

TOLNet Technical Document (TD-1): TOLNet Profile Data Format Description

Document History:

2013/07/16 Document version 1.0, released for official use by TOLNet

Introduction

The TOLNet data file format was discussed during the last quarter of 2012. Based on the agreed format, prototype input/output subroutines in IDL language were produced in early 2013 for use by the TOLNet data originators and users. The present documentation describes the TOLNet Profile Data Format and reviews the basic steps necessary to write and read the data, illustrated by examples using the prototype IDL routines. A similar (shorter) documentation exists for the TOLNet Surface Data Format. ***The TOLNet format version at the time the present documentation was last updated is v1.0.*** It will be incremented each time a “physical” format change is implemented and updated documentation will be provided accordingly.

General concept

The agreed format was built upon past experience and achievements from various atmospheric observation networks or projects, principally NDACC (Network for the Detection of Atmospheric Composition Change), and ICARTT (International Consortium for Atmospheric Research on Transport and Transformation). It was agreed that a quick, easy-to-access ASCII format had more advantages than a complex binary format. However TOLNet recognized the importance of keeping a strict consistency in data content for both metadata and primary (measured) and ancillary (*a priori*) data. Consequently, it was agreed that an ASCII format containing a unique general header followed by one or multiple blocks, each of which having a similar structure and reporting data and metadata of a specific profile, was the most appropriate choice for TOLNet.

Filename Structure

The TOLNet filename structure of the TOLNet data files is strongly inspired from that of the ICARTT data files. It is of the form: TOLNet-O3Lidar_S_YYYYMMDD_Rv[*c*].XXX, and is broken up as follows:

TOLNet-O3Lidar = Prescribed 14-character string to be used by all TOLNet lidars

S = Free length string for the TOLNet site/station name (PI's choice)

YYYYMMDD = 8-character string for the date of the first profile in file (year, month, day)

R = Prescribed 1-character string referring to the next character (the revision number)

v = 1- or 2-character string for the data revision number (0-99)

c = Optional, free length string used by PI to report special (e.g., non-routine) data

XXX = extension depending on file content type: “dat” for data, “png” or “jpg” for images, etc...

Delimiters

Commas are used to delimit data in the file header and data section. At the end of each line in the header, a semicolon is used followed by a brief text describing the content on that line (see examples)

Profile data file general structure

The overall I/O and formatting rules applying to each TOLNet data file are as follows:

- 1) There may be one and only one data file per day (recorded in Universal Time) for a given TOLNet instrument.
- 2) There may be one or several profiles in each data file
- 3) Each data file must be built of 4 data/metadata blocks, the structure of which is being prescribed by the TOLNet Data Format version. For version 1.0, these blocks are:
 - a. The general header
 - b. The general comments section
 - c. The profile header
 - d. The data section
- 4) Each file must contain a unique “general header” followed by a unique “general comments section”, both located before all the profiles
- 5) The general header contains the following information:
 - a. The file structure (incl. format version and other formatting details)
 - b. The number of lidar profiles reported in the file
- 6) The general comments section contains the following information:
 - a. General meta-data information (e.g., PI/contact info, site name, geo-location, etc.)
 - b. Data revision information, applying to entire file (i.e., not profile-specific)
- 7) Each profile is then reported sequentially in the form of a “profile header” followed by the 2-dimensional (altitudes x variables) data (the “data section”)
- 8) Each profile header contains the following information:
 - a. Number of data lines in profile
 - b. Measurement geo-location
 - c. Measurement conditions
 - d. Data analysis version
 - e. Any other meta-data information pertaining to the profile and deemed critical by the TOLNet community (see details in profile header description)
- 9) The last line of each profile header serves as data column labels, and contains all data variables’ short names separated by commas
- 10) Each profile header is preceded by a separator line starting with: “#BEGIN PROFILE”
- 11) Each profile header is immediately followed by the profile data, reported in the form of multiple lines of floating-point numbers separated by the comma character “,”
- 12) The number of columns in the data section is constrained by the TOLNet format version
- 13) The number of lines and columns in the data section must always be consistent with the values reported in the profile header and general header respectively.

Structure and information content between line 1 and first data line

	Description	Type	Dim.	Prescription	Ref.
Line 1	Number of lines to follow in general header, <u>excluding</u> general comments	Integer	Scalar	Fixed by TOLNet format version	<i>ngh</i>
Line 2	TOLNet format version	String	Scalar	Fixed by TOLNet format version	
Line 3	Number of profiles in file	Integer	Scalar	Set by data originator	<i>nprof</i>
Line 4	Number of data columns for all profiles	Integer	Scalar	Fixed by TOLNet format version	<i>ncol</i>
Lines 5 to 5+ <i>ncol</i> -1	Description of the <i>ncol</i> data variables (see list in TOLNet format version description)	Strings	Scalars	Fixed by TOLNet format version	
Line 5+ <i>ncol</i> (also 1+ <i>ngh</i>)	Missing values used for each column, separated by commas	String	Scalar	Fixed by TOLNet format version	
Line 2+ <i>ngh</i>	Number of lines to follow in general comments section	Integer	Scalar	Fixed by TOLNet format version	<i>ngc</i>
Lines 3+ <i>ngh</i> to 6+ <i>ngh</i>	First 4 general comments lines (see list in TOLNet format version description)	Strings Integers Floats	Scalars	Fixed by TOLNet format version	
Line 7+ <i>ngh</i> 1 st character	Letter "R" ("R" for "revision")	String	Scalar	Fixed by TOLNet format version	
Line 7+ <i>ngh</i> next character(s)	Revision number ("0" if first time data)	Integer	Scalar	Set by data originator	
Lines 8+ <i>ngh</i> to 2+ <i>ngh</i> + <i>ngc</i>	Revision comments lines (mandatory if rev number is not "0")	Strings	Scalars	Set by data originator	
Line 3+ <i>ngh</i> + <i>ngc</i>	Profile separator: must contain "#BEGIN PROFILE"	String	Scalar	Fixed by TOLNet format version	
Line 4+ <i>ngh</i> + <i>ngc</i>	Number of lines to follow in profile header <u>including</u> all comments	Integer	Scalar	TOLNet format and PI	<i>nph</i>
Lines 5+ <i>ngh</i> + <i>ngc</i> to 5+ <i>ngh</i> + <i>ngc</i> + <i>nph</i> -2	Profile header lines (see list of metadata in Profile Header description)	Strings Integers Floats	Scalars	Set by data originator	
Line 5+ <i>ngh</i> + <i>ngc</i> + <i>nph</i> -1	Data variables short names for easy visual identification of data columns (separated by commas)	Strings	Scalars	Fixed by TOLNet format version	
Line 5+ <i>ngh</i> + <i>ngc</i> + <i>nph</i>	First data line (<i>ncol</i> scalars separated by commas)	Floats	Scalars	Set by data originator	

Prescribed data variables for TOLNet Format version v1.0

Number of data columns: 14

	Description	Unit	Short name	Nb. of significant decimals	Missing value
Column 1	Altitude above sea level (center of sampling bin)	m	ALT	<i>f0.1</i>	-9999
Column 2	Ozone number density (measured)	molec.m-3	O3ND	<i>e0.3</i>	-9999
Column 3	Ozone number density standard uncertainty	molec.m-3	O3NDUncert	<i>e0.3</i>	-9999
Column 4	Ozone number density (NDACC-standardized) vertical resolution	m	O3NDResol	<i>f0.1</i>	-9999
Column 5	Measurement Precision	%	Precision	<i>f0.2</i>	-9999
Column 6	Channel range (1.0 for near-field, 1.xx for mix, 2.0 for any farther field, etc...)	undimensional	ChRange	<i>f0.2</i>	-9999
Column 7	Ozone mixing ratio (derived)	ppbv	O3MR	<i>f0.2</i>	-9999
Column 8	Ozone mixing ratio uncertainty	ppbv	O3MRUncert	<i>f0.2</i>	-9999
Column 9	Pressure used for mixing ratio derivation	hPa	Press	<i>e0.3</i>	-9999
Column 10	Pressure uncertainty	hPa	PressUncert	<i>e0.3</i>	-9999
Column 11	Temperature used for mixing ratio derivation	K	Temp	<i>f0.2</i>	-9999
Column 12	Temperature uncertainty	K	TempUncert	<i>f0.2</i>	-9999
Column 13	Air number density used for mixing ratio derivation	molec.m-3	AirND	<i>e0.3</i>	-9999
Column 14	Air number density uncertainty	molec.m-3	AirNDUncert	<i>e0.3</i>	-9999

Prescribed meta-data in General Comments Section for TOLNet Format version v1.0

Number of metadata lines: 5

	Description	Unit	Type
Line <i>ngh</i> +3	Instrument name	N/A	String
Line <i>ngh</i> +4	PI and contact info (separated by commas)	N/A	Strings
Line <i>ngh</i> +5	Site name	N/A	String
Line <i>ngh</i> +6	Longitude, latitude, altitude (separated by commas)	deg. East deg. North m	Floats
Line <i>ngh</i> +7 character(s) following "R"	Revision number	Undimensional	Integer

Prescribed meta-data in Profile Header for TOLNet Format version v1.0

Number of metadata lines: 10

	Description	Unit	Type	Ref.
Line <i>ngh+ngc</i> +5	Number of data lines in profile	Undimensional	Integer	<i>nalt</i>
Line <i>ngh+ngc</i> +6	Data processing date and time	YYYY-MM-DD, HH:MM:SS	Strings	
Line <i>ngh+ngc</i> +7	Data processing software name and version	N/A	String	
Line <i>ngh+ngc</i> +8	Result quality: must be NOMINAL, FAIR or GOOD	N/A	String	
Line <i>ngh+ngc</i> +9	Start date/time of profile	YYYY-MM-DD, HH:MM:SS	Strings	
Line <i>ngh+ngc</i> +10	End date/time of profile	YYYY-MM-DD, HH:MM:SS	Strings	
Line <i>ngh+ngc</i> +11	Weighted mean date/time of profile	YYYY-MM-DD, HH:MM:SS	Strings	
Line <i>ngh+ngc</i> +12	<i>A priori</i> source of pressure, temperature and air number density used for ozone mixing ratio derivation	N/A	String	
Line <i>ngh+ngc</i> +13	<i>A priori</i> source date/time	YYYY-MM-DD, HH:MM:SS	Strings	
Line <i>ngh+ngc</i> +14	<i>A priori</i> source longitude, latitude, altitude (separated by commas)	deg. East deg. North m	Floats	

Optional comments for TOLNet Format version v1.0

Optional comments may be added at two locations:

- 1) Data revision comments may be added at the end of the general comments section, immediately following line $ngh+7$. If data is of revision number "0", there should be no added comment (and therefore no added line). Otherwise, each increment in revision number should be associated with at least one additional comment line, most recent comments being on top of older comments. These comments pertain to the description of the revisions that were made and led to an update of the data files
- 2) At the end of a profile header, comments may be added just before the line listing the short names (line $5+ngh+ngc+nph-1$). These comments pertain to each individual profile being reported

Illustrative example:

An illustrative example is provided in attachment (attached to PDF version of present document), where the data content is color-coded based on its level of prescription:

- 1) Content in red is mandatory, and prescribed by the TOLNet format version (i.e., must appear "as is" in actual files)
- 2) Content in blue is mandatory, and set by the data originator
- 3) Content in green is optional, and set by the data originator (optional comments)

It is the responsibility of the data originator to set the appropriate values of all dynamic formatting elements, in particular:

- 1) The number of profiles in file ($nprof$)
- 2) The number of lines in the general comments section (ngc)
- 3) The number of profile header lines for each profile in the file (nph)
- 4) The number of data lines for each profile in the file ($nalt$)

It is the responsibility of the data originator to verify that all static formatting elements prescribed by the current TOLNet Format version are appropriately set, in particular:

- 1) The number of general header lines (ngh)
- 2) The TOLNet format version
- 3) The number of data columns ($ncol$)

Bonus Tools: How to write and read data files with the provided IDL routines

The provided IDL routines are called “WriteTOLNet.pro” for creating or updating a data file, and “ReadTOLNet.pro” for reading a data file. Every data and metadata variable to be mandatory reported is included in the list of input parameters.

IDL calling syntax to create a new data file:

```
WriteTOLNet, /create, a1, a2, a3, a4, a5, a6, x1, x2, x3, n1 $  
    , a7, a8, a9, a10, a11, a12, a13, a14, a15, a16 $  
    , a17, a18, a19, x4, x5, x6, a20 $  
    , v1, v2, v3, v4, v5, v6, v7, v8, v9, v10, v11, v12, v13, v14 $  
    [, rev_comm=a21] [, prof_comm=a22] $  
    [, file_suffix=a23] [, dir_name=a24] [, ndata=n2]
```

(the dollar “\$” sign is the standard IDL syntax used to continue the same command on the next line, and the brackets shown above simply annotate optional keywords)

One or several profiles may be written at once for each IDL routine call. Regardless of the number of profiles, the metadata passed through the call and meant to be written in the general header + general comments section are always scalars (input parameters a1 through a6, x1 through x3, and n1).

If only one profile is to be written, the meta-data passed through the call and meant to be written in the profile header may either be scalars or vectors of length 1. Each data variable may be either vectors of length *nalt*, or 2D arrays of dimension 1 x *nalt*.

If several profiles are to be written, the metadata meant for the profile header are always vectors of length *nprof* (*nprof*>1). Each data variable is a 2D arrays of dimension *nprof* x *nalt*.

When the number of data points (*nalt*) is different from one profile to another, the keyword “*ndata=n2*” should be used to specify how many data lines should be written for each profile (*n2* is a vector of length *nprof* containing the number of data lines for each profile).

Keywords “*rev_comm=*” (respectively “*prof_comm=*”) should be used only when including revision (respectively profile) comments. Keyword “*file_suffix=*” should be used only for special measurements (e.g., validation campaigns, non-routine measurements), and the string *a23* should be of reasonable length since it will be directly appended to the filename.

IDL calling syntax to append data to an already existing data file:

```
WriteTOLNet, /append, a1, a2, a3, a4, a5, a6, x1, x2, x3, n1 $  
    , a7, a8, a9, a10, a11, a12, a13, a14, a15, a16 $  
    , a17, a18, a19, x4, x5, x6, a20 $  
    , v1, v2, v3, v4, v5, v6, v7, v8, v9, v10, v11, v12, v13, v14 $  
    , rev_comm=a21, prof_comm=a22 $  
    , file_suffix=a23, dir_name=a24, ndata=n2
```

The only difference with the “create” mode is the use of the keyword “/append” instead of “/create”. In “append” mode, the full content of the target data file (i.e., general header, general comments section, profile headers, and data sections) is kept unchanged, except for the reported number of profiles (line 4) and the data revision lines (revision comments), which should reflect the content of the input parameters passed through the call in “append” mode.

List of mandatory input parameters associated with WriteTOLNet.pro:

	Description	Unit	Type	Dimension	Length	Destination	Example
a1	Site ID	/	string	scalar	3*	filename	TMF
a2	Instrument name	/	string	scalar	60*	General Comments	Tropospheric Lidar
a3	PI name	/	string	scalar	60*	General Comments	T. Leblanc
a4	PI organization	/	string	scalar	60*	General Comments	Jet Propulsion Laboratory
a5	PI email	/	string	scalar	60*	General Comments	leblanc@tmf.jpl.nasa.gov
a6	Site name	/	string	scalar	60*	General Comments	Table Mountain Facility
x1	Site longitude	degE	float	scalar	/	General Comments	-117.7
x2	Site latitude	degN	float	scalar	/	General Comments	34.4
x3	Site altitude	m	float	scalar	/	General Comments	2285.
n1	Revision version	/	integer	scalar	/	General Comments	1
a7	Data processing date	/	string	[<i>nprof</i>]	10 [†]	Profile Header(s)	['2013-01-23','2013-01-23']
a8	Data processing time	/	string	[<i>nprof</i>]	8 [†]	Profile Header(s)	['12:34:56','12:40:12']
a9	Data processing version	/	string	[<i>nprof</i>]	48*	Profile Header(s)	['LidAna v1.2','LidAna v1.2']
a10	Data quality	/	string	[<i>nprof</i>]	48*	Profile Header(s)	['POOR','NOMINAL']
a11	Profile start date	/	string	[<i>nprof</i>]	10 [†]	Profile Header(s)	['2013-01-22','2013-01-22']
a12	Profile start time	/	string	[<i>nprof</i>]	8 [†]	Profile Header(s)	['06:12:05','08:40:59']
a13	Profile end date	/	string	[<i>nprof</i>]	10 [†]	Profile Header(s)	['2013-01-22','2013-01-22']
a14	Profile end time	/	string	[<i>nprof</i>]	8 [†]	Profile Header(s)	['08:12:45','09:40:10']
a15	Profile weighted mean date	/	string	[<i>nprof</i>]	10 [†]	Profile Header(s)	['2013-01-22','2013-01-22']
a16	Profile weighted mean time	/	string	[<i>nprof</i>]	8 [†]	Profile Header(s)	['07:12:34','09:10:21']
a17	A priori source name	/	string	[<i>nprof</i>]	10*	Profile Header(s)	['Radiosonde','NCEP']
a18	A priori source date	/	string	[<i>nprof</i>]	10*	Profile Header(s)	['2013-01-22','2013-01-22']

* Recommended max length; † Prescribed, fixed length

List of mandatory input parameters associated with WriteTOLNet.pro (cont.)

	Description	Unit	Type	Dimension	Max length	Destination	Example
a19	A priori source time	/	string	[<i>nprof</i>]	8 [†]	Profile Header(s)	['06:00:12','12:00:34']
x4	A priori source longitude	degE	float	[<i>nprof</i>]	/	Profile Header(s)	-115.0
x5	A priori source latitude	degN	float	[<i>nprof</i>]	/	Profile Header(s)	32.5
x6	A priori source altitude	m	float	[<i>nprof</i>]	/	Profile Header(s)	237.
a20	Operator comments	/	string	[<i>nprof</i>]	48*	Profile Header(s)	['Windy tonight','Clear']
v1	Altitude	m	float	[<i>nprof,nalt</i>]	/	Data col-1	
v2	Ozone number density	molec.m-3	float	[<i>nprof,nalt</i>]	/	Data col-2	
v3	Ozone number density uncertainty	molec.m-3	float	[<i>nprof,nalt</i>]	/	Data col-3	
v4	Vertical resolution	m	float	[<i>nprof,nalt</i>]	/	Data col-4	
v5	Precision	%	float	[<i>nprof,nalt</i>]	/	Data col-5	
v6	Channel code	/	float	[<i>nprof,nalt</i>]	/	Data col-6	
v7	Ozone mixing ratio	ppbv	float	[<i>nprof,nalt</i>]	/	Data col-7	
v8	Ozone mixing ratio uncertainty	ppbv	float	[<i>nprof,nalt</i>]	/	Data col-8	
v9	A priori pressure	hPa	float	[<i>nprof,nalt</i>]	/	Data col-9	
v10	A priori pressure uncertainty	hPa	float	[<i>nprof,nalt</i>]	/	Data col-10	
v11	A priori temperature	hPa	float	[<i>nprof,nalt</i>]	/	Data col-11	
v12	A priori temperature uncertainty	hPa	float	[<i>nprof,nalt</i>]	/	Data col-12	
v13	A priori air number density	hPa	float	[<i>nprof,nalt</i>]	/	Data col-13	
v14	A priori air number density uncertainty	hPa	float	[<i>nprof,nalt</i>]	/	Data col-14	

* Recommended max length; † Prescribed, fixed length

In the case of multiple profiles, the vertical dimension length “*nalt*” must always be set to the maximum length found among all profiles to ensure that all the data is properly transferred to the file (IDL works with rectangular arrays). The proper number of (valid) altitudes to be actually written in the file is transferred through the “*ndata*” keyword.

List of optional input keywords associated with *WriteTOLNet.pro*:

	Description	Unit	Type	Dimension	Max length	Destination	Example
a21	Revision comments	/	string	scalar	60*	General comments section	'Revised data due to bug in v0'
a22	Profile comments	/	string	[<i>nprof</i>]	48*	Profile Header(s)	['Pb. with analysis','Used NCEP']
a23	File suffix	/	string	scalar	free	Filename	'AJAX_Campaign'
a24	Output directory name	/	string	scalar	free	Filename	'C:\TOLNet_Data\2013'
n2	Number of data lines per profile	/	integer	[<i>nprof</i>]	/	Data section	[234,321]

IDL calling syntax to read a TOLNet data file:

```
ReadTOLNet, filename , a1, a2, a3, a4, a5, a6, x1, x2, x3, n1 $
, a7, a8, a9, a10, a11, a12, a13, a14, a15, a16 $
, a17, a18, a19, x4, x5, x6, a20 $
, v1, v2, v3, v4, v5, v6, v7, v8, v9, v10, v11, v12, v13, v14 $
[, rev_comm=a21] [, prof_comm=a22] [, ndata=n2] [, extract=n3]
```

In “create” or “append” mode, there were only input parameters and keywords, i.e., no returned variables. In the present case, all input parameters of the “WriteTOLNet.pro” routine have now become the returned variables. The only input parameter is the filename to read. Unlike “WriteTOLNet.pro” there is no “/create” or “/append” keyword. Instead there is an “extract” keyword that allows to extract a subset of the profiles contained in the file. For example, if the file contains 10 profiles and one wants only profiles #4 and #6, then the keyword syntax “extract=[4,6]” can be used.

Also, there is no “file_suffix” or “dir_name” keywords like in the “WriteTOLNet.pro routine because the complete filename (including path) is included in the one input parameter “filename”.

When multiple profiles are returned, the size of the altitude dimension is always set to the maximum size found among all profiles. The correct number of altitudes for each profile is then set by using the keyword “ndata”.

IDL routine for the definition of the TOLNet Format template: *TOLNetFormat.pro*:

This routine is called by “WriteTOLNet.pro”. Its role is to build up the proper TOLNet format template. The input parameter “version” is currently not used since only one TOLNet format exists (v1.0). It will be used in the future when format versions change. This routine should not be modified for any reason without prior consultation with the TOLNet group. Any change to this routine may result in a possible alteration of the format template, increasing the risk for non-compliance.