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 ".maf" 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	Alignment	Encryption
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

Field	Offset	Size	Type	Contents	Alignment	Encryption
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
Subject First Name	176	32	$(31)	subject first name	16	Subject
Subject Middle Name	208	32	$(31)	subject middle name	1	Subject
Subject Last Name	240	32	$(31)	subject last name	1	Subject
Subject ID	272	32	$(31)	subject ID	1	Subject
Session Password	304	16	$(15)	session password (15 character limit)	1	Subject

Field	Offset	Size	Type	Contents	Alignment	Encryption
Subject Password Validation Field	320	16	ui1	Pascal-style string encoding subject password, terminal unused bytes random	16	Subject
Protected Region	336	16		discretionary	16	unspecified
Session Password Validation Field	352	16	ui1	Pascal-style string encoding session password, terminal unused bytes random	16	Session
Number of Entries	368	8	ui8	total recorded samples in file	8	Session
Channel Name	376	32	$(31)	channel name	1	Session
Recording Start Time	408	8	ui8	time in uUTC time format (see below) 0 indicates no entry	8	Session
Recording End Time	416	8	ui8	time in uUTC time format (see below) 0 indicates no entry	8	Session
Sampling Frequency	424	8	sf8	sampling frequency -1 indicates no entry	8	Session

Field	Offset	Size	Type	Contents	Alignment	Encryption
Low Frequency Filter Setting	432	8	sf8	high-pass filter setting -1 indicates no entry	8	Session
High Frequency Filter Setting	440	8	sf8	low-pass filter setting -1 indicates no entry	8	Session
Notch Filter Frequency	448	8	sf8	notch filter setting 0 indicates no notch filter -1 indicates no entry	8	Session
Voltage Conversion Factor	456	8	sf8	microvolts per sample unit 0 indicates no entry negative values indicate voltage values are inverted	8	Session
Acquisition System	464	32	$(31)	name of acquisition system	1	Session
Channel Comments	496	128	$(127)	channel comments	1	Session
Study Comments	624	128	$(127)	study comments	1	Session

Field	Offset	Size	Type	Contents	Alignment	Encryption
Physical Channel Number	752	4	si4	physical channel number during acquisition -1 indicates no entry	4	Session
Compression Algorithm	756	32	$(31)	“RED 1.0” (range encoded differences)	1	Session
Maximum Compressed Block Size	788	4	ui4	Maximum bytes in compressed block (including block header)	4	Session
Maximum Block Length	792	8	ui8	Maximum number of samples in a decompressed block	8	Session
Block Interval	800	8	ui8	contains microseconds between blocks 0 indicates variable block intervals	8	Session
Maximum Data Value	808	4	si4	The largest data value in the file	4	Session
Minimum Data Value	812	4	si4	The smallest data value in the file	4	Session
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	Session

Field	Offset	Size	Type	Contents	Alignment	Encryption
Number of Block Index Entries	824	8	ui8	Total number of entries (triplets) in index data block	8	Session
Block Header Length	832	2	ui2	length of encoded data block header in bytes	2	Session
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 keysample is introduced flagged by the reserved value -128. The three bytes following the keysample 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
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

Field	Size (bytes)	Type	Contents
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.

Multiscale Annotation Format (MAF) XML 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 can be defined.
- File easily customized to needs of experiment and lab.

Event File Format

Element	Tag	Contents
XML Declaration	<?xml version="1.0" encoding="UTF-8"?>	None
XREDE Document Declaration	<XREDE>	Encompasses all subject, annotation, and channel information
Dataset	<Dataset>	Identifies individual datasets within the MAF file
Subject Information	<Subject>	Any subject-related information that may be persisted.
Episode (Session) Information	<Episode>	Any information pertaining to the recording session that may be persisted
Task	<Task>	Identifies the source of annotations

Subject Information

Syntax: <Subject [parameters]> … </Subject>

Element	Tag	Contents
Subject First Name	`name_first="Firstname"`	Subject's first name.
Subject Middle Name	`name_middle="Middlename"`	Subject's middle name.
Subject Last Name	`name_last="Lastname"`	Subject's last name.
Subject ID Number	`Subject_nbr="#####"`	Subject's identification number.
Data Directory	`data_dir = "/path/"`	Local directory containing MEF channels
Dataset ID	`DatasetID="#"`	Identifies dataset within MAF file to which subject information pertains.

Episode Information

Syntax: <Episode [parameters]> … </Episode>

Element	Tag	Contents
Institution	`institution = "name"`	Institution where recordings occurred.
Session Unique ID	`uid = "0.0.0.0.0.0.0.0"`	Eight-integer, unique ID code separated by decimal points.
Session Recording Start Time	`recording_start_time = "1145095591430062"`	Beginning of recording session
Time Units	`time_units = "uUTC"`	Units in which recording start time and other time notations are expressed

Element	Tag	Contents
Dataset ID	`DatasetID="#"`	Identifies dataset within MAF file to which episode information pertains.
Subject ID	`SubjectID="#"`	Identifies subject within MAF file to which episode information pertains.

Event Annotations

Syntax: <Event [parameters]> … </Event>

Element	Tag	Contents
Event type	`type = “event_type”`	Describes the type of event in the current annotation
Episode ID	`EpisodeID="#"`	Identifies episode within MAF file to which event information pertains.
Task ID	`TaskID="#"`	Identifies task within MAF file to which event information pertains.

Timestamps

Syntax: <Timestamp [parameters]/>

Element	Tag	Contents
Onset	`onset = “1082190114028809”`	Gives the onset, or start, of the timestamp, in the time units denoted in the episode tag.
Offset	`offset = “1082190114028809”`	Gives the offset, or end, of the timestamp, in the time units denoted in the episode tag.

Element	Tag	Contents
Vector	`vector = “1082190114028809, 1082190119119348, 1082190132921644”`	Vectors are stored as type-specific information followed by comma-separated values.
Event ID	`EventID="#"`	Identifies event to which time-stamps information pertains.
Source ID	`SourceID="#"`	Identifies data source within MAF file to which event information pertains.

Source Information

Syntax: <Source [parameters]\>

Element	Tag	Contents
Name	`name = “channel1.mef”`	Name of MEF file
Label	`label = “channel1”`	Label used to refer to the current channel
Episode ID	`EpisodeID="#"`	Identifies episode within MAF file to which source belongs.

Task Information

Syntax: <Task [parameters]\>

Element	Tag	Contents
Name	`name = “task_info”`	Description or name of task linked to current Dataset.
Dataset ID	`DatasetID="#"`	Identifies dataset within MAF file to which task information pertains.

Example XML Session/Event File:

```
<?xml version="1.0" encoding="UTF-8"?>
<XREDE>
   <Dataset id="1">
      <Subject DatasetID="1" Subject_nbr="9-999-001" data_dir="/Volumes/Server/EEG_Data/Patient_1/" id="1"
         name_first="Firstname" name_last="Lastname">
         <Episode SubjectID="1" id="1" recording_start_time="1081883637196616" time_units="uUTC">
            <Event EpisodeID="1" TaskID="1" id="1" type="seizure">
               <Timestamp EventID="1" SourceID="1" id="1" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="2" id="2" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="3" id="3" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="4" id="4" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="5" id="5" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="6" id="6" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="7" id="7" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="8" id="8" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="9" id="9" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="10" id="10" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="11" id="11" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="12" id="12" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="13" id="13" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="14" id="14" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="15" id="15" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="17" id="16" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="18" id="17" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="19" id="18" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="20" id="19" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="21" id="20" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="22" id="21" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="23" id="22" offset="1082190132044160" onset="1082190114028809" />
               <Timestamp EventID="1" SourceID="24" id="23" offset="1082190132044160" onset="1082190114028809" />
            </Event>
            <Event EpisodeID="1" TaskID="1" id="10" type="Note: Patient pressed call button">
               <Timestamp EventID="10" id="25" onset="1082190116117843" />
            </Event>
            <Event EpisodeID="1" TaskID="1" id="10" type="artifact">
               <Timestamp EventID="10" id="26" onset="1082190605119412" />
            </Event>
            <Event EpisodeID="1" TaskID="1" id="10" type="spike">
               <Timestamp EventID="1" SourceID="18" id="27" onset="1082190674122539" />
               <Timestamp EventID="1" SourceID="19" id="28" onset="1082190674122539" />
               <Timestamp EventID="1" SourceID="22" id="29" onset="1082190674122539" />
               <Timestamp EventID="1" SourceID="23" id="30" onset="1082190674122539" />
            </Event>
            <Source EpisodeID="1" id="1" label="LAG1" name="LAG1.mef" />
            <Source EpisodeID="1" id="2" label="LAG2" name="LAG2.mef" />
            <Source EpisodeID="1" id="3" label="LAG3" name="LAG3.mef" />
            <Source EpisodeID="1" id="4" label="LAG4" name="LAG4.mef" />
            <Source EpisodeID="1" id="5" label="LAG5" name="LAG5.mef" />
            <Source EpisodeID="1" id="6" label="LAG6" name="LAG6.mef" />
            <Source EpisodeID="1" id="7" label="LAG7" name="LAG7.mef" />
            <Source EpisodeID="1" id="8" label="LAG8" name="LAG8.mef" />
            <Source EpisodeID="1" id="9" label="LAG9" name="LAG9.mef" />
            <Source EpisodeID="1" id="10" label="LAG10" name="LAG10.mef" />
            <Source EpisodeID="1" id="11" label="LAG11" name="LAG11.mef" />
            <Source EpisodeID="1" id="12" label="LAG12" name="LAG12.mef" />
            <Source EpisodeID="1" id="13" label="LAG13" name="LAG13.mef" />
            <Source EpisodeID="1" id="14" label="LAG14" name="LAG14.mef" />
            <Source EpisodeID="1" id="15" label="LAG15" name="LAG15.mef" />
            <Source EpisodeID="1" id="16" label="LAG16" name="LAG16.mef" />
            <Source EpisodeID="1" id="17" label="LAG17" name="LAG17.mef" />
            <Source EpisodeID="1" id="18" label="LAG18" name="LAG18.mef" />
            <Source EpisodeID="1" id="19" label="LAG19" name="LAG19.mef" />
            <Source EpisodeID="1" id="20" label="LAG20" name="LAG20.mef" />
            <Source EpisodeID="1" id="21" label="LAG21" name="LAG21.mef" />
```

```
            <Source EpisodeID="1" id="22" label="LAG22" name="LAG22.mef" />
            <Source EpisodeID="1" id="23" label="LAG23" name="LAG23.mef" />
            <Source EpisodeID="1" id="24" label="LAG24" name="LAG24.mef" />
            <Source EpisodeID="1" id="25" label="LAS1" name="LAS1.mef" />
            <Source EpisodeID="1" id="26" label="LAS2" name="LAS2.mef" />
            <Source EpisodeID="1" id="27" label="LAS3" name="LAS3.mef" />
            <Source EpisodeID="1" id="28" label="LAS4" name="LAS4.mef" />
            <Source EpisodeID="1" id="29" label="LAS5" name="LAS5.mef" />
            <Source EpisodeID="1" id="30" label="LAS6" name="LAS6.mef" />
            <Source EpisodeID="1" id="31" label="LAS7" name="LAS7.mef" />
            <Source EpisodeID="1" id="32" label="LAS8" name="LAS8.mef" />
          </Episode>
        </Subject>
        <Task DatasetID="1" id="1" name="user annotations" />
    </Dataset>
</XREDE>
```