

**GEOMS compliant HDF files for FTIR
Changes, Versioning + Discussion
Ian Boyd (BCSC) and Bavo Langerock (BIRA)**

2019 Joint NDACC-IRWG and TCCON Meeting 20th to 24th May, 2019 Wanaka New Zealand

Aims

- *What is GEOMS?*
- *Data Versioning changes*
- *Discussion on proposed update of the reporting template*

What is GEOMS?

GEOMS – Generic Earth Observation Metadata Standard is a metadata standard used for archiving data from ground-based networks, like NDACC, so it can be used for the validation of NASA and ESA satellite data

Facilitate portability and accessibility of data and make contents self-describing

Documentation and Tools – available on the AVDC and EVDC web-sites

<https://avdc.gsfc.nasa.gov/> and <http://evdc.esa.int/>

Tools – include IDL programs to create and read GEOMS-compliant files, and to convert between HDF4, HDF5 and netCDF

On-line Tools

- QA and template checker on the AVDC web-site (use before submitting to NDACC) <https://avdc.gsfc.nasa.gov/index.php?site=1829327959>
- GEOMS file creation tool to be available on the EVDC web-site – input data as single column ASCII file(s) and the on-line converter will create GEOMS-compliant files

Data Versioning

GOAL – Allow a clear way to delineate datasets from a single instrument with different processing algorithms.

This will mean, for example that you can have two or more different processing versions from the same instrument on NDACC at the same time e.g.

groundbased_ftir.co_bira.iasb003_standard_....

vs.

groundbased_ftir.co_bira.iasb003_stratomeso_....

groundbased_ftir.co_bira.iasb003_velazco2007_....?

HOW – add an optional third field to the *DATA_SOURCE* global attribute

*This field will then directly identify a particular variant of a dataset and will mean any files with that name can be treated as a continuous time-series. The *DATA_FILE_VERSION* now points to a sub-version within that specific data/algorithm variant. It refers to small/minor changes e.g. fixing an error in previously submitted data*

Data Versioning

GUIDELINES:

- *Working Groups work with providers to determine suitable variant names*
- *Standardized names vs. Flexibility*
 - *Standard names associated with common applications e.g. CALVAL, INHOUSE, HIRES, WEEKLY*
 - *Other names will be specific to a particular instrument type or processor, so may only be used by a single group or limited number of groups*
- *Database of version names to be maintained by NDACC and GEOMS to avoid possible conflicts*

IMPLEMENTATION: *The new Data Versioning rules will be presented at the next NDACC SC meeting (October) for approval, with the aim to put the guidelines in place at that time*

Why update the template?

- *Add value/information (e.g. source variables)*
- *Current template is not GEOMS compliant*
- *Clearer definitions of variables*

Timeline

- *Coincide with the introduction of the new 'Data Versioning' rules*
- *Combine with any other changes proposed by the WG (HITRAN 2016)*
- *October/November 2019 (after NDACC SC meeting)*

Proposed FTIR Template

DATETIME	DATETIME	MJD2K	DOUBLE	Mean of the zero path difference times of the individual scans that were averaged to produce the fitted spectrum	x
DATETIME.START	DATETIME	MJD2K	DOUBLE	Start time of the measurement	x
DATETIME.STOP	DATETIME	MJD2K	DOUBLE	End time of the measurement	x
INTEGRATION.TIME	DATETIME	s	REAL	Duration of the measurement corresponding to the retrieved datapoint	x
LATITUDE.INSTRUMENT	CONSTANT	deg	REAL	Inst. geolocation (+ for north; - for south)	x
LONGITUDE.INSTRUMENT	CONSTANT	deg	REAL	Inst. geolocation (+ for east; - for west)	x
ALTITUDE.INSTRUMENT	CONSTANT	m	REAL	Inst. Geolocation	x
SURFACE.PRESSURE_INDEPENDENT	DATETIME	hPa	REAL	Surface/ground pressure	x
SURFACE.PRESSURE_INDEPENDENT_SOURCE	[CONSTANT DATETIME]	[empty]	STRING	Surface/ground pressure profile source (e.g. Mercury barometer etc.) if SURFACE.PRESSURE_INDEPENDENT is provided	x
SURFACE.TEMPERATURE_INDEPENDENT	DATETIME	K	REAL	Surface/ground temperature	x
SURFACE.TEMPERATURE_INDEPENDENT_SOURCE	[CONSTANT DATETIME]	[empty]	STRING	Surface temperature profile source (e.g. Meteorological thermometer etc.) if SURFACE.TEMPERATURE_INDEPENDENT is provided	x
HUMIDITY.RELATIVE.SURFACE_INDEPENDENT	DATETIME	%	REAL	Relative humidity at the station	0
HUMIDITY.RELATIVE.SURFACE_INDEPENDENT_SOURCE	[CONSTANT DATETIME]	[empty]	STRING	Relative humidity source (e.g. Anemometer; ECMWF etc.) if HUMIDITY.RELATIVE.SURFACE_INDEPENDENT is provided	x
WIND.DIRECTION.SURFACE_INDEPENDENT	DATETIME	deg	REAL	Wind direction at the station using WMO definition (wind from the north is 360; from the east is 90 and so on no wind (calm) is 0)	0
WIND.DIRECTION.SURFACE_INDEPENDENT_SOURCE	[CONSTANT DATETIME]	[empty]	STRING	Wind direction source (e.g. Anemometer; ECMWF etc.) if WIND.DIRECTION.SURFACE_INDEPENDENT is provided	x
WIND.SPEED.SURFACE_INDEPENDENT	DATETIME	m s-1	REAL	Wind speed at the station	0
WIND.SPEED.SURFACE_INDEPENDENT_SOURCE	[CONSTANT DATETIME]	[empty]	STRING	Wind speed source (e.g. Anemometer; ECMWF etc.) if WIND.SPEED.SURFACE_INDEPENDENT is provided	x
ALTITUDE	[ALTITUDE DATETIME;ALTITUDE]	km	REAL	Retrieval effective altitude vector (if DATETIME independent) or matrix (if DATETIME dependent). Values are monotonically increasing	x
ALTITUDE.BOUNDARIES	[ALTITUDE;INDEPENDENT DATETIME;ALTITUDE;INDEPENDENT]	km	REAL2D	(if DATETIME independent) or 3D (if DATETIME dependent) matrix with on each row the lower and upper boundaries of the layers for which the partial columns are reported. In layer-based retrieval these are equal to the lower and upper boundaries of the respective retrieval layers	x
PRESSURE_INDEPENDENT	DATETIME;ALTITUDE	hPa	REAL	Effective air pressure at each altitude	x
PRESSURE_INDEPENDENT_SOURCE	[CONSTANT DATETIME DATETIME;ALTITUDE]	[empty]	STRING	Pressure profile source (e.g. Lidar; NCEP; Sonde; ECMWF etc.)	x
TEMPERATURE_INDEPENDENT	DATETIME;ALTITUDE	K	REAL	Effective air temperature at each altitude	x
TEMPERATURE_INDEPENDENT_SOURCE	[CONSTANT DATETIME DATETIME;ALTITUDE]	[empty]	STRING	Temperature profile source (e.g. NCEP; Sonde; ECMWF etc.)	x
DRY.AIR.COLUMN.PARTIAL_INDEPENDENT	DATETIME;ALTITUDE	Zmolec cm-2	REAL	Vertical profile of partial columns of air number densities (for conversion between VMR and partial column profile)	x
DRY.AIR.COLUMN.PARTIAL_INDEPENDENT_SOURCE	[CONSTANT DATETIME DATETIME;ALTITUDE]	[empty]	STRING	Partial columns of air source (e.g. NCEP etc)	x

Proposed FTIR Template

* Reporting of surface pressure and temperature mandatory (IRWG 190523)

Only temperature and pressure included in v2.
 * All optional. If reported then it is mandatory to report the source (note: optional dependencies). If reporting surface wind both speed and direction must be included

Variable	Units	Real	Description	Optional
MEAN_ZERO_PATH_DIFFERENCE	Mean of the zero path difference times of the individual scans that were averaged to produce the fitted spectrum	DOUBLE		x
SURFACE.PRESSURE_INDEPENDENT				x
SURFACE.PRESSURE_INDEPENDENT_SOURCE				x
SURFACE.TEMPERATURE_INDEPENDENT				x
SURFACE.TEMPERATURE_INDEPENDENT_SOURCE				x
HUMIDITY.RELATIVE.SURFACE_INDEPENDENT				o
HUMIDITY.RELATIVE.SURFACE_INDEPENDENT_SOURCE				x
WIND.DIRECTION.SURFACE_INDEPENDENT				o
WIND.DIRECTION.SURFACE_INDEPENDENT_SOURCE				x
WIND.SPEED.SURFACE_INDEPENDENT				o
WIND.SPEED.SURFACE_INDEPENDENT_SOURCE				x

- SURFACE.PRESSURE_INDEPENDENT
- SURFACE.PRESSURE_INDEPENDENT_SOURCE
- SURFACE.TEMPERATURE_INDEPENDENT
- SURFACE.TEMPERATURE_INDEPENDENT_SOURCE
- HUMIDITY.RELATIVE.SURFACE_INDEPENDENT
- HUMIDITY.RELATIVE.SURFACE_INDEPENDENT_SOURCE
- WIND.DIRECTION.SURFACE_INDEPENDENT
- WIND.DIRECTION.SURFACE_INDEPENDENT_SOURCE
- WIND.SPEED.SURFACE_INDEPENDENT
- WIND.SPEED.SURFACE_INDEPENDENT_SOURCE

Proposed FTIR Template

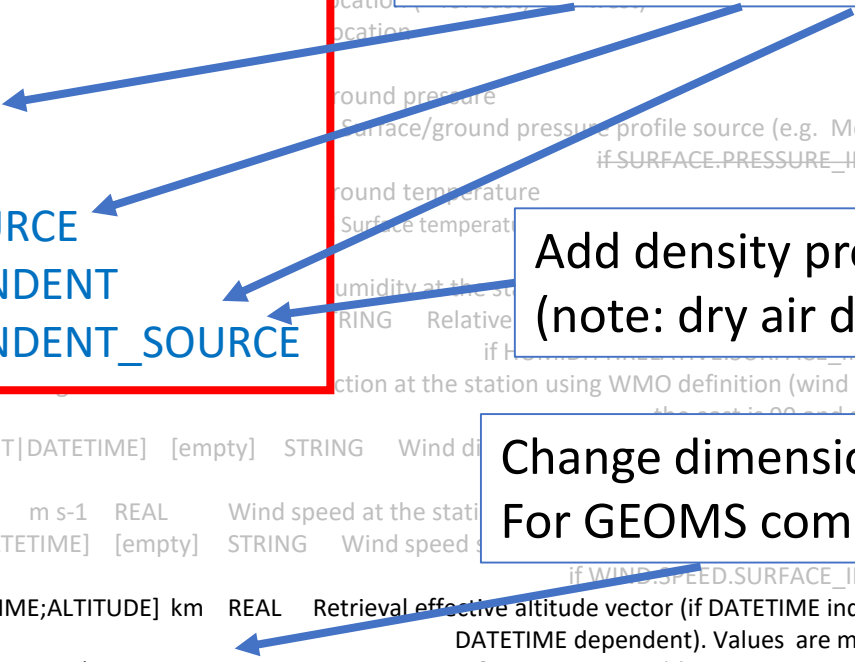
DATETIME	DATETIME	MJD2K	DOUBLE	Mean of the zero path difference times of the individual scans that were averaged to produce the fitted spectrum	
DATETIME.START	DATETIME	MJD2K	DOUBLE	Start time of the	
ALTITUDE	[ALTITUDE DATETIME;ALTITUDE]	km	REAL	Retrieval effective altitude vector (if DATETIME independent) or matrix (if DATETIME dependent). Values are monotonically increasing	
ALTITUDE.BOUNDARIES	[ALTITUDE;INDEPENDENT DATETIME;ALTITUDE;INDEPENDENT]	km	REAL2D (if DATETIME independent) or 3D (if DATETIME dependent)	matrix with on each row the lower and upper boundaries of the layers for which the partial columns are reported. In layer-based retrieval these are equal to the lower and upper boundaries of the respective retrieval layers	
PRESSURE_INDEPENDENT	DATETIME;ALTITUDE	hPa	REAL	Effective air pressure at each altitude	
PRESSURE_INDEPENDENT_SOURCE	[CONSTANT DATETIME DATETIME;ALTITUDE]	[empty]	STRING	Pressure profile source (e.g. Lidar; NCEP; Sonde; ECMWF etc.)	
TEMPERATURE_INDEPENDENT	DATETIME;ALTITUDE	K	REAL	Effective air temperature at each altitude	
TEMPERATURE_INDEPENDENT_SOURCE	[CONSTANT DATETIME DATETIME;ALTITUDE]	[empty]	STRING	Temperature profile source (e.g. NCEP; Sonde; ECMWF etc.)	
DRY.AIR.COLUMN.PARTIAL_INDEPENDENT	DATETIME;ALTITUDE	Zmolec	cm-2	REAL	Vertical profile of partial columns of air number densities (for conversion between VMR and partial column profile)
DRY.AIR.COLUMN.PARTIAL_INDEPENDENT_SOURCE	[CONSTANT DATETIME DATETIME;ALTITUDE]	[empty]	STRING	Partial columns of air source (e.g. NCEP etc)	

ALTITUDE
ALTITUDE.BOUNDARIES
PRESSURE_INDEPENDENT
PRESSURE_INDEPENDENT_SOURCE
TEMPERATURE_INDEPENDENT
TEMPERATURE_INDEPENDENT_SOURCE
DRY.AIR.COLUMN.PARTIAL_INDEPENDENT
DRY.AIR.COLUMN.PARTIAL_INDEPENDENT_SOURCE

Add source variables
(note: optional dependencies)

Add density profile
(note: dry air designation)

Change dimension ordering
For GEOMS compliance



Proposed FTIR Template

[GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]	DATETIME;ALTITUDE	[ppmv ppbv pptv]	REAL	o
[GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]_APRIORI	DATETIME;ALTITUDE	[ppmv ppbv pptv]	REAL	x
	if [GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR] is provided			
[GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]_AVK	DATETIME;ALTITUDE;ALTITUDE	1	REAL	x
	if [GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR] is provided			
[GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.RANDOM.COVARIANCE	DATETIME;ALTITUDE;ALTITUDE	[ppmv2 ppbv2 pptv2]	REAL	x
	x if [GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR] is provided			
[GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.SYSTEMATIC.COVARIANCE	DATETIME;ALTITUDE;ALTITUDE	[ppmv2 ppbv2 pptv2]	REAL	x
	x if [GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR] is provided			
[GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.COMBINED.COVARIANCE	DATETIME;ALTITUDE;ALTITUDE	[ppmv2 ppbv2 pptv2]	REAL	x
	x if [GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR] is provided			
[GAS].COLUMN.PARTIAL_ABSORPTION.[SOLAR LUNAR]	DATETIME;ALTITUDE	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
	x if [GAS].MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR] is provided			
[GAS].COLUMN.PARTIAL_ABSORPTION.[SOLAR LUNAR]_APRIORI	DATETIME;ALTITUDE	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]	DATETIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_APRIORI	DATETIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_AVK	DATETIME;ALTITUDE	1	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.RANDOM.STANDARD	DATETIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.SYSTEMATIC.STANDARD	DATETIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.COMBINED.STANDARD	DATETIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
ANGLE.[SOLAR LUNAR]_ZENITH.ASTRONOMICAL	DATETIME	deg	REAL	x
ANGLE.[SOLAR LUNAR]_AZIMUTH	DATETIME	deg	REAL	x
LATITUDE	[DATETIME DATETIME;ALTITUDE]	deg	REAL	o
LONGITUDE	[DATETIME DATETIME;ALTITUDE]	deg	REAL	o
H2O.MIXING.RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]	DATETIME;ALTITUDE	[ppmv ppbv pptv]	REAL	x
H2O.COLUMN_ABSORPTION.[SOLAR LUNAR]	DATETIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
SOURCE.PRODUCT	INDEPENDENT	[empty]	STRING	o

Original archived FTIR file name in form Original_Archive;Original_FileName;Original_File_Generation_Date

Proposed FTIR Template

[GAS].COLUMN.PARTIAL_ABSORPTION.[SOLAR|LUNAR]_APRIORI
[GAS].COLUMN_ABSORPTION.[SOLAR|LUNAR]
[GAS].COLUMN_ABSORPTION.[SOLAR|LUNAR]_APRIORI
[GAS].COLUMN_ABSORPTION.[SOLAR|LUNAR]_AVK
[GAS].COLUMN_ABSORPTION.[SOLAR|LUNAR]_UNCERTAINTY.RANDOM.STANDARD
[GAS].COLUMN_ABSORPTION.[SOLAR|LUNAR]_UNCERTAINTY.SYSTEMATIC.STANDARD
~~[GAS].COLUMN_ABSORPTION.[SOLAR|LUNAR]_UNCERTAINTY.COMBINED.STANDARD~~

REAL o
REAL x
provided
REAL x
provided
pvtv2] REAL
is provided
[pptv2] REAL
is provided
[pptv2] REAL
is provided
cm-2) REAL

x if [GAS].MIXING_RATIO.VOLUME.DRY_ABSORPTION.[SOLAR|LUNAR] is provided

[GAS].COLUMN.PARTIAL_ABSORPTION.[SOLAR LUNAR]_APRIORI	DATE TIME;ALTITUDE	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]	DATE TIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_APRIORI	DATE TIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_AVK	DATE TIME;ALTITUDE	1	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.RANDOM.STANDARD	DATE TIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.SYSTEMATIC.STANDARD	DATE TIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
[GAS].COLUMN_ABSORPTION.[SOLAR LUNAR]_UNCERTAINTY.COMBINED.STANDARD	DATE TIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x

Add (IRWG 190523)
~~UNCERTAINTY_COMBINED_STANDARD~~

Change from conditional mandatory to mandatory

ANGLE.[SOLAR LUNAR]_ZENITH		deg	REAL	x
ANGLE.[SOLAR LUNAR]_AZIMUTH		deg	REAL	x
LATITUDE		ALTITUDE	REAL	o
LONGITUDE	[DATE TIME DATE TIME;ALTITUDE]	deg	REAL	o
H2O.MIXING_RATIO.VOLUME.DRY_ABSORPTION.[SOLAR LUNAR]	DATE TIME;ALTITUDE	[ppmv ppbv pptv]	REAL	x
H2O.COLUMN_ABSORPTION.[SOLAR LUNAR]	DATE TIME	molec cm-2 (or a unit scaled by 1E3*n ex. Pmolec cm-2)	REAL	x
SOURCE.PRODUCT	INDEPENDENT	[empty]	STRING	o

Original archived FTIR file name in form Original_Archive;Original_Filename;Original_File_Generation_Date

Example of changes required to become compliant with the new template

- ~~• Add DATETIME.START and DATETIME.STOP variables~~
- Change the units for ALTITUDE.INSTRUMENT to m
- Change the ALTITUDE.BOUNDARIES dimension ordering
- Add an air density profile and source
(DRY.AIR.COLUMN.PARTIAL_INDEPENDENT[_SOURCE])
- Add the source of the pressure and temperature profiles, and surface pressure and temperature fields
- ~~• Add combined uncertainty for the measurement gas =
 $\text{sqrt}((\text{random uncertainty})^2 + (\text{systematic uncertainty})^2)$~~

Final Comments

- Ensure the FILE_META_VERSION value aligns with the TAV version that contains the metadata definitions used in the HDF file (this will need to be done when changing to the new reporting template)
- Writing STRING datasets to an HDF4 file is not straightforward. [idlcr8hdf](#) is set up to do this. If you have problems with this, or anything else GEOMS related then ...

Contact Ian:

iboyd@astro.umass.edu