INTERMAGNET Technical Note

Title: INTERMAGNET CDF data format - ImagCDF

Document number: 8	**Version number:** 1.0	**Creation date:** 12/09/2017

Related documents: *Technical Note 6 – INTERMAGNET Definitive One Second Data Standard*

Keywords: Geomagnetic data; data format

Lead author: Simon Flower	**Contributors:**

Purpose of document: Describe a standard for a new geomagnetic data format

Terms of Reference

1. Initially motivated by the need to store 1-second data in higher precision than is possible using the IAGA-2002 format
2. This format has been designed to be usable with any geomagnetic time series data
3. It may also be useful as an archive storage format

Date due:	**Date submitted:** **Date revised:**

Outcomes: *ImagCDF adopted as the format used for 1-second definitive data from 2015 onwards.*

Note for information

INTERMAGNET Technical Notes are designed to record information on how to accomplish technical tasks. They will often describe alternative approaches, and make recommendations on the practice to be adopted by INTERMAGNET. In many cases a Technical Note will be used as the basis for an entry in the Technical Manual.

The Secretary of the Operations Committee assigns the Document Number, maintains the metadata on the cover sheet, and keeps an index of all documents. Documents to be archived and indexed will be given integer version numbers. Working documents will have fractional version numbers. Only the lead author may edit the main text.

This document describes how NASA's Common Data Format (CDF- http://cdf.gsfc.nasa.gov/) will be used to store geomagnetic data. This format is called ImagCDF. Version 1.2 of ImagCDF is described here.

1. Design details and CDF concepts

1.1 General design details

Geomagnetic data is held in CDF variables, one variable per geomagnetic element. Additional variables hold time stamp data. Each variable has one or more records, an individual record holding a single sample value. The index numbers of the records in the geomagnetic data variables correspond with the index numbers in the relevant time stamp variable's records. A typical file might contain 6 variables: 3 geomagnetic elements (such as HDZ or XYZ); one scalar element (F); and two timing variables, one for the vector data, the other for the scalar (if the scalar data has a different sample rate to the vector data). The format does not mandate that these variables are present – there may be fewer (for example only scalar data) or more (for example additional temperature data).

Metadata is held in CDF attributes in two ways: global attribute entries concern all the data in a file; variable attributes have entries that concern a single variable (e.g. the 'H' variable). An entry holds an individual item of metadata. An attribute name (whether global or variable) must be unique, so when an attribute needs to be used more than once (e.g. the element type for a geomagnetic variable must be used once for each geomagnetic element), then multiple entries are created in a single attribute. Global attributes in ImagCDF will have only one entry (with two exceptions described below). Variable attributes will have an entry for each of the relevant variables in the file. For variable attributes, the multiple entries are indexed using the variable's numeric identifier, so that the metadata 'belongs' to the variable. For example, in an ImagCDF file holding HDZF data, the attribute *Units* will have four entries (in this order): "nT"; "Degrees of arc"; "nT"; "nT".

1.2 Data types used for variables and attributes

1.2.1 Real Numbers

ImagCDF uses double precision CDF_DOUBLE (8-byte) floating point numbers (to the IEEE 754 standard) for all numeric values. These numbers provide about 14.5 (decimal) digits of precision - http://flint.cs.yale.edu/cs422/doc/art-of-asm/pdf/CH14.PDF . In order to use floating point numbers successfully, the dynamic range of the quantity being represented (ratio of smallest to largest value) must be smaller than 14.5 digits. For geomagnetic field strengths, assume that the ratio of smallest difference to largest value that we need to represent in geomagnetic field data is 0.1pT in 80,000nT. This equates to 1 part in 80,000 x 1,000 x 10, or 1 part in 800,000,000, or 9 digits of precision, so is within the dynamic range available.

1.2.2 Dates / times

All date / time values in ImagCDF are held as CDF TT2000 dates, which are based on 8-byte integers. TT2000 uses an epoch (midday on 1st January 2000) to store dates and times, has a precision of 1 nanosecond which gives a range in excess of ±280 years from the epoch date. The TT2000 type can correctly handle leap seconds.

1.2.3 Strings

Text metadata is held using CDF_CHAR data type in ImagCDF.

1.3 Compression

The CDF library allows the user to specify whether all, part or none of a CDF file is compressed at the time it is written. Once this choice has been made, the CDF library handles compression and decompression of the data in the file automatically (reading and any further writing to the file will decompress or compress as required). Choosing whether or not to compress a file is simply a matter of specifying which (if any) compression method to use when the file is created. ImagCDF allows any of the compressions provided by CDF.

Significant compression is achieved using the CDF compression option. A day file of four element minute data can occupy under 15Kb.

2. Where to obtain the CDF software

Before you can use any of NASA or INTERMAGNET's tools for working with ImagCDF, you will need to download and install the CDF software from NASA: http://cdf.gsfc.nasa.gov/html/sw and docs.html. Software that has been written to work with CDF will need the libraries that are installed. For details of other software that may be useful see section 6.

3. ImagCDF data

All variables holding geomagnetic data have the following features:

- Units used must be nT for geomagnetic field values, degrees for angles or celsius for temperatures.
- Lengths of time series are arbitrary (e.g. a file may be used to store an entire day of data or a small fragment of a day down to a single sample).

Geomagnetic data is held in variables called *GeomagneticField<E>* where <E> represents the code for the geomagnetic element recorded - see section 4.6 for a list of valid codes. The variable has 0 dimensions, each consecutive record holding individual consecutive data samples in CDF double data type, starting at record 1. Missing data values are represented by a data sample that contains the same number as is present in the *FILLVAL* metadata attribute. The value of FILLVAL must exceed any valid geomagnetic field strength or angle. Typically this would be 99999.0. The variable attribute *FIELDNAM* must be set to "Geomagnetic Field Element <E>".

An ImagCDF file must include a set of geomagnetic field variables that describe the vector field in a recognised orientation (such as 'HDZ', 'XYZ' or 'DIF). All vector variables must have the same number of records. An ImagCDF file may also include an additional field element from an independent scalar instrument, with an element code of 'G' or 'S'.

Temperature data is optional (unless otherwise stated in the data standard that the data conforms to, for example, temperature is mandatory for data that conforms to the INTERMAGNET 1-second data standard). The first temperature variable is called *Temperature1*, then *Temperature2* and so on. The *FIELDNAM* attribute describes, in free text, for each temperature variable, the location at which the temperature was recorded.

Time stamps for the data are held in separate variables from the data. Time stamps must represent a regular time series with no missing values in the series. Time stamp variables are referenced from data variables using the DEPEND_0 metadata attribute entry for the data variable. All geomagnetic field vector variables must reference the same time stamp variable with their DEPEND_0 attribute. The scalar data variable, if present, may reference the same time stamp variable as the vector data if the sample rates and data length are the same, otherwise it must reference a different time stamp variable that contains the correct time stamps for the scalar data. Similarly, temperature data may reference an existing time stamp variable if the sample rate and data length is correct, otherwise they must reference a separate time stamp variable with the correct time stamps for the temperature data. All time stamp variables have 0 dimensions and must have the same number of records as the data variables that they apply to. Each record in a time stamp variable holds a CDF TT2000 epoch time. Time stamps must always refer to the start of each sample period (e.g. for minute data, the seconds and milliseconds will always be set to zero).

Recommended names for time stamp variables are:

Situation	**Names**
The same time stamps can be used for all data in the file (i.e. there is a single time stamp variable in the file)	DataTimes
Different time stamps for vector, scalar and temperature data	*GeomagneticVectorTimes*, *GeomagneticScalarTimes, Temperature1Times, Temperature2Times, …*

Additional variables and metadata may be carried in an ImagCDF (e.g. it may be convenient to include meteorological data in the same file). The format of these variables and metadata is left to the user to define. The CDF system means that these variables can be included without causing problems to software reading ImagCDF files. Software writing ImagCDF files should preserve additional variables and attributes read from an ImagCDF file.

4. ImagCDF "global" attributes

The following attributes apply to all the data in an ImagCDF file. The "Entries" column shows whether the attribute has:

- A single mandatory entry — Number of entries is exactly 1
- A single optional entry — Number of entries may be 0 or 1
- Multiple mandatory entries — Number of entries is between 1 and N
- Optional mandatory entries — Number of entries is between 0 and N

Superscript numbers following the attribute name show:

1. That the attribute is a recommended attribute for use with NASA's CDF tools
2. That the attribute is part of the ISTP/IACG guidelines - http://spdf.gsfc.nasa.gov/istp_guide/gattributes.html

4.1 Attributes that describe the data format

These are 'constant' values that will be the same for all ImagCDF files. They allow 'generic' CDF programs to understand and process the data correctly.

Attribute Name	Type	Entries	Description
FormatDescription	String	1	Always set to "INTERMAGNET CDF Format"
FormatVersion	String	1	Set to the current version of the format – "1.2"
Title[1]	String	1	Always set to "Geomagnetic time series data"

4.2 Attributes that uniquely identify the data

The attributes in this section are sufficient, along with the start date and duration of the time series, to uniquely identify a piece of geomagnetic data.

Attribute Name	Type	Entries	Description
IagaCode	String	1	The IAGA code for the observatory
ElementsRecorded	String	1	A string consisting of single character codes, each describing one of the geomagnetic field elements that is recorded in this data file. This might typically be a three or four digit code such as HDZ, XYZG or DIFG. The codes in this attribute determine the names of the data variables (see the section on geomagnetic data). Valid codes are defined in section 4.6
PublicationLevel	String	1	Choose one of the following codes to describe the level that the data has been processed to: • *1*: The data is unprocessed and as recorded at the observatory with no changes made.

			• *2*: Some edits have been made such as gap filling and spike removal and possibly a preliminary baseline added. • *3*: The data is at the level required for production of an initial bulletin or for quasi-definitive publication. • *4*: The data has been finalised and no further changes are intended Only these values are allowed. This field provides a quick description of the point the data has reached in the publication process. For detailed information on the standards that the data conforms to see section 4.4.
PublicationDate	Date/time	1	Date and time on which the data was published. This attribute is used to distinguish multiple publications of the same data.

4.3 Attributes that describe the observatory

These attributes are available from other metadata systems (given an IAGA code), but are included for convenience of the user.

Attribute Name	Type	Entries	Description
ObservatoryName	String	1	The full name for the observatory
Latitude	Double	1	The latitude of the observing position in degrees
Longitude	Double	1	The longitude of the observing position in degrees
Elevation	Double	1	The height of the observing position in metres above sea level. Set to 99999.0 if not known.
Institution	String	1 or more	The name of the responsible institute.
VectorSensOrient	String	0 - 1	The orientation code of the sensor at the original recording of the vector data. A string consisting of single character codes, each describing one of the geomagnetic field elements that was recorded by the vector instrument. Valid codes are the same as for the ElementsRecorded attribute.

4.4 Attributes that relate to data standards and quality

These attributes describe the standards, if any, that the data meets.

Attribute Name	Type	Entries	Description
StandardLevel	String	1	Describe whether the data conforms to a standard. Choose from one of the following codes: • *None*: The data does not conform to any standards. When using this, the *StandardName* attribute does not need to be set.

			• *Partial*: The data partially conforms to the relevant standard for this data product. • *Full*: The data fully conforms to the relevant standard for this data product. Only these values are allowed. If *StandardsLevel* is set to *Partial*, then the *PartialStandDesc* attribute must also be set.
StandardName	String	0 - 1	The name of the relevant standard. See section 4.7 for a description of how to use this attribute.
StandardVersion	String	0 – 1	If the standard has a version, put the version number of the standard in this attribute.
PartialStandDesc	String	0 - 1	See section 4.7 for a description of how to use this attribute.

4.5 Attributes that relate to publication of the data

These attributes are needed when that data is published.

Attribute Name	**Type**	**Entries**	**Description**
Source	String	1	Set to one of: “institute” (if the named institution provided the data- see section 4.3 for the institution); “INTERMAGNET” (if the data file has been created by INTERMAGNET from another data source); “WDC” (if the World Data Centre has created the file from another data source)
TermsOfUse	String	0 – 1	The terms of use for the data. This could be text describing the terms of use or a link to a web page. INTERMAGNET has a recommended wording for data provided through INTERMAGNET.
UniqueIdentifier	String	0 – 1	A string that can be used to uniquely identify this data. This could be a Digital object identifier or could be an identifier created according to local rules. Note this is optional and will not be present if this data is not covered by an identifier.
ParentIdentifiers	String	0 or more	The unique identifiers of the parent data sets (if any), one identifier per entry. The parent data set is the data set from which this data set’s values have been derived. E.g. if a one minute data set has been created from a one second data set, the one second data set is the parent. Another example is where definitive data is created from provisional data.
ReferenceLinks	String/ URL	0 or more	URLs pointing to (e.g.) information about the data creator, information about the data repository… One URL per entry.

4.6 Valid codes for elements recorded

- ‘X’, ‘Y’, or ‘Z’ indicate that the corresponding geomagnetic variable holds the strength of the magnetic field vector in the standard geographic coordinates in nT.
- ‘H’ indicates that the variable holds the strength of the magnetic field vector in the horizontal plane along the magnetic meridian in nT.

- D' indicates that the variable holds the angle between the magnetic vector and true north, in degrees of arc, positive east.
- 'E' indicates that the variable holds a field strength in the horizontal plane perpendicular to 'H' in nT. 'E' is only valid for data that is not baseline corrected.
- 'V' indicates that the variable holds the field strength along the direction of the inclination.
- 'I' indicates that the variable holds the angle between the magnetic vector and the horizontal plane, in degrees of arc, positive below the horizontal.
- 'F' indicates that the variable holds the geomagnetic field strength in nT, calculated from and consistent with XYZ or HDZ field elements.
- 'S' indicates that the variable holds the geomagnetic field strength in nT, measured by an independent scalar instrument
- 'G' indicates that the variable holds delta- F values, defined as F(vector) –S(scalar) in nT. When calculating values for the G element, if F(vector) is missing, G is set to –S (scalar)

Other codes are allowed, but may lead to data not being understood.

4.7 Relevant data standards

Different geomagnetic data products have different standards associated with them. This table shows what standards are being referred to in the *StandardsLevel* attribute and describes what to put into the *StandardName* attribute in the case where *StandardsLevel* is set to *Partial* or *Full*.

Data product	**Relevant Standard (see below for references)**	**Contents of *StandardName* attribute**
One second definitive data	INTERMAGNET Definitive One-second Data Standard[1.]	*INTERMAGNET_1-Second*
One minute definitive data	INTERMAGNET magnetic observatory – specifications[2].	*INTERMAGNET_1-Minute*
One minute quasi-definitive data	INTERMAGNET magnetic observatory – specifications[2], modified for baseline accuracy[3].	*INTERMAGNET_1-Minute_QD*
Hourly means	No relevant standard	
Daily means	No relevant standard	
Monthly means	No relevant standard	
Annual means	No relevant standard	

References to standards:

1 INTERMAGNET Technical Note 6
2 Section 2.1 of the INTERMAGNET technical manual version 4.
3 Quasi-definitive definition on the INTERMAGNET web site: http://www.intermagnet.org/faqs-eng.php#quasi-definitive

If a standard is met in full or not met at all, PartialStandDesc can be omitted. Where a standard is partially met (e.g. the time stamp accuracy is within tolerance, but the data is not baseline corrected), the name of the relevant standard should be put in the *StandardName* attribute, *StandardLevel* should be set to *Partial* and the *PartialStandDesc* attribute should be filled in with a comma separated list of the sub-sections from the standard that the data meets. E.g. if the data meets the time stamp accuracy and thermal stability sections of the 1-minute data standard, enter *IMOM-01,IMOM-16* in *PartialStandDesc*.

Value to put in *PartialStandDesc*	**Description**
One-minute Definitive Data: General specifications	

IMOM-01	Time-stamp accuracy (centred on the UTC minute): 5s
One-minute Definitive Data: Vector Magnetometer specifications	
IMOM-11	Absolute Accuracy: ±5nT
IMOM-12	Resolution: 0.1nT
IMOM-13	Dynamic Range: ≥±4000nT High Lat., ≥±3000nT Mid/Equatorial Lat.
IMOM-14	Band pass: D.C. to 0.1Hz
IMOM-15	Minimum sampling rate: 1Hz
IMOM-16	Thermal stability: 0.25nT/°C
IMOM-17	Long term stability: 5nT/year
IMOM-18	Filtering to one-minute data: INTERMAGNET Gaussian
One-minute Definitive Data: Scalar Magnetometer specifications	
IMOM-21	Resolution: 0.1nT
IMOM-22	Absolute Accuracy: ±1nT
IMOM-23	Minimum sampling rate: 0.033Hz (30 sec)
One-second Data: General Specifications	
IMOS-01	Time-stamp accuracy (centred on the UTC second): 0.01s
IMOS-02	Phase response: Maximum group delay: ±0.01s
IMOS-03	Maximum filter width: 25 seconds
IMOS-04	Instrument amplitude range: ≥±4000nT High Lat., ≥±3000nT Mid/Equatorial Lat.
IMOS-05	Data resolution: 1pT
IMOS-06	Pass band: DC to 0.2Hz
One-second Data: Specifications in the Pass Band [DC to 8mHz (120s)]	
IMOS-11	Noise level: ≤100pT RMS
IMOS-12	Maximum offset error (cumulative error between absolute observations): ±2. 5 nT
IMOS-13	Maximum component scaling plus linearity error: 0.25%
IMOS-14	Maximum component orthogonality error: 2mrad
IMOS-15	Maximum Z-component verticality error: 2mrad
One-second Data: Specifications in the Pass Band [8mHz (120s) to 0.2Hz]	
IMOS-21	Noise level: ≤10pT/√Hz at 0.1 Hz
IMOS-22	Maximum gain/attenuation: 3dB
One-second Data: Specifications in the Stop Band [≥ 0.5 Hz]	
IMOS-31	Minimum attenuation in the stop band (≥ 0.5Hz): 50dB
One-second Data: Auxiliary measurements:	
IMOS-41	Compulsory full-scale scalar magnetometer measurements with a data resolution of 0.01nT at a minimum sample period of 30 seconds.
IMOS-42	Compulsory vector magnetometer temperature measurements with a resolution of 0.1°C at a minimum sample period of one minute.

ImagCDF variable attributes

The following attributes apply to individual variables - there is an attribute entry for each geomagnetic field element or temperature in an ImagCDF file and the value of that entry applies only to that field element or temperature. The “Entries” column shows whether the attribute is:

- Mandatory Number of entries per variable is exactly 1
- Optional Number of entries per variable may be 0 or 1

Superscript numbers following the attribute name show:

1. That the attribute is a recommended attribute for use with NASA's CDF tools
2. That the attribute is part of the ISTP/IACG guidelines - http://spdf.gsfc.nasa.gov/istp_guide/vattributes.html

None of these attributes are required for the time stamp variables.

Attribute name	Type	Entries	Description
FIELDNAM[1]	String	1	Set to "Geomagnetic Field Element " + the element code (e.g. H, D, Z,… - see section 4.6 for a list of valid codes); or set to "Temperature " + the name of the location where the temperature was recorded.
UNITS[1]	String	1	Must be one of "nT", "Degrees of arc" or "Celsius"
FILLVAL[1]	Double	1	The value used to show that a data sample is missing. Set to 99999.0 for compatibility with other formats. The values must either less than VALIDMIN or greater than VALIDMAX.
VALIDMIN[1]	Double	1	The smallest allowed numeric value for the data in the corresponding variable.
VALIDMAX[1]	Double	1	The largest allowed numeric value for the data in the corresponding variable.
DEPEND_0[2]	String	0 - 1	For geomagnetic and temperature data, set this to the name of the variable that holds time stamps for this time series. For records containing time stamps, do not set this variable.
DISPLAY_TYPE[2]	String	1	Set to "time_series"
LABLAXIS[2]	String	1	Set to the element code (as defined in section 4.6)

5. ImagCDF file names

ImagCDF files are named using the convention:

```
<iaga-code>_<date-time>_<cadence>_<publication-level>.cdf
```

- Iaga-code is the three letter IAGA code for the observatory that the data is from. This should match the IAGA code in the attribute described in section 4.2.
- Date-time is the start date/time of the data in the file. The format for the date/time is described below.
- Cadence is the sample period of the vector data, expressed as an ISO 8601 duration. ISO 8601 duration strings for common geomagnetic sample periods are shown below.
- Publication-level is the *PublicationLevel* attribute from section 4.2.

Filenames are in lower case. Files may contain arbitrary amounts of data, however the amount of data is not coded into the filename.

5.1 ISO 8601 duration strings for common geomagnetic sample periods

Sample Period	ISO 8601 duration string
1 second	PT1S
1 minute	PT1M
Hourly means	PT1H
Daily means	P1D
Monthly means	P1M
Annual means	P1Y

The table above is a set of examples. Other sample periods may be used provided that the sample period used represents the vector data and conforms to ISO 8601.

5.2 Format of date/time portion of filename and examples

The date/time portion of the filename is formatted differently for different data with different data intervals:

Data Interval	Date/time format	Example filename
Annual means	YYYY	*esk_2000_P1Y_4.cdf* – final annual mean data from Eskdalemuir starting in the year 2000.
Monthly means	YYYYMM	*ott_201401_P1M_4.cdf* – final monthly mean data from Ottawa starting in January 2014.
Daily means	YYYYMMDD	*gua_20100101_P1D_4.cdf* – final daily mean data from Guam starting at the beginning of 2010.
Hourly means	YYYYMMDD_HH	*naq_20020201_00_PT1H_4.cdf* – final hourly mean data from NAQ in January 2002
Minute means	YYYYMMDD_HHMM	*naq_20020120_0000_PT1M_3.cdf* – 'bulletin' or quasi-definitive minute mean data from NAQ for 20th January 2002 starting at midnight
Second	YYYYMMDD_HHMMSS	*naq_20020120_012300_PT1S_1.cdf* – raw 1-second data from NAQ for 20th January 2002 starting at 01:23:00

- YYYY = four digit year (i.e. 2002)
- MM = two digit month (01 for January - 12 for December)
- DD = two digit day of month (01-31)
- HH = two digit hour (0-23)
- MM = two digit minute (0-59)
- SS = two digit second (0-59)

6. Tools to look at CDF data

Once you have some CDF data you will want to look at it. The CDF toolset (that is installed for you when you install CDF) provides programs to do this. The simplest way is (using a command shell or DOS prompt):

```
cdfdump <filename> | more
```

Which will display the entire contents of the file. The example file in Appendix A is formatted using cdfdump. Other tools from the CDF toolset that may be useful include cdfexport and cdfedit.

Autoplot is a useful tool that can plot the time series data in ImagCDF data files. You can download it from http://autoplot.org/.

A number of packages provide access to CDF data. These include Matlab and IDL. A list is maintained on the NASA CDF website: http://cdf.gsfc.nasa.gov/html/FAQ.html#cdfsw

Octave may be able to read CDF data: http://lists.gnu.org/archive/html/help-octave/2007-04/txtyRdYjKU85H.txt. The Wolfram Language can use CDF data via its NASACDF data format - http://reference.wolfram.com/language/guide/NumericalDataFormats.html.