Clinical Neurophysiology 114 (2003) 1755–1761
www.elsevier.com/locate/clinph

European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data

Bob Kemp[a,c,*], Jesus Olivan[b]

[a]Leiden University Medical Centre, Department of Neurology, Leiden, The Netherlands
[b]Hospital Carlos III, Clinical Neurophysiology Unit, Madrid, Spain
[c]Medical Centre Haaglanden, Sleep Centre, P.O. Box 432, NL-2501 CK Den Haag, The Netherlands

Accepted 15 April 2003

Abstract

The European data format (EDF) is a widely accepted standard for exchange of electroencephalogram and polysomnogram data between different equipment and labs. But it hardly accommodates other investigations.

EDF+ is a more flexible but still simple format which is compatible to EDF except that an EDF+ file may contain interrupted recordings. Also, EDF+ supports time-stamped annotations for the storage of events such as text annotations, stimuli, averaged signals, electrocardiogram parameters, apnoeas and so on.

When compared to EDF, EDF+ can not only store annotations but also electromyography, evoked potentials, electroneurography, electrocardiography and many more types of investigations. Further improvements over EDF include the use of standard electrode names. EDF+ is so much like EDF that existing EDF viewers still display the signals in EDF+ files. Software development is limited mainly to implementing the annotations.

EDF+ offers a format for a wide range of neurophysiological investigations which can become a standard within a few years.

Keywords: Quantitative analysis European data format

1. Introduction

Different manufacturers of neurophysiological equipment and software apply different file formats for their data. This hinders collaboration, consultation and archival. This problem has been solved for electroencephalogram (EEG) and Sleep recordings by a standard format (Kemp et al., 1992), also known as European data format (EDF). EDF is the only digital format that has become widely accepted in commercially available equipment and in (multi-center) scientific studies (Kemp, 2002). Its main limitation is that it does not support events and annotations and cannot handle interrupted recordings such as made in electromyography (EMG), ElectroNeurography and Evoked Potentials (ENMGEP). We wanted to develop a new standard that meets these limitations, while still maintaining EDF compatibility.

The main difficulty is that EDF recordings can not be interrupted, while ENMGEP studies often record data discontinuously. So, we simply allowed interrupted recordings but kept all other specifications of EDF intact. In this way, existing EDF software will still work, though it will treat interrupted recordings as if they were continuous. And we added a few EDF-compatible extra's, in particular standard electrode names and time-stamped annotations for the archival of events and annotations. Details of the specification and further improvements were discussed in the year 2002 in Yahoo's EDF users group. The result is EDF+.

2. The EDF+ protocol

EDF+ is based on EDF (Kemp et al., 1992). In order to understand EDF+, one must be familiar with EDF first. Section 2.1 describes how EDF+ differs from EDF. Section 2.2 describes how one of the EDF+ signals can be specially

* Corresponding author. Tel.: +31-70-330-2205/71-526-2188; fax: +31-70-388-2636.
E-mail address: bk@hsr.nl (B. Kemp).

1388-2457/03/$30.00
doi:10.1016/S1388-2457(03)00123-8

coded in order to store text annotations, clock time, events and stimuli. Section 2.3 describes how analysis results derived from original recordings can be stored, and also a filename convention for such derived data. Finally, Section 2.4 specifies some standard texts to be used in EDF+. Throughout this article, square borders around text indicate that this text is represented as US-ASCII encoded characters in the file. For instance, [ab2] means that the file contains the 3 subsequent characters 'a', 'b' and '2' represented (according to US-ASCII) by their successive byte values 97, 98 and 50.

EDF+ files must have .edf or .EDF as filename extension.

2.1. EDF+ and EDF

A standard EDF file consists of a header record followed by data records. The header record identifies the patient and specifies the technical characteristics of the recorded signals. The data records contain consecutive fixed-duration epochs of the recorded signals. A standard EDF+ file also consists of a header record followed by data records. The structure of these records is compatible to EDF but contains additional specifications as described in the following paragraphs.

2.1.1. The EDF+ header

The EDF+ header record identifies the patient and specifies the technical characteristics of the recorded signals exactly according to the EDF specs, except for the first 'reserved' field whose 44 characters must start with [EDF+C] if the recording is uninterrupted, thus having contiguous data records, i.e. the starttime of each data record coincides with the end (starttime + duration) of the preceding one. In this case, the file is EDF compatible and the recording ends (number × duration) seconds after its startdate/time. The 'reserved' field must start with [EDF+D] if the recording is interrupted, so not all data records are contiguous. In both cases, the time must be kept in each data record as specified in Section 2.2.4.

The only incompatibility with EDF is, that signals may be recorded discontinuously. Therefore, we have decided that the EDF+ 'version' field must still read [0   ] like in EDF. In this way, existing EDF viewers will still work and display EDF+ files (be they continuous or discontinuous) as continuous EDF files. EDF+ software will know the difference between continuous and discontinuous files from the mentioned 'reserved' field.

2.1.2. The EDF+ data records

A signal in an EDF data record is a series of 2-byte samples, the subsequent samples representing subsequent integer values of that signal, sampled with equal time intervals. We will refer to this kind of signal as an 'ordinary signal' from now on. Such ordinary signals are in the data records exactly according to the EDF specs (including the size limit of 61,440), except for the fact that the data records may unconditionally be shorter than 1s, and subsequent data records need not form a continuous recording. However, as in EDF, data records that follow up in time must also follow up in the file. The samples of an ordinary signal must have equal sample intervals inside each data record, but the interval to the first sample of the next data record may be different.

For instance, in a motor nerve conduction study with a number of stimuli, each data record would hold the ordinary signals corresponding to one stimulus. In this case, the duration of a data record corresponds to the 'window size' in ENMGEP studies.

Of course, in the extreme case in which each ordinary signal only occupies one sample in each data record, while the file is discontinuous, then the duration (and any derived sampling frequency) makes no sense. The same is true if the EDF+ file does not contain any ordinary signal but only time-stamped events (see Section 2.2 and an example in Section 3). In such cases, specify the 'duration of a data record' to be [0   ].

2.1.3. Additional specifications in EDF+

These specifications clarify and tighten a few rather loose definitions of EDF, thus avoiding any ambiguity and making the life of the programmers more easy. As is custom in EDF applications, EDF+ must store the ordinary signal samples (2-byte two's complement integers) in 'little-endian' format, that is the least significant byte first. This is the default format in PC applications. And in the EDF+ header:

1. Use only printable US-ASCII characters with byte values 32 . . 126.
2. The 'startdate' and 'starttime' fields should contain only characters 0–9, and the period (.) as a separator, for example [02.08.51]. In the 'startdate', use 1985 as a clipping date in order to avoid the Y2K problem. So, the years 1985–1999 must be represented by yy = 85–99 and the years 2000–2084 by yy = 00–84. After the year 2084, yy must be 'yy' which implies that only item 4 below defines the correct year.
3. The 'local patient identification' field must start with the following subfields in this order (subfields must not contain, but are separated by, spaces):
– the code by which the patient is known in the hospital administration.
– sex (English, so [F] or [M]).
– birthdate in dd-MMM-yyyy format using the English 3-character abbreviations of the month in capitals. [02-AUG-1951] is OK, while 2-AUG-1951 is not.
– the patients name.
Any space inside the hospital code or the name of the patient must be replaced by a different character, for instance an underscore. As an example, the 'local patient identification' field could start with [MCH-0234567 F 02-MAY-1951 Haagse_Harry]. Subfields whose contents are unknown, not applicable or

must be made anonymous are replaced by a single character X . Additional subfields may follow the ones described here.
4. The 'local recording identification' field must start with the following subfields in this order (subfields must not contain, but are separated by, spaces):
– The text Startdate .
– The startdate itself in dd-MMM-yyyy format using the English 3-character abbreviations of the month in capitals.
– The hospital administration code of the investigation, i.e. EEG number or polysomnography (PSG) number.
– A code specifying the responsible investigator or technician.
– A code specifying the used equipment.
Any space inside any of these codes must be replaced by a different character, for instance an underscore. As an example, the 'local recording identification' field could start with: Startdate 02-MAR-2002 PSG-1234/2002 NN Telemetry03 . Subfields whose contents are unknown, not applicable or must be made anonymous are replaced by a single character X . So, if everything is unknown then the 'local recording identification' field would start with: Startdate X X X X . Additional subfields may follow the ones described here.
5. The 'Digital maximum' of each signal must be larger than its 'Digital minimum'. In case of a negative amplifier gain, the corresponding 'Physical maximum' is smaller than the 'Physical minimum'. Check Section 2.4.1 on how to apply the 'negativity upward' rule of Clinical Neurophysiology. 'Physical maximum' must differ from 'Physical minimum'. In case of uncalibrated signals, physical dimension must be left empty (that is 8 spaces), while 'Physical maximum' and 'Physical minimum' must still contain different values (this is to avoid 'division by 0' errors by some viewers).
6. Never use any digit grouping symbol in numbers. Never use a comma ',' for a decimal separator. When a decimal separator is required, use a dot ('.'). So, do not code numbers as 1.234.567,23 but note this number as 1234567.23 instead.
7. The 'starttime' should be local time at the patients location when the recording was started.
8. The 'number of data records' can only be −1 during recording. As soon as the file is closed, the correct number is known and must be entered.
9. If the ordinary signals were filtered (for instance HighPass, LowPass or Notch) then, preferably automatically, specify them like HP : 0.1 Hz LP:75 Hz N:50 Hz in the 'prefiltering' field of the header. If the file contains an analysis result, the prefiltering field should mention the relevant analysis parameters.
10. The 'transducertype' field should specify the applied sensortype, such as AgAgCl electrode or thermistor .

2.2. Annotations for text, time-keeping, events and stimuli

This section describes how one of the EDF+ signals can be specially coded to store text annotations, time, events and stimuli. In this way, annotations and events are kept in the same file as the ordinary signals that they refer to. The coding is EDF compatible in the sense that old EDF software would simply treat this 'EDF Annotations' signal as an ordinary (be it strange-looking) signal.

2.2.1. The 'EDF Annotations' signal

Besides the ordinary signals as described in Section 2.1.2, EDF+ introduces one other kind of signal which contains annotations of events that can occur at any arbitrary point of time. This signal is identified by giving it (in the EDF+ header) the label EDF Annotations . As in EDF, the 'nr of samples in each data record' field in the header specifies how many 2-byte integers this 'EDF Annotations' signal occupies. But instead of storing 'ordinary signal' samples, the bytes in the data records are filled with characters. The character-bytes are stored byte-by-byte without changing their order. For instance, the text abc is represented by successive byte values 97, 98 and 99 in the 'EDF Annotations' signal. At least one 'EDF Annotations' signal must be present, be it only for keeping the time. Of course, the label EDF Annotations is not allowed for ordinary signals.

An 'EDF Annotations' signal only has meaningful header fields 'label' and 'nr of samples in each data record'. For the sake of EDF compatibility, the fields 'digital minimum' and 'digital maximum' must read − 32768 and 32767 , respectively. The 'Physical maximum' and 'Physical minimum' fields must contain numbers that differ from each other. The other fields of this signal are filled with spaces.

Additional 'EDF Annotations' signals may be defined according to the same specification.

2.2.2. Time-stamped Annotations Lists (TALs) in an 'EDF Annotations' signal

Text, time-keeping, events and stimuli are coded as text annotations in 'EDF Annotations' signals. These annotations are listed in TALs as follows.

Each TAL starts with a timestamp 'Onset'***[21]*** 'Duration'***[20]*** in which ***[21]*** and ***[20]*** are single bytes with value 21 and 20, respectively (unprintable US-ASCII characters) and 'Onset' as well as 'Duration' are coded using US-ASCII characters with byte values 43, 45, 46 and 48–57 (the '+', '−', '.' and '0'–'9' characters, respectively). 'Onset' must start with a '+' or '−' character and specifies the amount of seconds by which the onset of the annotated event follows ('+') or precedes ('−') the startdate/time of the file, that is specified in the header. 'Duration' must not contain any '+' or '−' and specifies the duration of the annotated event in seconds. If such a specification is not relevant, 'Duration' can be skipped in which case its preceding ***[21]*** must also be skipped. Both 'Onset' and 'Duration' can contain a dot ('.') as a decimal separator but only if the fraction of a second is specified, which may be done up to arbitrary accuracy.

After the time stamp, a list of annotations all sharing the same 'Onset' and 'Duration' may follow. Each annotation is followed by a single ***[20]***. A ***[0]*** byte (the unprintable US-ASCII character with byte value 0) follows after the last ***[20]*** of this TAL. So the TAL ends with a ***[20]*** followed by a ***[0]***.

In each data record, the first TAL must start at the first byte of the 'EDF Annotations signal'. Subsequent TALs in the same data record must follow immediately after the trailing ***[0]*** of the preceding TAL. A TAL, including its trailing ***[0]***, may not overflow into another data record. Each event is annotated only once, even if its duration makes it extend into the time period of other data records. Unused bytes of the 'EDF Annotations' signal in the remainder of the data record are also filled with ***[0]*** bytes.

For example, if the technician switches off the lights and closes the door 3 min after startdate/time, this can be stored as the 28-bytes TAL

+180[***20***]Lights off[***20***]Close door[***20***][***0***]

Alternatively, the two events can be stored as two subsequent shorter TALs

+180[***20***]Lights off[***20***][***0***] + 180[***20***]Close door[***20***][***0***]

A 25.5 s apnea that begins 30 min and 0.2 s after the starttime of the file is coded as the TAL

+1800.2[***21***]25.5[***20***]Apnea[***20***][***0***]

2.2.3. Annotations in a TAL

The part between ***[20]*** and the next ***[20]*** is called one annotation. These annotations may only contain characters from the 'Universal Character Set' (UCS, ISO 10646, the 'Universal Character Set', which is identical to the Unicode versions 3+ character set) encoded by UTF-8. This encoding is supported by the major operating systems, compilers and applications. The first 127 UCS characters are identical to those in US-ASCII and are encoded in the corresponding single byte values. US-ASCII characters that are represented by byte values 0–31 are allowed in the annotations only if explicitly prescribed by this EDF+ protocol. In order to enable multi-line texts and tables, US-ASCII characters that are represented by byte values 9 (TAB), 10 (LF) and 13 (CR) are allowed in the annotations. The first 65,534 characters (the Basic Multilingual Plane: BMP) of the UCS contain virtually all characters used in any language in the world including Asian languages. Each BMP character is encoded in up to 3 byte-values. Remember that this encoding applies to the 'EDF Annotations' signal only: in the EDF+ header only US-ASCII characters with byte values 32–126 are allowed.

In order to support automatic averaging and superimposition, identical events or stimuli that occur several times in one file must be coded each time by the same, unique annotation. Annotations of different events/stimuli must differ from this unique annotation.

Annotations that are related to information in only one particular data record, must be in that same data record. Even any annotations describing events preceding the start of that data record, for instance a pre-interval stimulus, must follow the time-keeping annotation.

2.2.4. Time keeping of data records

Because data records need not be contiguous, the starttime of each data record must be specified in another way. So, the first annotation of the first 'EDF Annotations' signal in each data record is empty, but its timestamp specifies how many seconds after the filestartdate/time that data record starts. So, if the first TAL in a data record reads +567[***20***][***20***], then that data record starts 567 s after the startdate/time of the file. If the data records contain 'ordinary signals', then the starttime of each data record must be the starttime of its signals. If there are no 'ordinary signals', then a non-empty annotation immediately following the time-keeping annotation (in the same TAL) must specify what event defines the starttime of this data record. For example, +3456.989[***20***][***20***]R-wave[***20***] indicates that this data record starts at the occurrence of an R-wave, which is almost 3457 s after file start.

The startdate/time of a file is specified in the EDF+ header fields 'startdate of recording' and 'starttime of recording'. These fields must indicate the absolute second in which the start of the first data record falls. So, the first TAL in the first data record always starts with +0.X[***20***][***20***], in which X is the fraction of the second (following the startdate/time) after which the first data record starts. If X = 0, then the .X may be omitted.

2.3. Analysis results in EDF+

Ideally, all data (signals, annotations, events) recorded in one session using one recording system are in one EDF+ file. Data from the same patient but from other sessions or equipment will usually be kept in separate files. Ideally, all these files have an identical 'local patient identification' field. In this way, accurate synchronicity between signals and events is kept within the files and it is exactly known to what period in which patients life the data apply.

In practice, this will not always be possible. However, it *is* easy to maintain synchronicity and patient identification between a recording and data that are derived from that recording. Such derived data can be analysis results such as averages, electrocardiographic QRS parameters, peak latencies or sleep stages or simply a subset of the recording. If such analysis results are stored in EDF+ then this must be done as follows.

If the original recording is in file R.edf (R can be any string), then the derived-file name must be RA.edf in which A can be any string. For instance a PSG would be recorded in file PSG0123_2002.edf and its sleep stage analysis in PSG0123_2002_hyp.edf. *The 'local patient identification' in the header (80 characters) must be copied from the original file into the derived file.*

Make sure that startdate, starttime, and number and duration of data records, are correct. So, if the analysis contains a period from 01.05.00 h till 01.25.00 h of a 24-h recording that was started on 02-AUG-1999, 23.00.00 h, then the analysis file should have startdate 03-AUG-1999 and starttime 01.05.00. In this way both files refer to the correct time period in the patient's life. Because the analysis may reduce or increase the amount of data, the durations of analysis-file data records and recording-file data records may differ.

If analysis results are stored as ordinary signals, then linearly scale them in such a way that a large part of the available range of −32768 till 32767 is used. Such scaling can be applied even after the analysis was done. Specify this linear scaling in the header (digital and physical minimum and maximum) of the analysis file. If linear scaling is really impossible because the useful dynamic range of the analysis result is too large, but only then, apply the standardised logarithmic transformation (Kemp et al., 1998) to store floating point values. EDF+ software must detect thus transformed data and apply the inverse transformation before further processing of the data. Note that old EDF software is not aware of this transformation, and will show the analysis results on a logarithmic scale. So really try linear scaling first!

If a hypnogram is stored as an ordinary signal, sleep stages W, 1, 2, 3, 4, R, M should be coded in the data records as the integer numbers 0, 1, 2, 3, 4, 5, 6, respectively. Unscored epochs should be coded as the integer number 9. If a hypnogram is stored as annotations, use the standard texts described in Section 2.4.3 (example in Section 3).

Automatically document the analysis principle and parameters in the header field 'local recording identification' and, in case of ordinary signals, also in 'label', 'transducer type', 'physical dimension' and 'prefiltering'.

2.4. Standard texts and polarity rules

Standard texts for the header fields 'label' and 'physical dimension' are specified in Sections 2.4.1 and 2.4.2. Section 2.4.1 also specifies polarity rules for EEG, EP and EMG. Standard annotation texts are in Section 2.4.3. Additional standard texts may be developed later and will be kept at http://www.hsr.nl/edf/standard_text.htm.

2.4.1. The 'label' field

The EDF+ header field 'label' (for example `EEG Fpz-Cz`) must start with the 'type of signal' (in this example 'EEG'), followed by a space, followed by a further 'specification of the signal' (in this example 'Fpz-Cz'). Do not specify the type of sensor because this should be done in the 'transducer type' field.

EDF+ prescribes the following standard texts for 'type of signal': 'EEG', 'ECG', 'EOG', 'ERG', 'EMG', 'NC', 'MEG', 'MCG', 'EP', 'Temp', 'Resp', 'SaO2', 'Light', 'Sound', 'Event', 'Freq'.

The 'specification of the signal' for an EEG, EP or EMG signal consists of the names of the two recording electrodes (as defined in the international 10/20 system), separated by a '−' (minus) character. The order of the two names is important because of the following. The recorded voltage (expressed in units of 'physical dimension', usually uV) is computed from the digital values in the EDF+ data records. It equals [(physical minimum) + (digital value in the data record − digital minimum) × (physical maximum − physical minimum)/(digital maximum − digital minimum)]. This voltage must equal the potential at the first electrode (before the '−' character) minus the potential at the second electrode. For example, if the 'Specification' is 'Fpz-Cz' (i.e. the standard label reads `EEG Fpz-Cz`), then the voltage in the file must be the potential at Fpz minus the potential at Cz. In case of a concentric needle electrode recording, a positivity at the centrally insulated wire relative to the cannula of the needle must be stored as a positive voltage in the file. If the electrode locations cannot be accurately specified in short form, like in some EMG recordings, the 'Specification' may be replaced by a less accurate indication such as the name of the muscle.

Note that, in case of an inverting electrode amplifier, a positive voltage is represented by a negative digital value in the file. Any such negative amplifier gain is clear from the digital and physical maximum and minimum in the header, so these header fields enable correct computation of the recorded voltage. In many standard procedures in Clinical Neurophysiology, a relative negativity at the first electrode must be displayed as an upward deflection on the screen. The displaying software must implement any such 'negativity upward' rule by simply upwardly displaying a negative voltage.

In standard EEG investigations, EEG voltages are usually referenced to one, common, electrode, for example A1. The file then contains, in this example, the signals C1-A1, C2-A1, C3-A1, C4-A1, F1-A1, F2-A1, F3-A1, and so on. In some cases, the reference electrode is an average over more than one electrode. In that case, define this average between round brackets. For instance, the EEG between C3 and linked earlobes has label `EEG C3-(A1 + A2`). If the common electrode can not be specified in this format, then use the text 'Ref', for instance in `EEG C3-Ref`. If more of such commons exist, then use texts 'Ref1', 'Ref2', and so on.

If a standard ECG derivation I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, V6, or -aVR, or V2R, V3R, V4R, V7, V8, V9, or X, Y, Z is recorded, then the 'Specification' of the ECG signal must equal the name of that derivation, for instance resulting in label `ECG V2R`.

2.4.2. The 'physical dimension' field

The EDF+ header field 'physical dimension' (for example `uV      `) must start with a 'prefix' (in this example u), followed by the 'basic dimension' (in this example V). The 'basic dimension' for the EXG's is 'V', for temperature it is 'K', 'degC' or 'degF' and for SaO2 it is '%', all without the quotes. The prefix scales the 'physical dimension' according to Table 1. Powers in a 'basic dimension' (for instance the basic dimension used in frequency analysis can be Volts to the power 2 per Hertz) are noted by ^, in this example `V^2/Hz  `. Some basic dimensions involve complicated mathematical expressions. The evaluation order of such expressions is: [everything between the most inner brackets] – prefix – powers – multiplication – division. For example, `uV^2    ` means (0.000001V)^2, not 0.000001(V^2).

2.4.3. Annotations

The annotations `Recording starts` and `Recording ends` help to avoid analysis of irrelevant parts in the beginning (possibly due to a calibration procedure) or the end (after disconnecting the electrodes) of the file. The same is true for the annotations `Lights off` and `Lights on` in a PSG. No durations of these 4 events must be specified in their TAL.

The annotations `Sleep stage W`, `Sleep stage 1`, `Sleep stage 2`, `Sleep stage 3`, `Sleep stage 4`, `Sleep stage R`, `Sleep stage M` and `Sleep stage ?` store hypnograms. The annotations `Obstructive apnea`, `Central apnea`, `Apnea`, `Mixed apnea`, `Hypopnea`, `Limb movement` and `EEG arousal` store other sleep quality assessments. The durations of these events must be specified in their TAL.

3. Examples

For clarity, each TAL in the below examples starts at a separate line. Note however, that the TALs within a data record follow each other immediately: without any space, linefeed or carriage return. The last TAL in some data records is followed by more than one *[0]* in order to complete that data record.

An EDF+ file of an auditory EP recording typically contains the recorded EP signal(s) and an 'EDF Annotations' signal. The example below shows the 'EDF Annotations' signal in the first two data records. In each data record, the first TAL includes the (mandatory) time keeping annotation, while the second TAL specifies a pre-interval stimulus.

```
+0[20][20]Stimulus click 35 dB both ears[20] Free text[20][0]
−0.065[20]Pre-stimulus beep 1000 Hz[20][0]
```

```
+0.3[20][20]Stimulus click 35 dB both ears[20][0]
+0.235[20]Pre-stimulus beep 1000 Hz[20][0][0][0][0][0][0][0][0][0]
```

In this example, averaging can be triggered by the unique texts 'Stimulus click 35 dB both ears' and/or 'Pre-stimulus beep 1000 Hz'.

The next example shows annotations related to a sleep recording. A 8–24 h sleep recording takes about 30–300 MB when stored in EDF or EDF+. The recording can be analysed manually, resulting in apnea's, leg movements and sleep stages. These results are best kept in a separate EDF+ file (about 10–100 kB) which, in this example, only contains one data record with one 'EDF Annotations' signal and no 'ordinary' signals. The table below shows the first half hour and the last few minutes in this data record. This patient fell asleep 9 min after switching off the light and had limb movements (Right and/or Left leg) and apneas after reaching sleep stages 2 and 3, respectively. Another technician might also score apnea's, leg movements or sleep stages from the same sleep recording. These scorings are kept in another separate EDF+ file.

```
+0[20][20]Recording starts[20][0]
+0[21]660[20]Sleep stage W[20][0]
+120[20]Lights off[20][0]
+660[21]300[20]Sleep stage 1[20][0]
+742[20]Turning from right side on back[20][0]
+960[21]180[20]Sleep stage 2[20][0]
+993.2[21]1.2[20]Limb movement[20]R + L leg[20][0]
+1019.4[21]0.8[20]Limb movement[20]R leg[20][0]
+1140[21]300[20]Sleep stage 3[20][0]
+1526.8[21]30.0[20]Obstructive apnea[20][0]
+1603.2[21]24.1[20]Obstructive apnea[20][0]
+1410[21]210[20]Sleep stage 4[20][0]
+1620[21]270[20]Sleep stage 3[20][0]
+1634[20]Turning from back on left side[20][0]
+1890[21]30[20]Sleep stage 2[20][0]
.:.:.:.:.:.:.:.
.:.:.:.:.:.:.:.
+30100[20]Lights on[20][0]
+30210[20]Recording ends[20][0][0][0][0][0][0][0]
```

The EDF+ website at http://www.hsr.nl/edf describes a few examples in which EDF+ stores originally recorded as well as analysed data from various neurophysiological studies such as routine EEG, EMG, F responses, Motor Nerve Conduction Velocities, various EPs, Intra-operative monitoring, Sleep and MSLT. The website also gives an example of a complete EDF+ file.

4. Discussion

At present there is no open and widely accepted standard for the archival and exchange of digital recordings from

Table 1
Dimension prefixes that multiply or divide the basic dimension by factors of 10 till 10^{24}

Prefix		Factor power of 10
Multiplying	Dividing	
Y	y	24
Z	z	21
E	a	18
P	f	15
T	p	12
G	n	9
M	u	6
K	m	3
H	c	2
D	d	1

For example, the dimension uV (microVolt) is 10^{-6} V or 0.000001 V, ms is millisecond and Ms is Megasecond. Note that the prefix 'u' is the US-ASCII letter 'u', not the extended-ASCII letter 'μ'.

ElectroNeurography, ElectroMyoGraphy and Evoked Potential studies (further named ENMGEP). EDF+ accommodates such studies because, contrary to EDF, it can store interrupted recordings as well as annotations and events. Also, a few details have been specified more precisely than in EDF, which avoids ambiguity and makes automatic processing possible in more cases.

Using EDF+, all signals, annotations and events that are recorded in one session using one recording system can be kept safely together in one file. EDF+ can also store events and annotations only, without any signals. This flexibility allows the user to choose an optimal mix. For instance, a sleep centre might store all on-line recorded data (signals and annotations) in one file, its hypnogram and apnea detections in another EDF+ file, a hypnogram made by another technician in a third file. An EEG laboratory might store the on-line obtained raw EEG traces with stimulus events from an EP investigation in one file and the averaged curves with detected latencies in a second file. In Cardiology, the raw ECG with annotations about the patients exercises can be in one file, the detected QRS parameters in another file.

ENMGEP studies are technically more complex than EEG or PSG studies because of two reasons. Firstly, ENMGEP studies apply a larger variety of techniques (such as Conduction velocity, Evoked potentials, EMG) and study parameters such as stimulation intensity and location. Secondly, several rather standardised types of data reduction (like averaging, peak amplitude and latency detection, firing rate computation) are carried out during many ENMGEP investigations. This can all be coded in the EDF+ header and annotations. In this way, EDF+ can store ENMGEP studies, including the applied techniques, technical parameters and most analysis results.

The additional possibilities of EDF+ are also useful for ECG, EEG and PSG recordings. However, EDF+ allows storage of several NON-CONTIGUOUS recordings into one file. This is a definite INCOMPATIBILITY with EDF. This non-contiguousness is the only incompatibility. All other features are EDF compatible. In fact, EDF viewers still work and display EDF+ recordings as if they were continuous. Therefore, we recommend EDF+ files of EEG or PSG studies to be continuous.

We had to accept a few compromises for the sake of EDF compatibility. For instance, EDF+ can only distinguish dates from different centuries by checking more than just the two-digit 'yy' field in the startdate. A special coding scheme was required in order to store floating point analysis results in the two-byte signal samples. And many languages can be expressed in full richness only in the EDF+ annotations, not in the ASCII header. More elegant solutions would have been possible in a completely new format. However, we chose EDF compatibility over elegance because EDF is the only format that is used in practice to collaborate between many countries and labs. The solutions in EDF+, although requiring some programming effort, further increase the possibilities for such collaboration.

Because EDF+ is very close to EDF, and equally simple, EDF+ software can relatively easily be developed based on available EDF software. Because EDF is already widely accepted, EDF+ offers the opportunity to the ENMGEP world to have a standard for archival and exchange of data which can be used immediately.

The EDF+/EDF website at http://www.hsr.nl/edf will continue to support the use of both EDF and EDF+ by supplying further examples, files and free software.

Acknowledgements

Many EDF users suggested to develop something like EDF+. We made a proposal in the summer and the specification was finalised in December 2002. We very much appreciate the constructive discussion with Stig Hanssen, Peter Jacobi, Kevin Menningen, Garðar Þorvarðsson, Thomas Penzel, Marco Roessen, Andreas Robinson and Alpo Värri, mainly on and around Yahoo's EDF users group.

References

Kemp B. European data format (EDF): current availability and additional applications. J Sleep Res 2002;11(Suppl. 1):120.

Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings. Electroenceph clin Neurophysiol 1992;82:391–3.

Kemp B, Penzel T, Värri AO, Sykacek P, Roberts SJ, Nielsen KD. EDF: a simple format for graphical analysis results from polygraphic SIESTA recordings. J Sleep Res 1998;7(Suppl. 2):132.