1. Introduction

This note is an implementation note on the Mini-SEED file format for InSight/SEIS raw data exchanges. The Mini-SEED file format will be used on InSight mission to communicate the raw data taken by the SEIS instrument and associated APSS sensors onboard the spacecraft to the Science team. The intent of this document is to define exactly the current and planned implementation of the Mini-SEED format by the SISMOC development team.
The note is starting from the SEED definition manual (as read in SEED Format Version 2.4 and reminded in dark-blue in the document) and tries to answer the question how do we fill that field in, with the data coming from the spacecraft?

2. Few words on SISMOC Mini-SEED tool implementation

The SISMOC Mini-SEED generation tool is a set of Python scripts that make an extensive use of the python library obspy. Because some of the necessary functions were missing in obspy, the library has been adapted to fulfill all the SEIS needs. These modifications have been reported in the library and will be part of the next release.

3. Inputs of SISMOC Mini-SEED generation tool

SEIS/APSS data are coming down to earth from the spacecraft in the frame of CCSDS packets. The content of each of these packets is defined in the document SEIS Flight to ground ICD (Ref JPL D-78505). At different extent four major types of packets are considered in the SISMOC Mini-SEED generation process:

- Packets containing time information,
- Packets containing “meta-data” information indicating that a particular event occurred on board (calibration, recentering, leveling, etc.),
- Packets containing compressed telemetry blocks of data generated by the flight software (FSW) on board data processing function, also named science packets,
- Packets containing housekeeping data.

The time information packets are used to convert the data timestamps made by the two different clocks on board the spacecraft (AOBT for APSS, LOBT for SEIS) into UTC time. At SISMOC, this conversion is done thanks to a dedicated tool named Time Correlation Tool (TCT). The SISMOC Mini-SEED generation tool uses TCT as a library in order to make all its necessary time conversion.

Packets containing “meta-data” information are used by the Mini-SEED generation tool to fill in the Mini-SEED header flags as detailed in this note.

The content of the housekeeping packets and the science packets are converted into Mini-SEED records. In the later case, a header in the compressed telemetry block allows to identify which channel is stored in the packet. To do so, the SISMOC Mini-SEED generation tool reads 2 items from the block header:

- The configuration files identifier, which identifies the configuration that was used by the data processing part of the instrument FSW to produce this data.
- The Channel ID in this file.

Thus a set data processing configuration files is also taken as input to the Mini-SEED generation tool.
It has to be noticed that the science packets are of two types: packets containing continuous channels and packets containing events data which are processed slightly differently (see section 4).

The Mini-SEED generation tool is also using as input several configuration files. In one of them, the channel codes, and location IDs for the various on board configuration are defined, reflecting the content of the SEIS Mini-SEED convention codes Excel file (ref. INS-ST-GRDS-1500-IPGP).

Finally, the tool takes as input an argument to discriminate the source of data in order to distinguish flight data from test data.

4. File granularity definition and naming convention

For continuous science packets, one single file is created per day per Mini-SEED channel and location code.
For data coming from the event buffers, one file per event and per channel and location code is created.

The filename convention for continuous data is *NET.STA.LOC.CHA.YYYY.DDD.REV.mseed*, where:

- *NET* is the Network ID,
- *STA* is the Station ID,
- *LOC* is the location ID,
- *CHA* is the channel ID,
- *YYYY* is the earth year of the measurement,
- *DDD* is the Julian day from earth,
- And *REV* is the revision of the file.

The filename convention for event data is *NET.STA.LOC.CHA.EVID.REV.mseed,* where:

- *NET* is the Network ID,
- *STA* is the Station ID,
- *LOC* is the location ID,
- *CHA* is the channel ID,
- *EVID* the event ID as expressed in the event request and identified in the Mars Quake Service catalogue,
- And *REV* is the revision of the file.

5. Mini-SEED Fixed header

	Note	Field name	Type	Length	Mask or Flags
	1	Sequence number	D	6	"######"
V2.4 -	2	Data header/quality indicator ("D"I"R"I"Q"I"M")	A	1	
	3	Reserved byte ("Δ")	A	1	
	4	Station identifier code	A	5	[UN]
	5	Location identifier	A	2	[UN]
	6	Channel identifier	A	3	[UN]
V2.3 -	7	Network Code	A	2	[ULN]
	8	Record start time	B	10	
	9	Number of samples	B	2	
	10	Sample rate factor	B	2	
	11	Sample rate multiplier	B	2	
	12	Activity flags	B	1	
	13	I/O and clock flags	B	1	
	14	Data quality flags	B	1	
	15	Number of blockettes that follow	B	1	
	16	Time correction	B	4	
	17	Beginning of data	B	2	
	18	First blockette	B	2	

1 Sequence number

Data record sequence number (Format "######"). The data record sequence number starts with value "000001" for each Mini-SEED file and is incremented for every record in the file. If it ever goes beyond "999999", it will roll over to "000001".

2 Data header/quality indicator

Previously, this field was only allowed to be "D" and was only used to indicate that this is a data header. As of SEED version 2.4 the meaning of this field has been extended to also indicate the level of quality control that has been applied to the record.

D — The state of quality control of the data is indeterminate.
R — Raw Waveform Data with no Quality Control.
Q — Quality Controlled Data, some processes have been applied to the data.
M — Data center modified, time-series values have not been changed.

Constant value "R" is set, since no quality control has been applied.

3 Space (ASCII 32)

Reserved; do not use this byte.

Set as ASCII = 32, blank.

4 Station identifier code

Station identifier designation. Left justify and pad with spaces.

Station identifier is coded on 5 characters representing the data origin. Please refer to SEIS Mini-SEED convention codes document (INS-ST-GRDS-1500-IPGP) for detailed definition of this field value.

5 Location identifier

Location identifier designation. Left justify and pad with spaces.

The location is derived from the FSW configuration file and the SEIS Mini-SEED convention codes configuration file. Please refer to SEIS Mini-SEED convention codes document (INS-ST-GRDS-1500-IPGP) for the detailed definition of this field value..

6 Channel identifier

Channel identifier designation (see Appendix A). Left justify and pad with spaces.

The channel identifier is derived from the FSW configuration file and the SEIS Mini-SEED convention codes configuration file. Please refer to SEIS Mini-SEED convention codes document (INS-ST-GRDS-1500-IPGP) for the detailed definition of this field value.

7 Network

A two characters alphanumeric identifier that uniquely identifies the network operator responsible for the data logger. This identifier is assigned by the IRIS Data Management Center in consultation with the FDSN working group on the SEED format.

The Network name cannot be guessed from the input data, it is an input of the Mini-SEED generation program. Please refer to SEIS Mini-SEED convention codes document (INS-ST-GRDS-1500-IPGP) for the detailed definition of this field value.

8 Record start time

The UTC start time of the record.

L0 products

These are converted from input date (AOBT or LOBT) using the SISMOC Time Correlation Tool. If the input data is spread across several records, the start time of each record is recalculated to minimize the effects of the clock drift.

If a new time information packet is received at the SISMOC, a new revision of the product is created with updated timestamps.

TWINS Miniseed products

The date is read in AOBT field and converted to UTC using the SISMOC Time Correlation Tool. Just like L0 products, if a new time information packet is received at the SISMOC, a new revision of the product is created with updated timestamps.

Scientific Miniseed products

Scientific Miniseed are computed from Time Series format, which gives dates in UTC. No date conversion is performed during the generation of Scientific Miniseed products.

Date details

Date is 10 bytes long with:

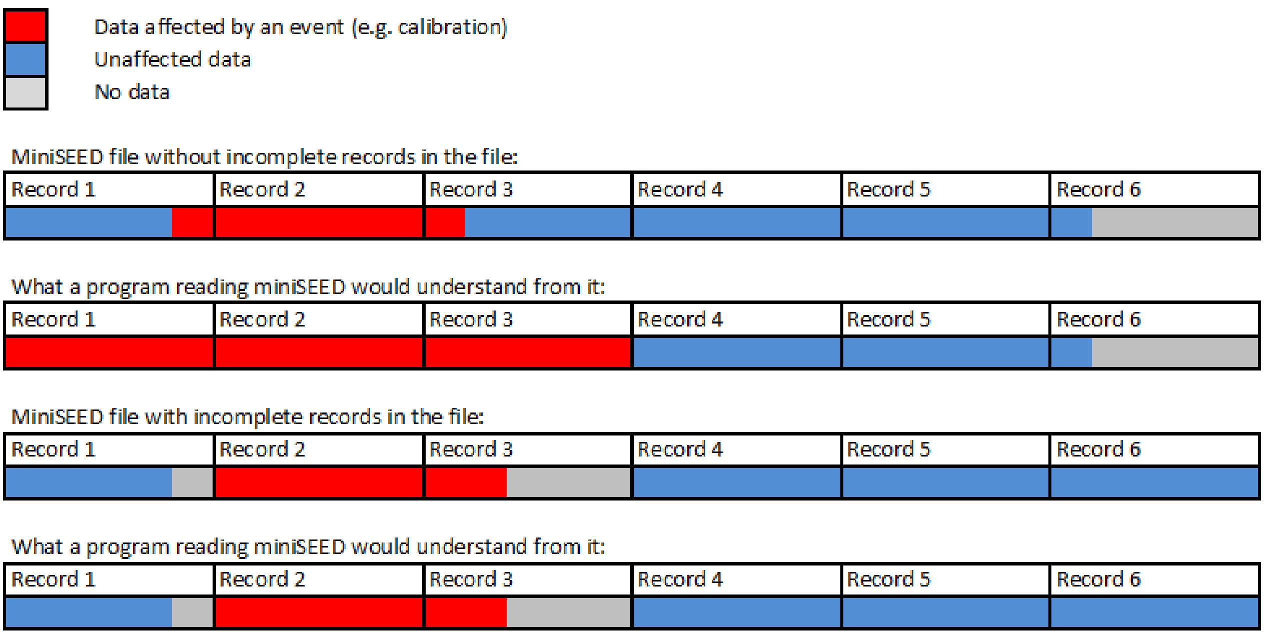
- Bytes 1 and 2: year (big endian format).
- Bytes 3 and 4: day of the year (big endian format). Jan 1 is 1. Values between 1 and 366.
- Byte 5: hours. Values between 0 and 23.
- Byte 6: minutes. Values between 0 and 59.
- Byte 7: seconds. Values between 0 and 59, occasionally 0 to 60 (leap second).
- Byte 8: unused for data (required for alignment).
- Bytes 9 and 10: number of 0.1 millisecond (10^{-4} second).

9 Number of samples in record.

This field contains the number of samples in the current record. Incomplete records data are padded with zeroes.

Incomplete records may occur:

- at the end of the file (last record),
- because a too important clock drift was detected (>1% of the sampling rate),
- because header flags are defined for a whole record, an event may cause incomplete records in the middle of a file to limit the effects of the event flags:



As shown in the above example, without splitting the data records, a great part of the record 1 and 3 would be incorrectly flagged as affected by an event.

10 Sample rate factor

> 0 — Samples/second
< 0 — Seconds/sample
= 0 — Seconds/sample Use this for ASCII/OPAQUE DATA records (see Blockette 2000 description)

The Sample rate field is filled automatically by the underlying library obspy from a given frequency. The frequency is derived from the FSW configuration file.

11 WORD: Sample rate multiplier

>0 — Multiplication factor
<0 — Division factor

Referring to the SEED manual (chapter 8, page 110):

If Sample rate factor > 0 **and** Sample rate Multiplier > 0,
Then nominal Sample rate = Sample rate factor X Sample rate multiplier

If Sample rate factor > 0 **and** Sample rate Multiplier < 0,
Then nominal Sample rate = -1 X Sample rate factor / Sample rate multiplier

If Sample rate factor < 0 **and** Sample rate Multiplier > 0,
Then nominal Sample rate = -1 X Sample rate multiplier / Sample rate factor

If Sample rate factor < 0 **and** Sample rate Multiplier < 0,
Then nominal Sample rate = 1/ (Sample rate factor X Sample rate multiplier)

The Sample multiplier field is filled automatically by the underlying library obspy from a frequency. The frequency is derived from the FSW configuration file.

12 Activity flags

Activity flags are filled independently for each record, depending on the "metadata" information known by the Mini-SEED generation tool at the time it runs. If a new "metadata" information is received, a new revision of the Mini-SEED file is created with updated flags.

[Bit 0] — Calibration signals present

This bit is set to one for the mini-SEED records containing data corresponding to a calibration, i.e. excitation of a sensor with a known signal.

L0 products

Calibration events are known from a SEIS_FSW_SPECIAL_EVR packet with "Sensor status" field indicating a calibration change (bits 6, 4, 2, 0 set to one).

All VBB channels are considered impacted by any VBB calibration. All SP channels are considered impacted by any SP calibration. All other channels are considered unimpacted by any calibration.

TWINS Miniseed products

This field is always set to zero.

Scientific Miniseed products

This field is always set to zero.

[Bit 1] — Time correction applied.

Set this bit to 1 if the time correction in field 16 has been applied to field 8. Set this bit to 0 if the time correction in field 16 has not been applied to field 8.

Always set to 0. Header's field 8 (record start time) is precise enough to store 0.1 milliseconds, UTC-translated clocks (AOBT or LOBT) are not precise enough to go beyond that precision.

[Bit 2] — Beginning of an event, station trigger

L0 Products

Used to flag the following event:

- Recentering occurred on board. Recentering events are known from a SEIS_FSW_SPECIAL_EVR packet with "Sensor status" field indicating a recentering (bits 5, 3, 1 set to one)
- Leveling occurred on board. Leveling events are known from a SEIS_FSW_SPECIAL_EVR packet with field "auxiliary data" set to one.

A recentering event affects all VBB channels. A leveling event affects all BB, SP or SCIT channels.

TWINS Miniseed products

Always set to zero

Scientific Miniseed products

Always set to zero

[Bit 3] — End of the event, station detriggers

Same as bit 2

[Bit 4] — A positive leap second happened during this record (a 61 second minute).

L0 products

The Mini-SEED generation tool uses the standard leap seconds file distributed by the US National Institute of Standards and Technology (available for download at ftp://time.nist.gov/pub/or ftp://tycho.usno.navy.mil/pub/ntp/) to detect leap seconds event and set this flag to 1 if needed.

TWINS Miniseed products

Always set to zero.

Scientific Miniseed products

Always set to zero.

[Bit 5] — A negative leap second happened during this record (A 59 second minute).

A negative leap second clock correction has not yet been used, but the U.S. National Bureau of Standards has said that it might be necessary someday.

Same as bit 4

[Bit 6] — Event in progress

Same as bit 2, in case of an event spread across several records.

13 I/O flags and clock flags

[Bit 0] — Station volume parity error possibly present

Not used for SEIS/APSS Mini-SEED files, always set to 0.

[Bit 1] — Long record read (possibly no problem)

Not used for SEIS/APSS Mini-SEED files, always set to 0.

[Bit 2] — Short record read (record padded)

Not used for SEIS/APSS Mini-SEED files, always set to 0.

[Bit 3] — Start of time series

Not used for SEIS/APSS Mini(SEED files, always set to 0.

[Bit 4] — End of time series

Not used for SEIS/APSS Mini-SEED files, always set to 0.

[Bit 5] — Clock locked

Not used for SEIS/APSS miniSEED files, always set to 0.

14 Data quality flags

[Bit 0] — Amplifier saturation detected (station dependent)

L0 Products

This bit is set if the data is marked as "Invalid" in the telemetry. Invalidity events are known from a SEIS_FSW_SPECIAL_EVR packet with "<Sensor name> validity" field indicating invalid value.

TWINS Miniseed products

This bit is always set to zero.

Scientific Miniseed products

This bit is always set to zero.

[Bit 1] — Digitizer clipping detected

Same as bit 0.

[Bit 2] — Spikes detected

Unused for SISMOC Miniseed production, always set to 0.

[Bit 3] — Glitches detected

Unused for SISMOC Miniseed production, always set to 0.

[Bit 4] — Missing/padded data present

Unused for SISMOC Miniseed production, always set to 0.

[Bit 5] — Telemetry synchronization error

Unused for SISMOC Miniseed production, always set to 0.

[Bit 6] — A digital filter may be charging

Unused for SISMOC Miniseed production, always set to 0.

[Bit 7] — Time tag is questionable

Given the SEED precision and the clocks drifts, this flag is always set to 1.

15 Total number of blockettes that follow

This field is set automatically by obspy. Blockette 1000 and 1001 are always present, so this field has a constant value of 2.

16 Time correction

This field contains a value that may modify the field 8 record start time. Depending on the setting of bit 1 in field 12, the record start time may have already been adjusted. The units are in 0.0001 seconds.

Always set to 0.

17 Offset in bytes to the beginning of data

The first byte of the data records is byte 0.

Given the fixed header length and the 2 blockettes following the header, it is always set to 64.

18 Offset in bytes to the first data blockette in this data record

Enter 0 if there are no data blockettes. The first byte in the data record is byte offset 0.

This field is filled automatically by obspy. Given the fixed header length and the 2 blockettes following the header, it is alwaysset to 48.

NOTE: All unused bits in the flag bytes are reserved and must be set to zero.
The last word defines the length of the fixed header. The next-to-last word fixes the length reserved for the entire header.

6. BLockette 500

Note	Field name	Type	Length	Mask or Flags
1	Blockette type — 500	B	2	
2	Next blockette offset	B	2	
3	VCO correction	B	4	
4	Time of exception	B	10	
5	*μsec*	B	1	
6	Reception Quality	B	1	
7	Exception count	B	4	
8	Exception type	A	16	
9	Clock model	A	32	
10	Clock status	A	128	

1 UWORD : Blockette type [500]: Timing blockette

This field is always set to 500.

2 UWORD : Byte number of next blockette

(Calculate this as the byte offset from the beginning of the logical record - including the fixed section of the data header; use 0 if no more blockettes will follow.)

Since the blockette order is not guaranteed by the underlying library obspy, the value of this field may be 248 (blockette 500 is in first position) or 256 (blockette 500 is between blockette 1001 and 1000) or zero (blockette 500 is in last position).

3 FLOAT : VCO correction

This field is a floating point percentage from 0.0 to 100.0% of VCO control value, where 0.0 is slowest, and 100.0% is fastest.
This field cannot be filled by the underlying library obspy, therefore it is always set to 0.

4 BTIME: Time of exception

Same format as record start time.
This field cannot be filled by the underlying library obspy, therefore it is always set to 0.

Date details

Date is 10 bytes long with:

- Bytes 1 and 2: year (big endian format).
- Bytes 3 and 4: day of the year (big endian format). Jan 1 is 1. Values between 1 and 366.
- Byte 5: hours. Values between 0 and 23.
- Byte 6: minutes. Values between 0 and 59.
- Byte 7: seconds. Values between 0 and 59, occasionally 0 to 60 (leap second).
- Byte 8: unused for data (required for alignment).
- Bytes 9 and 10: number of 0.1 millisecond (10^{-4} second).

5 BYTE: µsec

This field has the clock time down to the microsecond. The SEED format handles down to 100µsecs. This field is an offset from that value. The recommended value is from -50 to +49 µsecs. At the users option, this value may be from 0 to +99 µsecs.
This field cannot be filled by the underlying library obspy, therefore it is always set to 0.

6 UBYTE: Reception quality

This field is a number from 0 to 100% of maximum clock accuracy based only on information from the clock.
This field cannot be filled by the underlying library obspy, therefore it is always set to 0.

7 ULONG: Exception count

This is an integer count, with its meaning based on the type of exception, such as 15 missing timemarks.
This field cannot be filled by the underlying library obspy, therefore it is always set to 0.

8 CHAR*16: Exception type

This field describes the type of clock exception, such as “Missing” or “Unexpected”.
This field cannot be filled by the underlying library obspy, therefore it is always set to 16 space caracter.

9 CHAR*32: Clock model

This field is an optional description of the clock, such as “Quanterra GPS1/QTS”.

L0 products

The value "LOBT" or "AOBT" depending on the clocks on board the spacecraft (AOBT for APSS, LOBT for SEIS).

TWINS Miniseed products

The value "AOBT" corresponding of the clock on board the spacecraft (AOBT for APSS).

Scientific Miniseed products

The value "UTC" from Time Series format, which gives dates in UTC.

10 CHAR*128: Clock status

This field is an optional description of clock specific parameters, such as the station for an Omega clock, or satellite signal to noise ratios for GPS clocks.

L0 products

The LOBT date of the first sample of the record in "centi-ticks" or the AOBT date of the first sample of the record is stored into this field. If present, warning messages during a UTC conversion are stored into this field. A separator is used to differentiate the date of the warning message.

For SEIS 100Hz acquisitions, there are 100 samples every 1024 LOBT ticks (hence 1 sample per 10,24 ticks). So there is no rounding errors (note 1) on the timestamps in LOBT as long as there are given in centi-ticks (1/100th of a tick):
one sample every 1024 centitick.
The LOBT range is 0 to 2^40 -1, which implies that 13+2 digits are required to represent the full range in centiticks.

For APSS 20Hz acquisitions, there are 20 samples and AOBT is reported in seconds (with a range from 0 to 2^32-1) + ms (000 to 999), so there is no rounding errors if AOBT are expressed directly in AOBT (one sample every 50ms).

TWINS Miniseed products

The AOBT date of the first sample of the record. If present, warning messages during a UTC conversion are stored into this field. A separator is used to differentiate the date of the warning message.

For APSS 20Hz acquisitions, there are 20 samples and AOBT is reported in seconds (with a range from 0 to 2^32-

1) + ms (000 to 999), so there is no rounding errors if AOBT are expressed directly in AOBT (one sample every 50ms).

*(note 1) In fact this is not true, because of the FIR EBOX delay is subtracted from the LOBT date in seconds. The FIR EBOX delay is expressed in 1/500s. The date LOBT displayed in blockette 500 is therefore rounded off. The LOBT value is computed as follows: (LOBTsec - FIRDelay) * 1024 * 100.*

Scientific Miniseed products

No date conversion is performed during the generation of Scientific Miniseed products. The default value is set to “999999999999999" (15 digits).

7. Blockette 1000

Note	Field name	Type	Length	Mask or Flags
1	Blockette type — 1000	B	2	
2	Next blockette's byte number	B	2	
3	Encoding Format	B	1	
4	Word order	B	1	
5	Data Record Length	B	1	
6	Reserved	B	1	

1 Blockette type [1000]: Data Only SEED

This field is always set to 1000.

2 Byte number of next blockette

Since the blockette order is not guaranteed by the underlying library obspy, the value of this field may be 56 (blockette 1000 is in first position) or 256 (blockette 1000 is after blockette 500) or 64 (blockette 1000 is after blockette 1001) or zero (blockette 1000 is in last position).

3 A code indicating the encoding format

This number is assigned by the FDSN Data Exchange Working Group. To request that a new format be included contact the FDSN through the FDSN Archive at the IRIS Data Management Center. To be supported in Data Only SEED, the data format must be expressible in SEED DDL. A list of valid codes at the time of publication follows.

CODES 0-9 GENERAL
0 ASCII text, byte order as specified in field 4
1 16 bit integers
2 24 bit integers
3 32 bit integers
4 IEEE floating point
5 IEEE double precision floating point
CODES 10 - 29 FDSN Networks
10 STEIM (1) Compression
11 STEIM (2) Compression
12 GEOSCOPE Multiplexed Format 24 bit integer
13 GEOSCOPE Multiplexed Format 16 bit gain ranged, 3 bit exponent
14 GEOSCOPE Multiplexed Format 16 bit gain ranged, 4 bit exponent
15 US National Network compression
16 CDSN 16 bit gain ranged
17 Graefenberg 16 bit gain ranged
18 IPG - Strasbourg 16 bit gain ranged
19 STEIM (3) Compression
CODES 30 - 39 OLDER NETWORKS
30 SRO Format
31 HGLP Format
32 DWWSSN Gain Ranged Format
33 RSTN 16 bit gain ranged

L0 products

The science packets are first decompressed using the SEIS specific compression algorithm (STEIM-like algorithm), and then compressed using STEIM2.
The housekeeping packets are converted to STEIM2 data records. This field has therefore a constant value of 11.

TWINS Miniseed products

All TWINS data store floating point numbers, but STEIM2 is only capable of storing integers. To ensure all precision is kept, TWINS Miniseed products are stored as uncompressed double precision floating point numbers. This field has therefore a constant value of 5.

Scientific Miniseed products

Same as TWINS Miniseed products.

4 Word order

The byte swapping order for 16 bit and 32 bit words.
A 0 indicates little-endian order and a 1 indicates big-endian word order. See fields 11 and 12 of blockette 50.

Big-endian is used. This field is always set to 1.

5 The exponent (as a power of two) of the record length for these data.

The data record can be as small as 256 bytes and, in Data Only SEED format as large as 2 raised to the 256 power.

The data record length is fixed to 512 bytes for all SEIS/APSS Mini-SEED files. Therefore this field will have a constant value of 9.

6 Reserved

This field is padded with 0.

8. Blockette 1001

Note	Field name	Type	Length	Mask or Flags
1	Blockette type — 1001	B	2	
2	Next blockette's byte number	B	2	
3	Timing quality	B	1	
4	µsec	B	1	
5	Reserved	B	1	
6	Frame count	B	1	

1 UWORD: Blockette type [1001]: Data extension blockette

This field is always set to 1001

2 UWORD: Byte number of next blockette.

Since the blockette order is not guaranteed by the underlying library obspy, the value of this field may be 56 (blockette 1001 is in first position) or 256 (blockette 1001 is after blockette 500) or 64 (blockette 1001 is after blockette 1000) or zero (blockette 1001 is in last position).

3 UBYTE: Timing quality

Timing quality is a vendor specific value from 0 to 100% of maximum accuracy, taking into account both clock quality and data flags.

This field is set to constant value 99, in order to indicate that on the data record, data are dated with a drift lower than 1% of the sampling rate.

4 BYTE: µsec has the data start time down to the microsecond.

The SEED format handles down to 100µsecs. This field is an offset from that value. The recommended value is from -50 to +49µsecs. At the users option, this value may be from 0 to +99µsecs.

Time conversion precision shall not be precise enough to reach the µsec for SEIS/APSS clocks. Therefore this field shall be set to 0.

5 Reserved byte.

This field is padded with 0.

6 UBYTE: Frame count

Frame count is the number of 64 byte compressed data frames in the 4K record (maximum of 63). Note that the user may specify fewer than the maximum allowable frames in a 4K record to reduce latency

This field cannot be filled by the underlying library obspy, therefore it is always set to 0.