have been set.
* TPLOT settings in the limits structure will override settings in the dlimits.
*The limits structure can and should be modified by individual user commands during an interactive plotting session.



3. CDF

3.1 Introduction:

THEMIS CDFs generally follow ISTP and SPDF guidelines for CDF construction. Detailed documentation can be found on their Web-Site. Examples can also be found in TDAS routines ssl_general/CDF/cdf2tplot.pro and themis/common/thm_load_xxx.pro, as well as in the various load routines in the TDAS distribution. There are two significant deviations from the standard.

3.2 DEPEND_TIME:

All CDF variables that are loaded into our software via cdf2tplot.pro require an attribute called DEPEND_TIME that specifies the name of a CDF variable in the same file which contains UNIX times. UNIX times are double precision offsets in unleaped seconds from the UNIX epoch: 1970-01-01/00:00:00.

3.3 DEPEND_0:

ISTP requires a DEPEND_0 attribute on time-varying variables. The value of the DEPEND_0 attribute must be a variable name in the same CDF, with type CDF_EPOCH or CDF_EPOCH16. For THEMIS CDFs, this variable will be marked VIRTUAL, and will not contain any values. It will have a "FUNCT" attribute specifying the name of an IDL function, and COMPONENT_0 and COMPONENT_1 attributes specifying the arguments to that function, which can be used by SPDF to fill in the CDF_EPOCH or CDF_EPOCH16 timestamps from the DEPEND_TIME variable and epoch. Default THEMIS values for "FUNCT" are "COMP_THEMIS_EPOCH" or "COMP_THEMIS_EPOCH16"

4. TDAS Branches

4.1 Introduction:

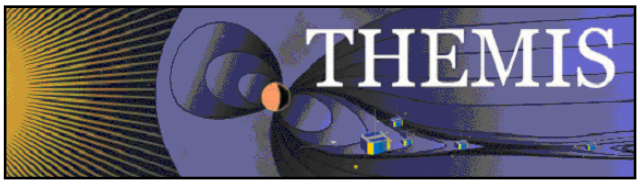
The TDAS release software currently has three different release branches. SSL_GENERAL, THEMIS, EXTERNAL. A fourth branch QA_Testing, is part of TDAS development, but not of software releases. Each of these branches serves a different purpose and has different dependencies. It is important that new software be placed in the correct branch.

4.2 SSL_GENERAL:

This contains general purpose library routines. These include routines for plotting, analysis, coordinate transformation, data structures, and programming tasks. Any routines in this branch should depend only upon other routines in this branch and on built in IDL routines. This branch contains many historical routines. Great care must be taken to ensure backwards compatibility and changes to existing routines should only be made if absolutely necessary. Note: Dependency rules are not enforced for cribs.

4.3 THEMIS:

This branch contains routines for the THEMIS mission. These include load routines, calibration routines, THEMIS specific analytical transformations, THEMIS specific coordinate transformations, routines to generate THEMIS summary plots, and the THEMIS GUI. This branch can depend upon routines in the SSL_GENERAL branch or the EXTERNAL branch.



4.4 External:

This branch contains packages external to the TDAS software, and any routines developed to interface with those packages. This includes the IDL_GEOPACK interface to the Tsyganenko models, the CDAlib, and external developers' routines. The developers' routines directory includes additions to the THEMIS software that were not developed by THEMIS programmers, but which are nonetheless useful enough to merit release with the package. The EXTERNAL branch may have dependencies to routines in the SSL_GENERAL branch, but not to the THEMIS branch. Note: Dependency rules are not enforced for cribs.

4.5 QA_TESTING:

This branch contains routines and suites to perform quality assurance testing before a release of the TDAS software. This include automated test scripts, human run test suites, and data for regression comparisons. Dependency restrictions are not enforced for the QA_TESTING branch. Dependencies are formed as needed to verify the correctness of release software.

5. Documentation

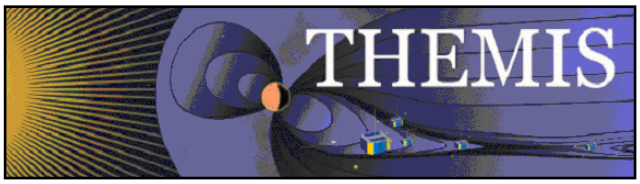
5.1 Introduction:

Nightly builds of TDAS automatically build HTML documentation for routines if documentation conventions are followed. It is also of the utmost importance that documentation be detailed and complete so that users can understand how to correctly use TDAS, to make sure that Scientists can trust the TDAS calculations, and to guarantee ease of code maintenance for future developers. TDAS code is documented in three ways: detailed routine header descriptions, clear code style, and cribs.

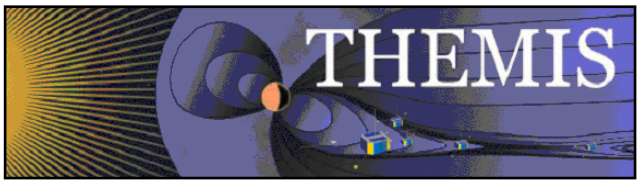
5.2 Routine Header Example:

From ssl_general/mini/calc.pro

```
;+
; Procedure: calc
;
; Purpose:  This routine takes a string as input and interprets the string as a mini-language.
;    This language can manipulate normal idl variables and tplot variables.  The idl variables
;    that can be modified are non-system variables in the scope in which the procedure is called.
;
; Inputs: s : A string that will be interpreted as the mini language
;
; Keywords: error:
;              If an error occurs during processing this keyword will return a struct
;              that contains information about the error.  If error is set at time of input
;              this routine will set it to undefined, before it interprets 's' to prevent internal errors.
;
;           function_list: return an array of strings that lists of the names/syntax of
;                          available functions in the mini_language.  Return without processing 's'
;                          if an argument in the returned name is in brackets it is optional
;                          for example:  min(x[,dim])
;           operator_list:  return an array of strings that lists of the names of available
;                           operators in the mini_language. Return without processing 's'
;
```



```
;           verbose:  set this keyword if you want the routine to print errors to screen rather
;                     than only return them via error
;
;           gui_data_obj: If 'calc' is being used inside the gui, then the loaded_data object will
;                         be passed in through this keyword.  NOTE: end users should not even set
;                         this argument.
;
; Outputs: none, but it will modify various variables in your environment
;
; Examples:
;    calc,'a = 5'
;    calc,'"pos_re" = "tha_state_pos"/6374'
;    calc,'a += 7'
;    calc,'"tvar" = "tvar" + var'
;    calc,'"tvar" = ln("tvar")'
;    calc,'"tvar" = total("tvar"+3,2)'
;    calc,'"tvar" = -a + 5.43e-7 ^ ("thb_fgs_dsl_x" / total("thb_fgs_dsl_x"))
;    calc,operator_list=o,function_list=f
;
; Notes:
;    1. The language generally uses a fairly straightforward computational syntax.  The main
;       difference from idl is that quoted strings are treated as tplot variables in this language
;    2. A full specification of language syntax in backus-naur form can be found
;       in the file bnf_formal.txt, the programmatic specification of this syntax
;       can be found in productions.pro
;    3. The language is parsed using an slr parser.  The tables required to do this parsing
;       are generated and stored ahead of time in the file grammar.sav and parse_tables.sav
;    4. The routines that describe the evaluation rules for the language can be found in the file
;       mini_routines.pro
;    5. If you want to modify the way the language works you'll probably need to modify productions.pro,
;       regenerate the slr parse tables using save_calc_tables and modify/add routines to mini_routines.pro
;    6. Arrays must have the same dimensions to be combines, and tplot variables must also have the same times.
;    7. Procedures: min,max,mean,median,count,total  all take a second argument that allow you to select the
;       dimension over which the operation is performed
;
; See Also:
;   All routines in the ssl_general/mini directory
;   The techniques used for this interpreter are based on two books:
;
;   1. Compilers:Principles,Techniques,and Tools by Aho,Sethi,& Ullman 1988 (esp. Ch3)
;
;   2. Structure & Interpretation of Computer Programs by Abelson & Sussman 1996 (esp. Ch4)
;
;   If you want to understand/modify this program it may help to use these books as
;   a reference.
;
;   Also see:  thm_crib_calc.pro for examples of usage
;
; ToDo: 1. Implement Constants
;       2. Implement 0 argument functions
;       3. Implement keywords for functions
;       4. Implement procedures
;       5. Implement control statements
;       6. Optional auto-interpolation
;
```



```
; $LastChangedBy: pcruce $
; $LastChangedDate: 2009-07-13 09:43:23 -0700 (Mon, 13 Jul 2009) $
; $LastChangedRevision: 6418 $
; $URL: svn+ssh://thmsvn@ambrosia.ssl.berkeley.edu/repos/ssl_general/trunk/mini/calc.pro $
;-
```

5. 3 Routine Header Guidelines

a. All routine headers should go at the very beginning of the file and describe the main routine (routine from which the file takes its name). Helper functions with sub-headers should follow, with the main routine being the last function/procedure listed in the file.

b. All routine headers should start with the line ";+" and end with the line ";-". All lines between the first and last line should begin with a semicolon(";"). This requirement is necessary to build the html documentation.

c. The four lines prior to the last line of the header should be:

```
; $LastChangedBy: $
; $LastChangedDate: $
; $LastChangedRevision: $
; $URL: $
```

These fields will be filled in automatically by SVN.

d. All headers should contain the procedure name and a description. Sections for inputs, outputs, keywords, notes, usage, examples, references (or see also:) should be added at the discretion of the developer, with the goal of being as thorough as possible.

5.4 Style Guidelines

a. In-line comments are left to the discretion of the developer. Sufficient detail should be provided so that someone who is unfamiliar with the code and has average familiarity with IDL, can understand the goals, methods, and function of a routine.

b. Developers should select names that use consist capitalization and underscoring across all names in a routine. Naming conventions should change only to indicate differences in type.

c. Names should be descriptive, not single letters or nonsense words. Good: (energy_min, mag_var_name, re_in_km,scpot_to_dens) Bad: (e,input1,foo,fx)

5.5 Crib Guidelines:

Cribs should be provided for a selection of useful or high level routines. They are intended to provide the user with an introduction to a routine, demonstrate common usage and pitfalls, and provide a base of code for users to start building off of. Documentation and comments should be detailed and several different usage examples should be provided.