DEPRI

Depri Data Format 1.3a

Reference Guide with Anemos / Anemos.plus / SafeWind specialties

GENERAL INFORMATION

Authors	Thomas Pahlke, Hans-Peter Waldl, Felix Dierich, Martin Strack
Affiliation	Overspeed GmbH & Co. KG
Address	Marie-Curie-Str. 1, 26129 Oldenburg, Germany
Telephone	+49-441-36116-301
Email address	info@overspeed.de
Further authors	
Reviewer	Hans-Peter Waldl
Approver	

DOCUMENT INFORMATION

Document type	Reference Guide
Document name	depri13a-uk.20110131.doc
Version	1.3a
Revision number	317
Revision date	2011-02-01
Status	Reviewed
Confidentiality	General Public

Table of Contents

CHANGE HISTORY

Version	Date	Author	Changes
1.2a	2009-10-09	S. Bremer / F. Dierich	converted Latex version to Word
1.3a beta	2010-06-16 – 2010-08-03	F. Dierich	major overhaul with many changes to update for current status and Depri additions for Anemos.plus and SafeWind (e.g. ensemble data extension), fix mistakes and to increase clarity, updated formatting, added title page
1.3a beta	2010-08-04	F. Dierich	Added ensemble data syntax extensions and examples, some other small changes
1.3a gamma	2010-09-10	F. Dierich / H.-P. Waldl	incorporated review by H.-P. Waldl, translated and added some sections previously commented out (in LaTeX version), added sections on grid-data and siteindex
1.3a	2011-01-31	F. Dierich	incorporated final review by H.-P. Waldl and some other input, updated ensemble and extra index custom headers, some minor beautification

1. Introduction

The Depri format is a data format for storing time series and statistical data including measurements, predictions and other suitable data, which is generally organized as data records with a number of channels. It is a plain ASCII format.

In the Depri file format meta information describing the data is included in the same file as the data, so that it should be always possible to interpret the data without further external information.

This reference guide introduces the Depri file syntax and gives some examples. This version of the guide also includes special conventions and changes made for the Anemos system.

2. General formatting

The header of a Depri file contains the meta information in a fixed format, which is described in the next chapter. Below this the actual data follows in data record lines. The following rules apply to all parts of the Depri file:

Data tokens in the meta data lines and in the lines with the actual data are separated by blanks (ASCII 32) or tab characters (ASCII 9). Data lines (meaning data records) and header lines have to be separated by CR or CR-LF (usually the operation system standard). Tools to read Depri data should always be able to read both line separator options!

Lines containing data may be as long as necessary, lines of the header are limited to 255 characters. Complete lines or trailing parts of lines can be commented out by `//` or `#` .

For Anemos it was agreed that Depri files should not contain any blank lines.

Tokens containing blanks:

Deprecated: In older Depri versions tokens containing blanks had to be quoted by "..." (ASCII 34). Now this is deprecated, most Depri parsing libraries do not support it anyway.

Many tokens that are parsed automatically must not contain blanks, this includes tokens in data lines and the individual tokens in the headers `datatype`, `site`, `channel`, `location`, `missingvalue` etc. Note that blanks are allowed as token separators, just not within tokens. For other header elements tokens **with blanks** are allowed, this includes `creator`, `fileinfo`, `filesource`, `copyright`, `origin`, `averagetime`, `comment` etc. See also the sections for the individual headers and the examples.

2.1. Dates and timestamps

All dates and timestamps must be given in ISO compliant format, in the order `YYYY-MM-DD` for dates and `YYYY-MM-DDThh:mm[:ss[.SSS]]` for timestamps (T being a literal 'T' / 't'). For Anemos the format for timestamps in data tokens has been fixed to `YYYYMMDDhhmm`, e.g. `200401011200`. Nevertheless timestamps with seconds not being zero should include the seconds also for Anemos, i.e. `YYYYMMDDhhmmss`.

3. File Header

The file header starts at the beginning of the data file. All keywords are preceded with `#` or `##` (ASCII 35). The standard Depri headers start with `#` (e.g. `#site`) whereas `##` is used for custom headers. Custom headers can be used to store user information not defined in the Depri standard in a standardized way. This includes some custom headers defined for Anemos (see section 7.1) and for ensemble data.

The standard Depri keywords need to be written in lowercase, although some Depri libraries will accept them written in a different way.

There must not be any whitespace between the `#` and the keyword, as a `#` followed by a blank starts a comment line.

The following subsections describe the different standard Depri headers.

3.1. File specific information

The header components described in the following describe meta information related to the Depri file itself.

3.1.1. Depri Version

`#version [DEPRI] versionNr`

Version number of Depri format (currently 1.3A). The token “DEPRI” is optional.

e.g.

`#version DEPRI V1.3A`

`#version V1.3a`

3.1.2. Language Information

`#language Language`

Language used for texts. Use English name (e.g. `german` | `english` | `spanish`) or abbreviation (e.g. `de` for German, `en` or `uk` for English etc.).

e.g. `#language en`

3.1.3. Creating program

`#creator program-name [version]`

Name of data storage, acquisition or filter program

e.g. `#creator Anemos TSDR V2.06`

3.1.4. File Name

`#fileinfo PathAndFileName [DateTime]`

Identification and creation time (optional) of this file, blank separated.

e.g. `#fileinfo  /data/depri/N1303.dep 2003-11-10T09:13:44`

3.1.5. Source File Name

`#filesource PathAndFileName [DateTime]`

Identification and creation time (optional) of the source file, which was used to generate current data. Can appear more than once, e.g. to indicate a chain of processing or multiple input files.

e.g.

```
#filesource /data/raw/N1303.dat 2003-11-10T09:13:44
#filesource /data/processed/N1303.20031110t0913.dat
```

3.1.6. Responsible Person

`#user Name`

Name of responsible person, email address etc. In case of TSDR or similar service: user login used to retrieve Depri file.

e.g.

```
#user James Blond, Email: James@moneypenny.net
#user anemos-ro
```

3.1.7. Copyright

`#copyright CopyrightInformations`

Information about copyright for the current data. It is possible to give more than one copyright line.

e.g. `#copyright (c)2010 Overspeed GmbH & Co. KG`

3.2. Data specific information

The header components described in the following describe meta information related to the data contained in a Depri file.

3.2.1. Data Type

`#datatype timeseries [quantiles | ensembles | grid | ...]`

`#datatype statistic [bin [MinValue Value], spec, psd, rfc, dist, phase]`

Specifying the data type is most important for data analysis. Possible types are `timeseries` and `statistics`.

Within the Anemos system Depri files are only used for timeseries data.

For the type `timeseries`, additional information (optional) may be:

`quantiles`	indicates the content is uncertainty data with quantiles
`ensembles`	indicates the file contains ensemble predictions (see section 5)
`scenarios`	indicates the file contains prediction scenarios (format similar to ensembles)
`grid`	indicates the file contains gridded data (see section 6)

For the type `statistics`, additional information may be:

`BIN`	BIN-averaged data with bin width
`SPEC`	spectrum (general)
`PSD`	Power spectral density
`RFC`	Rainflow count
`DIST`	Frequency distribution, histogram
`PHASE`	Phase spectrum

Examples:

```
#datatype timeseries
#datatype timeseries quantiles
#datatype statistic DIST
#datatype statistic BIN 0.5 0.25
```

3.2.2. Origin of data

`#origin origin of Data`

A short description of the origin of the data

e.g.

```
#origin Norddeich II
#origin TNO EMS
#origin ECMWF
```

3.2.3. Data Source and Processing

`#source original | processed [process1, process2, ...]`

Additional information concerning the origin and processing of data (“source” being a historical naming)

e.g.

```
#source original
#source processed average 10Min, Fourier Analysis
```

3.2.4. Data filtering

`#filter [Value <|<=] ChannelRef [<|<= Value]`

Indicates filtering applied to the data, i.e. the file only contains data that passed this filter. Note that the order may be relevant if more than one filter operation was applied, also in combination with `#source` headers and their order (see above). Value can also be a timestamp if applied to time channels.

E.g.

```
#filter 2.0 <= WSpdMeas_EMS <= 8.0
#filter PowerPred_WPP <= 1200
#filter 2010-09-10 <= TimeMeas_UTC < 2010-09-11T1200
```

3.2.5. Average Time and Sampling Rate

Recorded measurement values are usually averaged from several measured samples for a given time period. This timeframe for averaging must be stated in the header (`#averagetime`), the sampling rate can be stated as well e.g. for critical applications (`#samplerate`).

More generally `#averagetime` is used to specify the typical time step between individual data lines, even if the data is not necessarily averaged. It is also used for prediction data to denote the time resolution of the predictions.

`#averagetime value [unit]`

Averaging time of single data values or time resolution of the data.

Possible units are:

days (`d`), hours (`h`), minutes (`min`), seconds (`sec`), milliseconds (`msec`). Default is seconds. If using days or hours, the averagetime value must be an integer.

e.g. `#averagetime 60 sec`

`#samplerate value [unit]`

Sampling rate of the measurement before averaging. See above for possible units.

e.g. `#samplerate 2.5 msec`

3.3. Geographic Coordinates of Site

To describe the coordinates of the site to which the data is attributed to (e.g. the measurement site), different coordinate systems can be used. These include UTM, geographical latitude and longitude, Gauß-Krüger and user defined coordinates. For Anemos only UTM is to be used.

For all options the height above mean sea level (HeightAMSL) and the height above ground level (HeightAGL) need to be included as additional tokens. All heights are given in meters. If the heights are not known, `-9999` is to be stated. An optional description of the site may follow.

The position of the site is also the default position for all data channels, if no special `#location` statement is present (see section 3.5.1).

For coordinate conversions, we recommend the free tool Geotrans or similar → http://earth-info.nga.mil/GandG/geotrans/ .

3.3.1. UTM

#site UTM MapBase [Hemisphere] Zzone Eeasting Nnorthing HeightAMSL HeightAGL [Description]

For UTM the map base is the first parameter (usually `WGS-84`), then 'N' or 'S' for the hemisphere (optional, but highly recommended), then the zone, the easting and northing are stated, each preceded by a letter defining the type of parameter. The values end with the heightAMSL, heightAGL and description (see above).

e.g.

```
#site UTM WGS-84 N Z28 E591022 N5996316 20 10 Anemometer02
#site UTM WGS-84 S Z54 E437385 N5830561 -9999 -9999
```

3.3.2. GEO: geographic coordinates

#site GEO latitude longitude HeightAMSL HeightAGL [Description]

Parameters are geographic latitude (northern lat. is positive) and longitude in degrees (eastern longitude is positive). Geographic coordinates may be decimal (e.g. 53.1235) or in degrees, minutes, seconds, separated by % (ASCII 37). Seconds can have additional digits after a floating point. The values end with the heightAMSL, heightAGL and description (see above).

e.g.

```
#site GEO 56.1 7.1 110 12 Elabor Uni Ol
#site GEO 56%05%34.4 7%06%04.4 110 12
```

3.3.3. GK: Gauß-Krüger coordinates

#site GK right up HeightAMSL HeightAGL [Description]

Parameters are Gauß-Krüger right and up parameters in meters. The values end with the heightAMSL, heightAGL and description (see above).

e.g. `#site GK 3543500 5601000 110 12 Elabor Uni Ol`

3.3.4. USER: user defined coordinates

#site USER p1 p2 [...] HeightAMSL HeightAGL [Description]

This allows user defined parameters to be used to describe the coordinates. The values end with the heightAMSL, heightAGL and description (see above).

3.4. Channel definitions

`#channel physSymbol globName physName physUnit [dataType [subDataType]]`

or for ensemble data channels:

`#channel physSymbol globName/ensembleMemberNr physName physUnit [dataType [subDataType]]`

For every data channel (and additional channels like date, time, data quality or status etc.), there must be exactly one `#channel` line in the Depri header. These lines must have the exact same order as the columns in the data part of the file. For improved readability they should appear as a block not mixed with any other header elements.

The channel definition lines are vitally important to all software parsing Depri files, as they are used to interpret the values of the data lines in the data part of a Depri file. Therefore all five parameters of each `#channel` line must always be specified correctly.

These parameters must not contain whitespace. To compensate you can use CamelCase, i.e. capitalize every new word as in `DateAndTime`. It is strongly recommended to just use ”normal” characters (a-z, A-Z) and digits for all parameters, to at least stay within 7bit-ASCII. Therefore special characters like ‘°’ should be replaced by texts like “deg” (HTML entity codes may be a good inspiration, used without the ‘&’ and ‘;’).

The combination of physSymbol and globName must be unique (see below).

The parameters are:

physSymbol

A physical symbol as the channel’s primary name. This could be a symbol like you would use in formulas (e.g. `v`) or an abbreviation of a textual name (e.g. `Time`, `WSpd` or `AirPress`).

The physSymbol forms the first part of the unique channel name (see also globName).

For Anemos the physSymbols are fixed for most types of channels, see the corresponding documentation (originally Anemos Deliverable 1.3a ”Specification and pre-standardization of data”, section 2.2) and provided example Depri files. It is important to used these fixed names, as the physSymbol and globNames are usually used by software components to recognize certain input channels.

Examples of the Anemos physSymbols are: `TimeNWP`, `WSpdNWP`, `TempNWP`, `TimeMeas`, `PowerMeas`, `WDirMeas`, `TimePred`, `PTimePred`, `PowerPred` etc.

globName

A short description of the concise source of this particular channel’s data, e.g. the measurement point for measured data. Also used for differentiating several channels with the same physSymbol, e.g. for NWP data a reference to the NWP model level, e.g. for a GFS model “GFS10m” for 10 meter level, “GFS1000” for the 1000 mbar level. This could also be what you would use as index on the physical symbol, (e.g. `tower1` for v_{tower1}).

The globName forms the second part of the unique channel name, so that a channel can be uniquely referenced by physSymbol plus globName. In Depri V1.3a a channel is referenced with `physSymbol_globName` (e.g. `WSpdNWP_GFS10m`).

For Anemos there are commonly used globNames, which are also documented in the channel naming guidelines and Depri file examples. Even if there is no obvious globName for a channel, a globName

has to be specified.

ensembleMemberNr (only used for ensemble channels)

See section 5 for an overview on how to store ensemble, scenario or similar data in Depri files, see section 10.4 for an example Depri file with ensemble data.

For data channels of ensemble data (or data of similar structure like scenario forecasts), there needs to be one #channel line for each member instance of the channel, so that the number of #channel lines always matches the number of columns in the data block, even though the channel definition is the same for all member instances of a data channel. The number of the ensemble member is appended to the globName, separated by a forward slash ("globName/ensembleMemberNr"). That way the number of tab / blank separated tokens in a `#channel` line remains the same. The member number is padded with zeros to three or more digits (as defined in the custom header `##ensembleDigits`, see section 5.2).

The channel duplication is not used in the alternative normalized format for Depri files with ensemble data. The #channel definition for an ensemble channel then is the same as for other channels.

physName

The name of the physical quantity (e.g. `Windspeed, WindDirection, Power, DateTime` etc). For Anemos see the channel naming guidelines for typical names used here.

physUnit

The physical unit: For channels holding data values this should be an ASCII compatible notation like e.g. `m/s`, `kW` or `degC` (degrees Celsius). For channels holding times `d` is usually used, for error flag channels `1` (see examples).

dataType

The data type describes the type of values held by a channel in a standardized way to allow for automatic processing by Depri libraries. Although this parameter is optional in the original Depri format, it is strongly recommended to always provide it.

In the Depri files used within the Anemos system, this parameter is mandatory.

Table 1 shows the valid data types and as what type they are usually stored in the software (Java/C types used here). Note that in the Anemos system, the time is always to be specified in the `datetime` format and only a subset of types is usually supported.

keyword	description	Java/C type
mean	Mean value (default)	double
int	Integer value	int
min	Minimum	double
max	Maximum	double
stddev	Standard deviation	double
dir	Angle in degrees, (0…360°)	double
rad	Angle in radians, (0…2π)	double
number	Enumeration	int
index	Index	int
offset	Offset	double
void	used if no type can be specified	double
date	Date	date object
time	Time	time object
datetime	date and time (without whitespace) ISO ordering (e.g. `200401011200` or `2004-01-01T1200`)	datetime object

timestamp	date and time (without whitespace) ISO ordering (e.g. `200401011200` or `2004-01-01T1200`)	datetime object
doy	Day of year 1st of January has value 1	int
toy	Time of year, measured in days. XXXX-01-01T0000 has toy = 0.0000	double
status	Status, general or connected to another channel (DS for Anemos)	int
quality	DataQuality (DQ for Anemos)	int
error	Error indicator	int
itvmin	Lower bin range	double
itvmax	Upper bin range	double
itvcount	Number of counts in bin	int
density	Probability density	double
binbase	Base channel, e.g. for bin analysis	String / char[]

Table 1: Valid data types for channels

A secondary sub data type can be specified to define the channel type more precisely. It must be separated from the primary data type keyword by whitespace or `:` (see examples). It is e.g. used to specify the timezone of date/time channels → "datetime:utc" or that a mean is the mean of a direction → "mean:dir".

The following are examples of valid channel definitions as could be used in the Anemos system:

```
#channel TimeMeas      UTC   Datetime        d     datetime:UTC
#channel PowerMeas     EMS   Power           kW    mean
#channel PowerMeasDQ   EMS   DataQuality      1     quality
#channel PowerMeasDS   EMS   DataStatus      1     status
#channel WSpdMeas      EMS   WindSpeed       m/s   mean
#channel WspdMeasDQ    EMS   DataQuality      1     quality
#channel WspdMeasDS    EMS   DataStatus      1     status
```

3.5. Additional channel attributes

The following paragraphs explain how additional channel attributes can be specified in the Depri format.

These attribute definitions have to be written in a header line each and can only be made for one channel at a time. They should appear after all #channel lines. Therefore the complete unique channel name always has to be specified as first token in the format physSymbol + underscore + globName (e.g. `WSpdMeas_EMS`, new in Depri V1.2A). This is called ChannelRef in the following. For ensemble data (and the like), the additional channel attributes are only stated once for a channel, not for each ensemble member instance of it, i.e. the ensembleMemberNr is not part of the ChannelRef.

3.5.1. Geographic Coordinates of a Measurement Channel

`#location ChannelRef GEO | GK | UTM | USER GeoParams HeightAMSL HeightAGL Description`

The location of the associated site for a single channel can be specified in the same format as used for `#site` (see section 3.3), only that the channel reference needs to be included. This is useful for channels, which have other locations than the global site, e.g. another measurement height or position.

e.g.

```
#location PowerMeas_EMS UTM WGS-84 N Z28 E591022 N5996316 20.0 50.0 wust01
```

3.5.2. Data Range

#range ChannelRef Min Max [Delta]

To specify the possible value range of the given channel. These values can then be used for range checks etc. The optional third value Delta can be used to set the grid spacing for diagrams.

e.g. `#range WDirNWP_GFS10m 0.0 360.0 30.0`

3.5.3. Error / Missing Value

#errorvalue ChannelRef Value1 [Value2]

#missingvalue ChannelRef Value1

With these commands the special values are specified, which are used in the data to signal that data is missing, because of gaps in a measurement / input feed or other problems or because the feed doesn't supply it at all. The normative error- / missing-value is `-9999.0` and should only be changed in cases where this is in conflict with valid values. For integer channels, it should be specified without the `.0` part, i.e. `-9999` (for other channels it can also be specified without the `.0`).

When using `#errorvalue`, two different values can be specified to differentiate between missing data (Value1) and otherwise erroneous data (Value2). With `#missingvalue` only the value for missing data can be specified.

Within the Anemos system only `#missingvalue` is to be used and the value has been fixed to always be `-9999.0` and nothing else. Note that this was originally `-99.0`, which must no longer be used for Anemos.

e.g.

```
#errorvalue WSpd_Nacelle -9999 -9999
#missingvalue PowerMeas_EMS -9999.0
```

3.5.4. Resolution

#resolution ChannelRef value

Specifies the resolution of a channel, e.g. the resolution of a measurement instrument. The resolution is specified as the increment / step size, e.g. as `0.1`, in the physUnit defined in `#channel`.

E.g.

```
#resolution WSpdMeas_EMS 0.25
#resolution WDirMeas_EMS 0.1
```

3.5.5. Error (systematic, statistical, total)

#error ChannelRef systematicError statisticalError

or

#error ChannelRef totalError

The absolute error or uncertainty of the channel, either specified as systematic and statistical error separately (two values stated) or as total error, a mixture of both error types (one value).

Note that `#errorvalue` has a totally different meaning, see section 3.5.3.

E.g.

```
#error WSpdMeas_EMS 0.13 0.05
```

```
#error WDirMeas_EMS 0.4
```

3.5.6. Re-norming

`#renorm ChannelRef poly factor0 ... factorN-1`

or

`#renorm ChannelRef table filename`

These headers indicate that a channel contains uncalibrated data, which **needs to be re-normed** either by a polynomial or by a calibration table.

In the **poly**nomial variant factors are specified, where the first factor is an offset and the following factors are multiplied with different powers of the channel value. E.g.

```
#renorm WSpdMeas_Raw poly 0.5 1.2 0.24 0.03
```

means the values of the channel WSpdMeas_Raw need to be re-normed as

WSpd = 0.5 + 1.2 value + 0.24 value² + 0.03 value³ .

In the **table** variant the (path and) filename of a file with a calibration table is specified, which has the format **`Value Offset`**, e.g.:

```
#renorm WSpdMeas_Raw table wspd-cal1.dat
```

with wspd-cal1.dat in the format

```
5.6  0.3
7.2  0.2
9.5 -0.13
...
```

Note: If the data in the file **is already re-normed** (or otherwise filtered), this should be indicated with a `#filter` or `#comment` header (as suitable).

3.6. Depri Comments

`#comment Comment text`

A commenting string, which may have up to the length of a header line. `#comment`s should be copied if the Depri file is copied or processed to generate a new data file.

In contrast to that, commented lines (beginning with `#` or `//`, see section 2) can be discarded when processing or copying.

e.g. `#comment This file is magic!`

4. Data section

This section explains the format to be used for the actual data section of a Depri file.

The data block is started with a `#begindata` line, after the data lines an `#enddata` should follow to mark the end of the file. Note that in cases, where Depri files are written continuously, they usually do not contain the `#enddata` line. Depri parsers may be tolerant to this.

4.1. Start of Data

`#begindata [DateTime]`

The actual data lines start after a line with the keyword `#begindata`.

The optional parameter DateTime is the time of the first/earliest data record. See section 2.1 for possible date / time formats.

e.g.

```
#begindata 2004-01-01T00:00:00
```

```
#begindata 201008030000
```

```
#begindata
```

4.2. Data

The data block starts in the line after `#begindata`. The columns need to be in the same order as the `#channel` lines in the header. In each data line the values of the different columns need to be separated by blanks or tabs. Note that some old Depri parsers cannot handle tabs. Data lines may be commented out by a `#` as first character of the line.

Example matching the example channel definition in section 3.4:

```
201008030000    9223.0      0    33    7.11  0    33
201008030001    9426.131    0    33    5.96  0    33
201008030002    9979.932    0    33    5.06  0    33
201008030003    9525.803    0    33    5.04  0    33
201008030004    8804.396    0    33    4.86  0    33
201008030005    10015.529   0    33    3.86  0    33
201008030006    10414.467   0    33    4.71  0    33
201008030007    13065.264   0    33    5.11  0    33
201008030008    16802.267   0    33    5.01  0    33
201008030009    17941.332   0    33    4.55  0    33
201008030010    21494.135   0    33    4.92  0    33
201008030011    21557.936   0    33    4.68  0    33
201008030012    22929.26    0    33    4.87  0    33
201008030013    21775.726   0    33    4.08  0    33
201008030014    21504.332   0    33    4.72  0    33
201008030015    22131.203   0    33    4.38  0    33
201008030016    23127.729   0    33    5.00  0    33
```

4.3. End of Data

#enddata [DateTime]

After the last line of data, an `#enddata` line has to follow (see also note at start of section 4). Everything after this line is ignored.

The optional parameter DateTime is the time of the last/latest data record. See section 2.1 for possible date / time formats.

e.g.

```
#enddata 2004-12-31T23:59:59
```

```
#enddata 201008030016
```

```
#enddata
```

5. Ensemble / multi-variant data

With Depri version 1.3a a syntax extension has been introduced to allow the storage of ensemble data and other multi-variant data of similar structure (notably scenarios data) in Depri files. This extension allows to store the results of an ensemble model run with all ensembles in a single file.

This is achieved by multiplying the data columns in the data part of the file for all members (not the Time, PTime or DQ / DS columns), multiplying the `#channel` definition lines noting the ensembleMemberNr (see also section 3.4), and adding additional custom headers.

There is also the alternative possibility to store the ensemble data in normalized form in a standard Depri file with an additional channel containing the member number, which is then part of the unique identification of a data record (together with the time and ptime). This is also the way, ensemble data is internally stored in the Anemos TSDR, which can provide both formats as output.

See also the document “TSDR enhancements for ensembles – Anemos.Plus WP2 Design Memo” and the ensemble file examples in sections 10.4 (standard form) and 10.5 (normalized form).

5.1. Ensemble channel naming

The ensemble channels are defined in a similar way as the channels of normal contexts. Their full name consists of the usual physSymbol and globName parts, with an additional ensemble member index. It is padded with zeros to three digits (or more if necessary).

Channels that only exist once, e.g. the Time and PTime channels, do not have this index. While physSymbol and globName are usually separated by an underscore (physSymbol_globName), the member index is separated by a forward slash, i.e. physSymbol_globName/index. An exception to this are the Depri header channel definition lines (`#channel ...`), which use whitespace to separate physSymbol and globName. In these lines the member index is postfixed to the globName with a slash as well, also to ensure existing Depri libraries do not need to be updated in most cases.

An example channel list for the fictitious NWP feed “Yet Another Ensemble Model”, short “YAEM”, with eight ensemble wind speeds and directions at 10m height and the standard DS and DQ channels, could look like this:

```
TimeNWP_UTC
PTimeNWP_UTC
WSpdNWP_YAEM10m/000
WSpdNWP_YAEM10m/001
WSpdNWP_YAEM10m/002
WSpdNWP_YAEM10m/003
WSpdNWP_YAEM10m/004
WSpdNWP_YAEM10m/005
WSpdNWP_YAEM10m/006
WSpdNWP_YAEM10m/007
WDirNWP_YAEM10m/000
WDirNWP_YAEM10m/001
WDirNWP_YAEM10m/002
WDirNWP_YAEM10m/003
WDirNWP_YAEM10m/004
WDirNWP_YAEM10m/005
WDirNWP_YAEM10m/006
WDirNWP_YAEM10m/007
RecordDQ_DQ
RecordDS_DS
```

5.2. Additional Ensemble Depri file headers

A Depri header for a file containing ensemble data has definitions for all ensemble channels included in the Depri file. Additional channel information like missingValues or locations are however only specified once for all member variants of a channel.

Additional header keywords are defined for ensemble data (these are recommended but optional for module output; they are always included in Depri files generated by the TSDR):

- `##ensembleMembers` specifies the total number of ensemble members,
- `##ensembleDigits` specifies the number of digits used in the member indexes,
- `##ensembleStandardMember` denotes the standard member (if any).

To be able to interpret ensemble Depri files in normalized format (and for internal processing in the TSDR) two further custom headers are used in normalized ensemble Depri files:

- `##ensembleIndexChannel` the name of the channel with the member number,
- `##ensembleDataChannels` a comma-separated list of the channels that are different for each ensemble.

The list in `##ensembleDataChannels` usually contains all data channels, but not any time channels or RecordDQ/DS channels. All channels that are not in this list, need to have the same value for each ensemble for a point in time (or Time-PTime combination), otherwise information would be lost when converting from normalized ensemble Depri to standard ensemble Depri.

For **scenarios** data, these additional custom headers are named differently, “ensemble” is replaced by “scenario”, i.e. `##scenarioMembers`, `##scenarioDigits`, … The same should be considered if other multi-variant data with a structure similar to ensembles or scenarios is stored in such Depri files (see also 6).

6. Variants with extra index channels

While the normalized format of ensemble data (see above) uses a special index channel to indicate the ensemble member, there are other notable usage possibilities of extra index channels, which are described below. These have in common (also with the normalized ensemble variant), that they are syntactically standard Depri, but have more than one record per time step (and prediction run time, if applicable) and use extra custom headers.

The naming of the extra custom headers is chosen, so that it is always the variant code plus the same suffixes, i.e. `xxxIndexChannel`, `xxxDataChannels`, `xxxDigits`, `xxxStandardMember`, `xxxMembers` (as applicable) with `xxx` being `ensemble`, `scenarios`, `sitekey`, `grid` etc.

6.1. Gridded data

Gridded data can be stored in Depri files in a normalized fashion by using extra index channels. That way it is possible to store data like gridded NWPs, which typically have five to six dimensions as illustrated in the following table (together with the way in which it is stored):

dimension	**representation in Depri**
time	Time channel (standard)
prediction run time	PTime channel (standard)
x-dimension of grid (e.g. longitude)	extra index channel
y-dimension of grid (e.g. latitude)	extra index channel
predicted variable (wspd, wdir, temperature, …)	data channel (physSymbol indicating type)
[predicted level (z-dimension)	data channel (globName indicating level)]

While four dimensions are supported by standard Depri files with prediction data, the two additional x and y-dimensions are stored in additional index channels. These channels typically use a numbered index for the grid or a geographical reference like degrees (potentially floating point). The data part then has one data line for each time (and possibly prediction run time) and each grid point indexed by the two extra channels.

This normalized representation is naturally not as visually intuitive as standard Depri and it consumes more disk space than binary grid storage formats, but it allows processing with most Depri compatible libraries / software and easy extraction of single grid points.

There are extra custom headers to use for gridded Depri data:

`##gridIndexChannel ChannelRef1,ChannelRef2`

defines the channels used as row and column indexes in a comma-separated list. Note that the `#channel` and `#resolution` headers should be used to further specify the unit and resolution used (see example below). Note that col(umn) (x-dimension) and row (y-dimension) may be unclear in case of rotated grids, in this case `##gridRotation` should be specified to clarify (see below).

`##gridDataChannels ChannelRef1[,ChannelRef2,...]`

A comma-separated list of the data channels. Internally needed by TSDR, optional for Depri files.

##gridRotation Angle [deg | rad]

The angle of grid rotation with unit (° or radians).

Other custom headers may be added later to provide additional information on the nature of the grid.

Example of selected header and data lines in a grid Depri file:

```
#version 1.3A
#datatype timeseries grid
...
#channel TimeNWP         UTC      Datetime             d    datetime:UTC
#channel PTimeNWP        UTC      Datetime             d    datetime:UTC
#channel GridCol         Deg      GridColumnIndex      deg  index
#channel GridRow         Deg      GridRowIndex         deg  index
#channel WSpdNWP         GFS10m   Windspeed            m/s  mean
#channel WDirNWP         GFS10m   WindDirection        deg  mean:dir
#resolution GridCol_Deg 0.5
#resolution GridRow_Deg 0.5
##gridIndexChannel GridCol_Deg,GridRow_Deg
##gridRotation 0.0 deg
...
#begindata   200001010000
200912010000 200912010000 94.5   12.0   2.43   93.2
200912010000 200912010000 94.5   12.5   2.65   88.84
200912010000 200912010000 94.5   13.0   2.50   97.99
200912010000 200912010000 95.0   12.0   2.40   92.64
200912010000 200912010000 95.0   12.5   2.37   102.52
200912010000 200912010000 95.0   13.0   2.44   98.57
...
```

6.2. Siteindex

An index channel may also be used to store and distinguish the data of different sites in a single Depri file. This is valuable if there is data for a high number of different sites and it is not desired to have separate files for each site. While most Depri libraries only support numerical (or datetime) channels, an alphanumeric channel with a site identifier could be considered as a future option. In numeric site index channels existing site numbering can be used.

Note that in this variant it is not guaranteed that there is an entry for each time (potentially each time – ptime combination) and each site.

Custom header for siteindex data:

##sitekeyIndexChannel ChannelRef

Indicates the channel containing the extra site index.

##sitekeyDataChannels ChannelRef1[,ChannelRef2,...]

A comma-separated list of the data channels, which are different for each site. Internally needed by TSDR, optional for Depri files.

Other custom headers may be added later to provide additional information on the site index.

Example of selected header and data lines in a Depri file with site index:

```
...
#channel TimeMeas        UTC      Datetime             d    datetime:UTC
#channel SiteIndex       Synop    SiteIndex            1    index
#channel WSpdMeas        Synop    Windspeed            m/s  mean
```

```
#channel WDirMeas        Synop    WindDirection        deg  mean:dir
##sitekeyIndexChannel SiteIndex_Synop
##sitekeyDataChannels WSpdMeas_Synop,WDirMeas_Synop
...
#begindata
200912210000 3362   4.43   93.2
200912210000 3821   2.65   88.84
200912210000 4591   4.50   97.99
200912210000 4597   4.40   92.64
200912210000 6672   5.37   102.52
200912210000 6812   3.44   98.57
...
```

7. Anemos Depri specialties

Besides the specialties already noted above, this section explains some further mandatory standardizations that have been made for the Depri files used in the Anemos system. See also the channel naming guidelines, the DQ/DS channel documentation and Anemos Deliverable D1.3A, which includes a checklist for Depri files, which should be used on all Depri files.

7.1. Custom headers

To allow an automatic selection and processing of Depri files, especially in connection with the TSDR, a few custom headers have been introduced.

7.1.1. Sitename (optional)

`##sitename Name of site`

The long name of the site.

e.g. `##sitename Wusterhusen`

7.1.2. Site shortname (required)

`##shortname shortname of site`

The short name of the site as agreed upon. This is used to look up the correct data context in the TSDR, so it needs to match the site / farm name of the data context. This is usually the standard code/ID used by the end user. It is not always necessarily shorter than the “long name” in ##sitename.

e.g. `##shortname wus`

7.1.3. Data context / majorType and subType (required)

`##model MajorType [SubType]`

or:

`##majorType MajorType`

`##subType subType`

This specifies what general “flavor”/type of time series data is contained in this file (majorType) and what specific variant (subType) it is. This could be the kind of model as majorType (“nwp”, “wpp”, …) and the specific model provider and/or name (e.g. “gfs”, “ecmwf”, “ecmwf-ifs” for NWP data) or a specific configuration of a specific model (e.g. “ecmwf_pcmodel” for PCModel WPPs based on ECMWF IFS NWPs). For measurements “measured” is used as majorType and potentially some more information on the measurement source or processing of the measurements (e.g. “filtered”) as subType (or “standard”).

When writing data to the TSDR, the majorType and subType information is used to look up the data context, therefore the parameters must match. The default SubType is `standard`, if not specified.

Either variant can used: a single `##model` line or the two separate `##majorType` and `##subType` lines. All three may also be specified, but must then match naturally.

Some valid values for MajorType are

- `nwp`: Numerical Weather Predictions
- `wpp`: Wind Power Predictions
- `measured`: Measured data

See the Anemos context naming guidelines for more information.

e.g. (##model variant)

```
##model wpp ecmwf_pcmodel
```

```
##model nwp gfs4
```

```
##model measured
```

```
##model measured standard
```

The last two examples are equivalent.

e.g. (other variant)

```
##majorType wpp
##subType ecmwf_pcmodel
```

```
##majorType nwp
##subType gfs4
```

7.1.4. Further informational headers (optional)

Further custom headers may be added to provide further meta information. This is e.g. done by the TSDR to inform about certain context configuration aspects. The custom parameters added by the TSDR include (depending also on context configuration and TSDR version):

```
##description Longer text description of data context
##farmname Long name of the site (duplicate of ##sitename)
##windfarm shortname of the site (duplicate of ##shortname)
##timeColumn Name of channel with time referred to by data record (Time)
##predTimeColumn Name of channel with prediction run reference time (PTime)
##timespan Name of context timespan, if using time-split contexts
```

Note that these are added in addition to the standard headers like ##sitename, ##shortname, ##model etc. E.g.:

```
##description ECMWF NWP for the windfarm Example WF
##farmname Example WF
##windfarm exwf
##timeColumn TimeNWP_UTC
##predTimeColumn PTimeNWP_UTC
##timespan all
```

See also section 5.2 for the additional custom headers used for ensemble data.

7.2. Data format and channels

- The channels and channel names need to follow the conventions as stated in the Anemos channel naming guidelines, formerly Deliverable D1.3a.
- For predictions two time channels are used to reference a value: The time for which the prediction is made (TimeXXX) and the reference time of the prediction run (PTimeXXX). Further time channels may be given in addition like NTimeXXX to indicate the NWP run time on which the predictions are based or UTimeXXX to indicated the time of the last measured production data used in a forecast. See also Anemos channel naming guidelines.
- Data Quality and Data Status channels have to be included (see ”Anemos Platform Documentation – Data Quality And Data Status Channels”).
- Timestamps have to be in the format `YYYYMMDDhhmm`, if seconds are different to zero in the format `YYYYMMDDhhmmss`.
- The missingvalues have to be `-9999.0` (see above).
- Power is to be given in kW.
- Windspeed is to be given in m/s.
- The files may not contain empty lines anywhere except as last line.

8. Minimum Depri header for Anemos

May be added in future document versions.

9. Syntax overview

May be added in future document versions.

10. Example files

The following are example Depri file as used in the Anemos system for different data streams.

10.1. Measured wind farm power / wind speed

```
#version 1.3A
#datatype timeseries
#language uk
#creator Anemos TSDR V2.06
#site UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999
#user exampleuser
#averagetime 1 min
#channel     TimeMeas     UTC    Datetime     d     datetime:UTC
#channel     PowerMeas    PI     PowerActual  kW    mean
#channel     PowerMeasDQ  PI     DataQuality  1     quality
#channel     PowerMeasDS  PI     DataStatus   1     status
#channel     WSpdMeas     PI     Windspeed    m/s   mean
#channel     WSpdMeasDQ   PI     DataQuality  1     quality
#channel     WSpdMeasDS   PI     DataStatus   1     status
#location    PowerMeas_PI UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999 exwf
#missingvalue PowerMeas_PI -9999.0
#missingvalue PowerMeasDQ_PI      -9999
#missingvalue PowerMeasDS_PI      -9999
#location    WSpdMeas_PI  UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999 exwf
#missingvalue WSpdMeas_PI  -9999.0
#missingvalue WSpdMeasDQ_PI -9999
#missingvalue WSpdMeasDS_PI -9999
#comment Measured 1 min SCADA averages for Example WF with importer filtering
##sitename Example Wind Farm
##shortname exwf
##majorType measured
##subType filtered
#begindata   201008030427
201008030427 613.412      0      33     3.24   0      33
201008030428 151.803      0      33     3.16   0      33
201008030429 439.533      0      33     3.06   0      33
201008030430 676.291      0      33     3.25   0      33
201008030431 769.928      0      33     3.25   0      33
201008030432 1083.951     0      33     3.33   0      33
201008030433 1101.314     0      33     3.54   0      33
201008030434 1116.321     0      33     3.40   0      33
201008030435 1228.189     0      33     3.47   0      33
201008030436 1043.52      0      33     3.43   0      33
201008030437 1033.846     0      33     3.41   0      33
...
#enddata     201008030437
```

10.2. NWP (single site numerical weather prediction)

```
#version 1.3A
#datatype timeseries
#language uk
#creator Anemos TSDR V2.06
#site UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999
#user exampleuser
#averagetime 3 hours
#channel     TimeNWP       UTC    Datetime     d      datetime:UTC
#channel     PTimeNWP      UTC    Datetime     d      datetime:UTC
#channel     WSpdNWP       GFS10m Windspeed    m/s    mean
#channel     WDirNWP       GFS10m WindDirection deg    mean:dir
#channel     RecordDS      DS     DataStatus   1      status
#channel     RecordDQ      DQ     DataQuality  1      quality
#location    WSpdNWP_GFS10m      UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999 exwf
#missingvalue WSpdNWP_GFS10m     -9999.0
#location    WDirNWP_GFS10m      UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999 exwf
#missingvalue WDirNWP_GFS10m     -9999.0
#missingvalue RecordDS_DS  -9999
#missingvalue RecordDQ_DQ  -9999
#comment GFS NWP data for exwf
##sitename Example Wind Farm
##shortname exwf
##model nwp gfs-direct
#begindata   201008030000
201008030000 201008030000 1.68   320.33 1      0
201008030300 201008030000 4.76   289.27 1      0
201008030600 201008030000 5.98   290.65 1      0
201008030900 201008030000 6.15   294.69 1      0
201008031200 201008030000 6.14   285.69 1      0
201008031500 201008030000 6.31   284.87 1      0
201008031800 201008030000 6.97   238.91 1      0
...
#enddata     201008101200
```

10.3. WPP (wind farm power prediction)

```
#version 1.3A
#datatype timeseries
#language uk
#creator Anemos TSDR V2.06
#site UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999
#user exampleuser
#averagetime 3 hours
#channel     TimePred     UTC    Datetime     d      datetime:UTC
#channel     PTimePred    UTC    Datetime     d      datetime:UTC
#channel     NTimePred    UTC    Datetime     d      datetime:UTC
#channel     PowerPred    WPP    Power  kW    mean
#channel     PowerPredDQ  WPP    DataQuality  1      quality
#channel     PowerPredDS  WPP    DataStatus   1      status
#location    PowerPred_WPP UTM WGS-84 S Z11 E306904 N5486077 -9999 -9999 exwf
#missingvalue PowerPred_WPP -9999.0
#missingvalue PowerPredDQ_WPP     -9999
#missingvalue PowerPredDS_WPP     -9999
#comment PCModel WPP based on GFS4 NWP for Example Wind Farm
##sitename Example Wind Farm
##shortname exwf
##model wpp gfs4_pcmodel
#begindata   201008030000
201008030000 201007240000 201007240000 137834.19    0      33
201008030000 201007241200 201007241200 116934.59    0      33
201008030000 201007250000 201007250000 139750.0     0      33
201008030000 201007251200 201007251200 138396.13    0      33
201008030000 201007260000 201007260000 121448.76    0      33
201008030000 201007261200 201007261200 131072.09    0      33
201008030000 201007270000 201007270000 139462.09    0      33
...
#enddata     201008040000
```

10.4. Ensemble NWP (standard form)

```
#version 1.3A
#datatype timeseries ensemble
#language uk
#creator Anemos TSDR V1.91
#site UTM WGS-84 N Z12 E345678 N9012345 -9999 -9999
#averagetime 3 hours
#channel TimeNWP  UTC          Datetime      d    datetime:UTC
#channel PTimeNWP UTC          Datetime      d    datetime:UTC
#channel WSpdNWP  YAEM10m/000  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/001  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/002  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/003  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/004  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/005  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/006  Windspeed     m/s  mean
#channel WSpdNWP  YAEM10m/007  Windspeed     m/s  mean
#channel WDirNWP  YAEM10m/000  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/001  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/002  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/003  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/004  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/005  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/006  Winddirection deg  mean:dir
#channel WDirNWP  YAEM10m/007  Winddirection deg  mean:dir
#channel RecordDS DS           DataStatus    1    status
#channel RecordDQ DQ           DataQuality   1    quality
#missingvalue WSpdNWP_YAEM10m     -9999.0
#missingvalue WDirNWP_YAEM10m     -9999.0
#missingvalue RecordDS_DS  -9999
#missingvalue RecordDQ_DQ  -9999
##description YEAM (yet another ensemble model) ensemble NWP for Example WF A
##ensembleDigits 3
##ensembleMembers 8
##ensembleStandardMember 000
##model nwp-ensemble yeam
##shortname WFA
##sitename Example Wind Farm A
#begindata   200001010000
200912010000 200912010000 2.43   2.65   2.50   2.40   2.37   2.44   2.57   2.68
       93.2   88.84  97.99  92.64  102.52 98.57  91.12  100.38 1      0
200912010300 200912010000 1.51   1.64   1.42   1.35   1.39   1.44   1.45   1.52
       140.35 127.1  129.79 144.13 122.24 135.55 145.27 119.8  1      0
200912010600 200912010000 2.72   2.25   2.66   2.44   2.29   2.35   2.58   2.34
       148.04 153.84 149.81 132.94 145.61 155.52 138.97 134.21 1      0
200912010900 200912010000 3.54   3.84   3.86   3.79   3.47   4.10   4.13   3.54
       132.85 116.52 114.66 127.52 119.77 135.71 129.14 126.4  1      0
...
#enddata
```

10.5. Ensemble NWP (normalized form)

```
#version 1.3A
#datatype timeseries ensemble
#language uk
#creator Anemos TSDR V1.91
#site UTM WGS-84 N Z12 E345678 N9012345 -9999 -9999
#averagetime 3 hours
#channel TimeNWP         UTC      Datetime             d    datetime:UTC
#channel PTimeNWP        UTC      Datetime             d    datetime:UTC
#channel EnsembleMember  Nr       EnsembleMemberNumber 1    index
#channel WSpdNWP         YAEM10m  Windspeed            m/s  mean
#channel WDirNWP         YAEM10m  Winddirection        deg  mean:dir
#channel RecordDS        DS       DataStatus           1    status
#channel RecordDQ        DQ       DataQuality          1    quality
#missingvalue WSpdNWP_YAEM10m     -9999.0
#missingvalue WDirNWP_YAEM10m     -9999.0
#missingvalue RecordDS_DS  -9999
#missingvalue RecordDQ_DQ  -9999
##description YEAM (yet another ensemble model) ensemble NWP for Example WF A
##ensembleDataChannels WSpdNWP_YAEM10m,WDirNWP_YAEM10m
##ensembleDigits 3
##ensembleIndexChannel EnsembleMember_Nr
##ensembleMembers 8
##ensembleStandardMember 000
##model nwp-ensemble yeam
##shortname WFA
##sitename Example Wind Farm A
#begindata   200001010000
200912010000 200912010000 0     2.43   93.2   1     0
200912010000 200912010000 1     2.65   88.84  1     0
200912010000 200912010000 2     2.50   97.99  1     0
200912010000 200912010000 3     2.40   92.64  1     0
200912010000 200912010000 4     2.37   102.52 1     0
200912010000 200912010000 5     2.44   98.57  1     0
200912010000 200912010000 6     2.57   91.12  1     0
200912010000 200912010000 7     2.68   100.38 1     0
200912010300 200912010000 0     1.51   140.35 1     0
200912010300 200912010000 1     1.64   127.1  1     0
200912010300 200912010000 2     1.42   129.79 1     0
200912010300 200912010000 3     1.35   144.13 1     0
200912010300 200912010000 4     1.39   122.24 1     0
200912010300 200912010000 5     1.44   135.55 1     0
200912010300 200912010000 6     1.45   145.27 1     0
200912010300 200912010000 7     1.52   119.8  1     0
200912010600 200912010000 0     2.72   148.04 1     0
200912010600 200912010000 1     2.25   153.84 1     0
200912010600 200912010000 2     2.66   149.81 1     0
200912010600 200912010000 3     2.44   132.94 1     0
200912010600 200912010000 4     2.29   145.61 1     0
200912010600 200912010000 5     2.35   155.52 1     0
200912010600 200912010000 6     2.58   138.97 1     0
200912010600 200912010000 7     2.34   134.21 1     0
...
#enddata
```