

# Multiscale Electrophysiology File Format

## Multiscale Electrophysiology File:

- Contains EEG data of a single channel in lossless compressed, optionally encrypted format.
- Identified with the “.mef” file extension.
- EEG data are written in compressed, variable-length blocks.
- The file contains a header, EEG data, and block indices section
- The block indices section contains triplets of times (uUTC time - see below), file offsets, and sample indices of the EEG data in the file.

MEF File Structure
Header
EEG Data
Block Indices

## Session/Event File (XML):

- Contains session information and event records associated with sample times.
- Identified with the “.xml” file extension.
- There is one event file for all channels.
- Example record types include:
  - Video file synchronization data
  - Spike records
  - Seizure markers
  - Event related study data
  - Sleep stage / behavioral state
  - Miscellaneous notes

**Data Type Definitions:**

Type Name	Description
ui1	1 byte unsigned integer
si1	1 byte signed integer
ui2	2 byte unsigned integer
si2	2 byte signed integer
si3	3 byte signed integer, range $-2^{23}$ to $+(2^{23} - 1)$ : In two's complement format: sign extend the most significant bit to create an si4.
ui4	4 byte unsigned integer
si4	4 byte signed integer
sf4	4 byte signed floating point number
ui8	8 byte unsigned integer
si8	8 byte signed integer
sf8	8 byte signed floating point number
\$(n)	zero-terminated string of length "n" bytes (not including terminal zero)

**Header Encryption:**

- The header begins with a series of unencrypted bytes, including two text fields and a series of numeric values defining the file's format and characteristics.
- The remainder of the header can be encrypted with "subject" & "session" passwords. Encryption is not required, and the subject and session encryptions can be used together or individually. If both encryptions are used, the session password is stored in the subject-encrypted header block.
- The passwords are zero-terminated strings with a maximum 15 character limit.
- The subject password is used to encrypt subject identifying information and (if session encryption is used also) access the session password stored in the header for session decryption.
- The session password decrypts all technical information related to the EEG recording session.
- The encryption / decryption algorithm is the 128-bit Advanced Encryption Standard (AES). [ <http://www.csrc.nist.gov/publications/fips/fips197/fips-197.pdf> ], which satisfies the Health Insurance Portability and Accountability Act (HIPAA) 112-bit requirement for symmetric encryption of human data.

**Header Alignment:**

- Fields in the header have required byte alignments relative to its start.

- 16-byte alignment facilitates encryption/decryption beginning at that offset.
- Other alignment requirements are determined by the data-types: e.g. 8-byte alignment facilitates reading si8, ui8, and sf8 data types.

## Header Version 2.0

Field	Offset	Size	Type	Contents	Al- lign- ment	En- crypt- tion
Institution	0	64	\$(63)	institution	1	None
Unencrypted Text Field	64	64	\$(63)	unencrypted text field (general use)	1	None
Encryption Algorithm	128	32	\$(31)	"128-bit AES"	1	None
Subject Encryption Used	160	1	ui1	1 if subject encryption used, 0 if not	1	None
Session Encryption Used	161	1	ui1	1 if session encryption used, 0 if not	1	None
Data Encryption Used	162	1	ui1	1 if session encryption applied to statistical model in block header, 0 if not	1	None
Byte Order Code	163	1	ui1	0 ==> big-endian 1 ==> little-endian	1	None
Header Major Version	164	1	ui1	numeric value: 2	1	None
Header Minor Version	165	1	ui1	numeric value: 0	1	None
Header Length	166	2	ui2	length of header in bytes	2	None
Session Unique Identifier	168	8	ui1	8 numeric values (0-255) that are shared by all mef, and event files representing a particular recording session (zeroes if not entered)	1	None

Field	Offset	Size	Type	Contents	Al- lign- ment	En- cryp- tion
Subject First Name	176	32	\$(31)	subject first name	16	Sub- ject
Subject Middle Name	208	32	\$(31)	subject middle name	1	Sub- ject
Subject Last Name	240	32	\$(31)	subject last name	1	Sub- ject
Subject ID	272	32	\$(31)	subject ID	1	Sub- ject
Session Pass- word	304	16	\$(15)	session password (15 character limit)	1	Sub- ject
Subject Pass- word Validation Field	320	16	ui1	Pascal-style string en- coding subject pass- word, terminal unused bytes random	16	Sub- ject
Protected Re- gion	336	16		discretionary	16	un- speci- fied
Session Pass- word Validation Field	352	16	ui1	Pascal-style string en- coding session pass- word, terminal unused bytes random	16	Ses- sion
Number of En- tries	368	8	ui8	total recorded samples in file	8	Ses- sion
Channel Name	376	32	\$(31)	channel name	1	Ses- sion
Recording Start Time	408	8	ui8	time in uUTC time for- mat (see below)  0 indicates no entry	8	Ses- sion
Recording End Time	416	8	ui8	time in uUTC time for- mat (see below)  0 indicates no entry	8	Ses- sion

Field	Offset	Size	Type	Contents	Al- lign- ment	En- cryp- tion
Sampling Fre- quency	424	8	sf8	sampling frequency -1 indicates no entry	8	Ses- sion
Low Fre- quency Filter Setting	432	8	sf8	high-pass filter setting -1 indicates no entry	8	Ses- sion
High Fre- quency Filter Setting	440	8	sf8	low-pass filter setting -1 indicates no entry	8	Ses- sion
Notch Filter Frequency	448	8	sf8	notch filter setting  0 indicates no notch filter  -1 indicates no entry	8	Ses- sion
Voltage Con- version Factor	456	8	sf8	microvolts per sample unit  0 indicates no entry  negative values indi- cate voltage values are inverted	8	Ses- sion
Acquisition System	464	32	\$(31)	name of acquisition system	1	Ses- sion
Channel Comments	496	128	\$(127)	channel comments	1	Ses- sion
Study Com- ments	624	128	\$(127)	study comments	1	Ses- sion
Physical Channel Num- ber	752	4	si4	physical channel num- ber during acquisition  -1 indicates no entry	4	Ses- sion
Compression Algorithm	756	32	\$(31)	“RED 1.0” (range en- coded differences)	1	Ses- sion

Field	Offset	Size	Type	Contents	Al- lign- ment	En- cryp- tion
Maximum Compressed Block Size	788	4	ui4	Maximum bytes in compressed block (including block header)	4	Ses- sion
Maximum Block Length	792	8	ui8	Maximum number of samples in a decompressed block	8	Ses- sion
Block Interval	800	8	ui8	contains microseconds between blocks  0 indicates variable block intervals	8	Ses- sion
Maximum Data Value	808	4	si4	The largest data value in the file	4	Ses- sion
Minimum Data Value	812	4	si4	The smallest data value in the file	4	Ses- sion
Offset to Block Indices Data	816	8	ui8	Offset to start of block indices  Block indices are stored at the end of the mef file with 8-byte alignment	8	Ses- sion
Number of Block Index Entries	824	8	ui8	Total number of entries (triplets) in index data block	8	Ses- sion
Block Header Length	832	2	ui2	length of encoded data block header in bytes	2	Ses- sion
Unused	834	190	ui1	random bytes	16	None
EEG Data Start	1024			RED encoded data blocks	1	None

## Micro-UTC (uUTC) Time Format

- ui8 containing the elapsed microseconds since January 1, 1970 at 00:00:00 in the GMT (Greenwich, England) time zone.

- Simply converted to UTC time format (seconds since 1/1/1970 at 00:00:00 GMT)

## Multiscale Electrophysiology File Data Format

- Data are stored in compressed blocks, compressed with the algorithm specified in the header. In the current version this is the RED (range encoded differences) compression algorithm.
- The time interval of the blocks is specified in the block interval field of the header.
- Each data block contain a small header detailed by the compression algorithm, and whose size is specified the block header length field of the file header.
- Each block is indexed by the block indices for random access.

## RED Data Compression Format

- Data are stored in compressed independent blocks
- Raw data are differenced. Differences are encoded in a single signed byte. If there is overflow, i.e.  $> +127$  or  $< -127$ , then a key sample is introduced flagged by the reserved value -128. The three bytes following the key sample flag contain the value of the second data point generating the overflow difference as an si3.
- The differenced data are statistically modeled, the model is stored in the block header.
- Range encoding is used to compress the differences, using the statistical model.
- Blocks are required to be 8-byte boundary aligned.

## RED Data Compression Block Format

Field	Size (bytes)	Type	Contents
Cyclically Redundant Checksum	4	ui4	Checksum detects data corruption within the block header and data block
Compressed Block Length	4	ui4	Number of bytes in the compressed block, including block header & boundary alignment bytes added at the end
Block Start Time	8	ui8	uUTC time
Difference Length	4	ui4	Difference data length in bytes
Block Length	4	ui4	Number of data samples encoded in the block

Field	Size (bytes)	Type	Contents
Maximum Data Value	3	si3	The maximum raw value (not difference) encoded in the data block
Minimum Data Value	3	si3	The minimum raw value (not difference) encoded in the data block
Discontinuity Flag	1	ui1	0 indicates no discontinuity, 1 indicates that this block began after a discontinuity in recording, or is the first block in a file.
Block Statistics	256	ui1	Statistical model of difference values for the block. Session password may be used to encrypt this field
Compressed Data	varies	si1	Encoded data

## Block Indices Format

- uUTC time, followed by file offset in bytes, followed by sample number.
- Stored at end of EEG data
- 8-byte boundary aligned
- The offset points to the first byte of a compressed block header in the EEG data.

Field	Offset (bytes)	Size (bytes)	Type	Contents
Sample Time	0	8	ui8	uUTC time
File Offset	8	8	ui8	File offset in bytes, including header bytes
Sample Index	16	8	ui8	Index of sample in data file. First sample index is zero.



## Session/Event File Schema

- Transitional file containing information relevant to the acquisition, analysis and persistent storage of EEG annotations.
- XML chosen for flexibility, and general acceptance.
- XML formatted data are considered transient storage.
- Long-term (i.e., “persistent”) storage is handled by a database.
- Database import facilitated by use of XML.
- Custom events and notations are be defined.
- File easily customized to needs of experiment and lab.

## Session/Event File Format

Element	Tag	Contents
XML Declaration	<?xml version="1.0" encoding="UTF-8"?>	None
Subject Information	<subject>	Any subject-related information that may be persisted. This information is encoded according to the same rules governing MEF header encoding.
Session Information	<session>	Session UID, directory containing channel files, channel file names, session start time
Annotations	<annotations>	Computer- or user-generated (label-timestamp) pairs.

## Subject Information

Sub-elements and syntax:

Element	Tag	Contents
Subject First Name	<field name = "name_first">	Subject's first name.
Subject Middle Name	<field name = "name_middle">	Subject's middle name.
Subject Last Name	<field name = "name_last">	Subject's last name.

Element	Tag	Contents
Subject ID	<field name = "id">	Subject's identification number.

## Session Information

### Sub-elements and syntax:

Element	Tag	Contents
Institution	<field name = "institution">	Institution where recordings occurred.
Unencrypted Text Field	<field name = "recording_comment">	Comment entered when recordings occurred.
Session Unique ID	<field name = "uid">	Eight-integer, unique ID code separated by decimal points.
Session Recording Start Time	<field name = "start_time" units="uUTC">	Beginning of recording session in uUTC time
Session Recording End Time	<field name = "end_time" units="uUTC">	Ending of recording session in uUTC time
Data Directory	<field name = "data_directory">	Local directory containing channel (.mef) data files.
Data File	<field name = "data_file">	Name of file containing channel data.

## Annotations

Element	Tag	Contents
Entry	<entry>	Entry contains 0 or more time points and labels.
<field_name>	<field name="..."> value</field>	Child of an Entry containing a name, 0 or more attributes and one value.

### Example XML Session/Event File:

```
<?xml version="1.0" encoding="UTF-8"?>
<subject>
  <field name="name_first"> John </field>
  <field name="name_middle" />
  <field name="name_last"> Doe </field>
  <field name="id"> 0-000-000 </field>
</subject>
<session>
  <field name="institution"> Mayo Systems Electrophysiology Laboratory </field>
  <field name="recording_comment"> intracranial monitoring using the
    Stead-Worrell hybrid electrode for pre-surgical evaluation </field>
  <field name="uid"> 0.32.64.96.128.160.192.255 </field>
  <field name="start_time" units="uUTC"> 2450928598500 </field>
  <field name="end_time" units="uUTC"> 2450928598500 </field>
  <field name="data_directory"> /Volumes/data/session_1 </field>
  <field name="data_file"> channel_1.mef </field>
  <field name="data_file"> channel_2.mef </field>
  <field name="data_file"> channel_3.mef </field>
</session>
<annotations>
  <entry name="behavioral_state">
    <field name="time" units="uUTC"> 2450919644106 </field>
    <field name="description"> Awake </field>
  </entry>
  <entry name="video_sync">
    <field name="start" units="uUTC"> 2450926598801 </field>
    <field name="end" units="uUTC"> 2450939594720 </field>
    <field name="video_file"> onset.mpg </field>
    <field name="format"> MPEG </field>
    <field name="camera_view"> frontal </field>
  </entry>
  <entry name="seizure">
    <field name="onset" units="uUTC"> 2450928598700 </field>
    <field name="offset" units="uUTC"> 2450958598700 </field>
  </entry>
  <entry name="annotation">
    <field name="time" units="uUTC"> 2450928598900 </field>
    <field name="text"> left arm jerking </field>
  </entry>
</annotations>
```