

Západočeská univerzita v Plzni  
Fakulta Aplikovaných Věd  
Katedra informatiky a výpočetní techniky

## **Diplomová práce**

# **Segmentace EEG signálů a archivace vybraných segmentů**

Plzeň, 2012

Bc. Miroslav Vacek

*Prohlašuji, že jsem diplomovou práci zpracoval samostatně a výhradně s použitím citovaných pramenů.*

V Plzni 20. června 2012

Miroslav Vacek .....

# Poděkování

Na tomto místě bych rád poděkoval především vedoucí mé závěrečné práce za udělené konzultace a odborné rady při jejím zpracování. Dále bych rád vyjádřil své díky Neurologickému Oddělení Fakultní nemocnice Lochotín za poskytnutá data.

Plzeň, květen 2012

Miroslav Vacek

# Abstract

## Segmentation of EEG signal and archivation of selected samples

This master's thesis deals with possibilities of creating and storing samples of EEG signal gathered by neurological clinic of Pilsen's hospital. The thorough analysis of problem was made and based upon these results the solution was designed.

# Obsah

---

<b>2</b>	<b>ÚVOD .....</b>	<b>1</b>
<b>3</b>	<b>ZÁKLADNÍ POJMY .....</b>	<b>2</b>
3.1	NEUROANATOMIE V KOSTCE .....	2
3.2	EEG – ELEKTROENCEFALOGRAFIE .....	3
3.2.1	<i>Nástroje EEG</i> .....	4
3.2.2	<i>EEG signál</i> .....	6
3.2.3	<i>Artefakty</i> .....	7
3.3	EEG VYŠETŘENÍ .....	7
<b>4</b>	<b>SOUČASNÉ SITUACE.....</b>	<b>8</b>
4.1	VYBAVENÍ PRACOVNÍHO MÍSTA.....	8
4.2	ULOŽENÍ EEG DAT DO IS NEMOCNICE.....	8
4.3	VYMEZENÍ PROBLÉMU .....	9
<b>5</b>	<b>ANALÝZA PROBLÉMU .....</b>	<b>10</b>
5.1	ROZBOR MOŽNÝCH ŘEŠENÍ.....	10
5.2	ZPŮSOB ULOŽENÍ DAT Z EEG VYŠETŘENÍ .....	10
5.3	ALTERNATIVNÍ ZPŮSOB ULOŽENÍ DAT .....	11
5.3.1	<i>European Data Format</i> .....	11
5.4	JAKÝ ZVOLIT PROHLÍŽEČ? .....	12
5.4.1	<i>EEGLAB</i> .....	12
5.4.2	<i>jEDF</i> .....	14
5.4.3	<i>EDFbrowser</i> .....	15
5.5	TESTOVACÍ DATA .....	16
5.6	TEST ZVOLENÝCH EDF PROHLÍŽEČŮ .....	17
5.7	POŽADAVKY NA BUDOUCÍ PROGRAM .....	18
5.8	PREZENTACE NAVRŽENÉHO ŘEŠENÍ .....	18
<b>6</b>	<b>NÁVRH ALTERNATIVNÍHO ŘEŠENÍ .....</b>	<b>20</b>
6.1	PACS A JEHO VÝZNAM.....	20
<b>7</b>	<b>PŘEVOD FORMÁTU EDF NA DICOM .....</b>	<b>21</b>
7.1	SPECIFIKACE FORMÁTU EDF(+) <i>.....</i>	21
7.1.1	<i>Dodatečná specifikace EDF+</i> .....	23
7.1.2	<i>Příklad souboru uloženého v EDF formátu</i> .....	26
7.2	DICOM.....	26
7.2.1	<i>Co je to vlastně DICOM?</i> .....	26
7.2.2	<i>Datový formát DICOM</i> .....	28
7.2.3	<i>DICOM objekty</i> .....	33
7.3	POROVNÁNÍ EDF(+) A DICOM .....	34
7.4	POPIS DOSTUPNÝCH KNIHOVEN .....	35
7.4.1	<i>EDFlib</i> .....	35
7.4.2	<i>DCMTK</i> .....	36

<b>8</b>	<b>POPIS ALGORITMU KONVERZE .....</b>	<b>38</b>
8.1.1	<i>Načtení EEG souboru .....</i>	38
8.1.2	<i>Vytvoření obrázku.....</i>	39
8.1.3	<i>Vytvoření DICOM souboru.....</i>	40
<b>9</b>	<b>ZÁVĚR.....</b>	<b>41</b>

# 1 Úvod

---

Tato práce vznikla na základě požadavků Neurologické kliniky Fakultní nemocnice Lochotín v Plzni na segmentaci a následné zpracování dat z elektroencefalografu.

Základní standardní neinvazivní metodou funkčního vyšetření elektrické aktivity centrálního nervového systému je vyšetření pomocí elektroencefalografu. Elektroencefalogramy jednotlivých pacientů jsou důležitým zdrojem informace pro lékaře, kteří na jejich základě mohou odborně posoudit zdravotní stav pacienta. Je však důležité si uvědomit, že běžné vyšetření EEG trvá přibližně 45 minut. V případě komplikovaných onemocnění je nutno zvolit záznam aktivit pacientova mozku během spánku. Takováto EEG vyšetření samozřejmě trvají mnohem déle (řádově několik hodin – podle potřeby). Během vyšetření pomocí elektroencefalografu jsou zaznamenávány elektrické potenciály měřené elektrodami umístěnými na povrchu hlavy pacienta. Získaný záznam je ukládán pro další využití.

Z výše uvedeného je patrné, že objem dat s výsledky vyšetření na elektroencefalografu je obrovský a to i když uvažujeme pouze běžná (cca 45 minutová) vyšetření. Uložit je jako celek do nemocničního informačního systému proto nepřipadá v úvahu jak z důvodu objemu dat, tak času potřebného k uložení (několik desítek minut). Proto je nutné, aby z výsledného záznamu byly uchovány pouze úseky elektroencefalogramu, které lékař označil jako důležité. Tyto lékařovy poznámky jsou uloženy ve formě textového souboru, který má v porovnání s celým elektroencefalogramem zanedbatelnou velikost. Právě tento soubor je uložen do IS nemocnice.

Dosavadní praxe na neurologické klinice umožňuje zapsat do informačního systému pouze text s popisem průběhu vyšetření. Je-li zapotřebí podrobnějšího náhledu či porovnání s jiným záznamem, je nutno manuálně vyhledat konkrétní soubor elektroencefalogramu, jenž je uložen na magnetické páse. Tento pásek se dále vloží do speciálního počítače. Ten je schopen data z něj přečíst a zobrazit. Vzhledem k tomu, že tento počítač není připojen k síti, znamená uvedený postup značnou časovou prodlevu.

Cílem této práce je prozkoumat způsob uložení naměřených dat a navrhnout takové řešení, které by umožnilo do informačního systému nemocnice uložit konkrétní výřez (segment) z elektroencefalogramu místo souboru s textovým popisem.

## 2 Základní pojmy

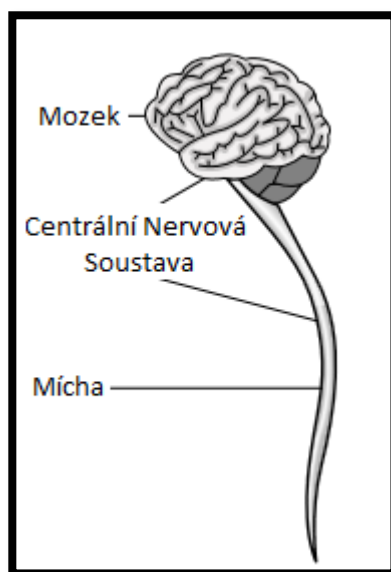
---

Pro pochopení problému předkládané práce je nezbytné vysvětlit několik základních pojmů z oblasti neurologie.

### 2.1 Neuroanatomie v kostce

Klíčovým je pojem neurologie. Z. Ambler ve své práci *Základy neurologie* (2006, s. 13) říká o neurologii toto: „Obor neurologie se zabývá poruchami centrálního i periferního nervového systému včetně svalového aparátu. Neurologické choroby mohou postihovat různé etáže a různé systémy centrálního i periferního nervstva, a mohou tak vznikat velmi rozdílné, často zcela specifické příznaky, např. poruchy řeší, hybností, citlivosti, zraku atd.“

O několik odstavců dále Ambler vysvětluje, že lidský nervový systém lze rozdělit na periferní (PNS) a centrální (CNS). CNS (viz obrázek 2.1) tvoří mozek a mícha, PNS se pak skládá z mozkových a míšních nervů a to včetně jejich jader a předních a zadních rohů. Pro tuto práci je nejpodstatnější CNS, a proto si ji blíže představíme.



Obrázek 2.1 - Schéma centrální nervové soustavy<sup>1</sup>

„Mozek se dělí na dvě hemisféry, tvořené jednotlivými mozkovými laloky a hlouběji uloženými bazálními ganglii, diencefalon, mozkový kmen a mozeček. Mozkový kmen tvoří mezencefalon, pons a medulla oblongata.“ (Ambler, 2006, s. 13).

„Mícha je tvořena jednak šedou hmotou (přední a zadní míšní rohy – obsahující buněčná těla, dendrity a, axony i gliální buňky), jednak bílou hmotou (provazce – tvořené myelinizovanými i nemyelinizovanými axony a gliálními buňkami).“ (Ambler, 2006, s. 13).

---

<sup>1</sup> Upraveno podle: [https://en.wikipedia.org/wiki/File:Central\\_nervous\\_system.svg](https://en.wikipedia.org/wiki/File:Central_nervous_system.svg)



Z. Abmler považuje za základní stavební kámen celé nervové soustavy neuron (či nervovou buňku). Neuron se skládá z buněčného těla (soma), ze kterého vychází jeden výběžek vedoucí vzruchy směrem od buňky – axon (neurit) a řada menších výběžků vedoucích vzruchy k buňce – dendrity.

Neuron má tři základní funkce (Ambler, 2006, s. 14):

- a) *trofickou*, která je vázána na vnitřek buňky a je nezbytná pro strukturní neporušenost a funkční výkonnost neuronu,
- b) *specifickou*, což je schopnost tvořit a přenášet vzruchy a je funkcí buněčné membrány,
- c) *sekreční*, která spočívá v uvolňování chemických látek.

Neurony lze rozdělit na aferentní, eferentní a interneurony (asociační neurony, které umožňují hlavně mezineuronovou komunikaci).

Neurony reagují na podněty tzv. nervovým vzruchem. Nervový vzruch je fyzikálně-chemická změna přenášená nervovým vláknem, resp. jeho membránou. V klidu je vnitřek neuronu elektricky negativní, má záporný náboj, povrch membrány je naopak pozitivní, membrána je polarizována. Při podráždění se polarita obrátí, povrch se stane negativním, dochází k depolarizaci a vlně elektrické negativity, která se šíří po celém povrchu membrány. Od vzruchu je třeba odlišovat podnět. Jde o energii, která vzruch vyvolává. Působení podnětu se nazývá stimulace (dráždění). Vzruch je fyziologický děj, v podstatě stejný ve všech nervových vláknech. Kvalita a intenzita podnětu rozhoduje o tom, jaký bude časový sled vzruchů a jejich prostorové upořádání v jednotlivých vláknech. Nervový systém je vlastně kybernetický systém, který přenáší a zpracovává řadu nejrůznějších informací. Každá informace musí být nejprve přeměněna na nervový vzruch. Určitý sled a uspořádání jednotlivých vzruchů tvoří kód, který pak představuje vlastní signál, jenž informuje o změně činnosti jednotlivých orgánů a systému těla. (Abmler, 2010, s. 14).

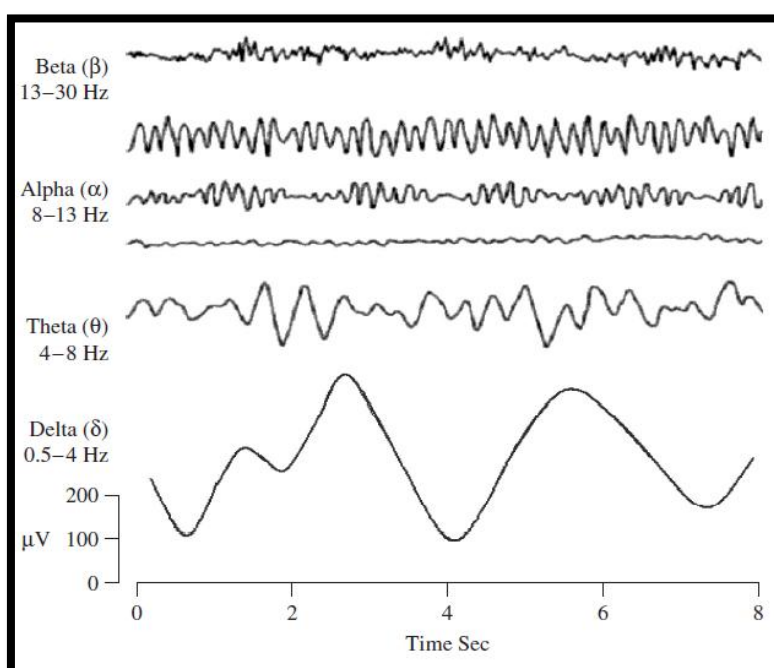
## 2.2 EEG – Elektroencefalografie

Z. Ambler ve své publikaci *Základy neurologie* (2006, s. 127) říká o EEG následující: „*Elektroencefalografie* (EEG) registruje elektrické mozkové potenciály z povrchu hlavy. Hodnotí funkční aktivitu mozkových struktur, detekuje ložiskovou poruchu (nemůže však zjistit její povahu) i difúzní poruchu u zánětlivých nebo metabolických afekcí a má velký význam při diagnostice epilepsie.“ Elektroencefalografie je tedy jednou z neinvazivních metod vyšetření mozkové aktivity člověka s cílem diagnostikovat specifické choroby (např. epilepsii).

Mozková aktivita je u různých věkových skupin různá. Závisí také na aktivitě člověka a na fázích spánku. Přehled frekvencí odpovídajících konkrétní věkové skupině a situaci zachycuje tabulka 2.1. To, jak jednotlivá pásma vypadají na elektroencefalogramu, ukazuje následující obrázek 2.2.

Pásmo	Frekvence [Hz]	Popis
Alfa	8-13	Vyskytuje se u dospělého člověka v klidovém stavu se zavřenýma očima.
Beta	13-40	Vyskytuje se u dospělého člověka při vyvíjení konkrétní činnosti.
Théta	4-7	Vyskytuje se u dětí. Dále je možné ji zaznamenat při spánku dospělého člověka.
Delta	>4	Vyskytuje se u kojenců. Dále je možné ji zaznamenat při hlubokém spánku dospělého

Tabulka 2.1 – Přehled frekvencí mozkových vln



Obrázek 2.2 – Grafické znázornění mozkový vln<sup>2</sup>

### 2.2.1 Nástroje EEG

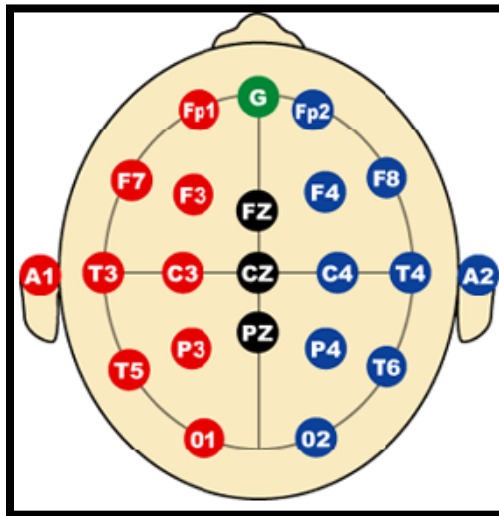
Produktem EEG je tzv. *elektroencefalogram*, což není nic jiného, než záznam časové změny elektrického potenciálu způsobeného mozkovou aktivitou. Takovéto záznamy jsou pořizovány elektroencefalografem a to z povrchu lebky<sup>3</sup>.

Aby bylo možno snímat data z povrchu lebky, je nutné pacientovy nasadit „EEG čepičku“ ve které jsou rozestaveny měřící elektrody. Tyto elektrody mohou být zapojeny v různých konfiguracích. Nejčastěji se používá zapojení elektrod podle tzv. 10/20 systému (čti deset - dvacet). Tento název vznikl podle způsobu, jakým jsou umístěovány elektrody do „EEG čepičky“ a v současné době se jedná o jedno ze základních rozestavení. Je vhodné zmínit, že „čepička“ může obsahovat různý počet elektrod a stále být zapojena podle 10/20

<sup>2</sup> Zdroj obrázku: Sanei, Chambers, EEG signal processing, s. 12

<sup>3</sup> Bc. Robin Horniak, 2010, s. 3

systemu. Přesné rozložení elektrod pro 10/20 systém s 21 elektrodami je uvedeno na obrázku 2.3. Před samotným začátkem měření je vhodné nanést na každou elektrodu gel, kterým se zvýší její vodivost, čímž se dosáhne přesnějších výsledků při měření.



Obrázek 2.3 - Rozložení elektrod v „EEG čepičce“ podle systému 10/20 s 21 elektrodami<sup>4</sup>

Přístroj zaznamenávající mozkovou aktivitu se nazývá *elektroencefalograf*. Snímaná mozková aktivita má napětí řádově desítek mikrovoltů, což není pro přesné zpracování dostatečné. Proto je potřeba daný signál zesílit. Tento postup má však několik nedostatků. Mezi největší z nich patří rušení signálu (špatné odstínění, špatný kontakt elektrod s pokožkou, elektromagnetické záření apod.). Přístroj se skládá ze dvou hlavních částí. První část tvoří „EEG čepička“ s rozložením elektrod a druhou část hlavní výkonová jednotka (viz obrázek 2.4).<sup>5</sup>



Obrázek 2.4 - Elektroencefalograf a jeho komponenty (EEG čepička a výkonová jednotka)<sup>6</sup>

<sup>4</sup> Zdroj obrázku: Bc. Robin Horniak, Analýza EEG signálu, s. 12

<sup>5</sup> Bc. Robin Horniak, 2010, s. 3

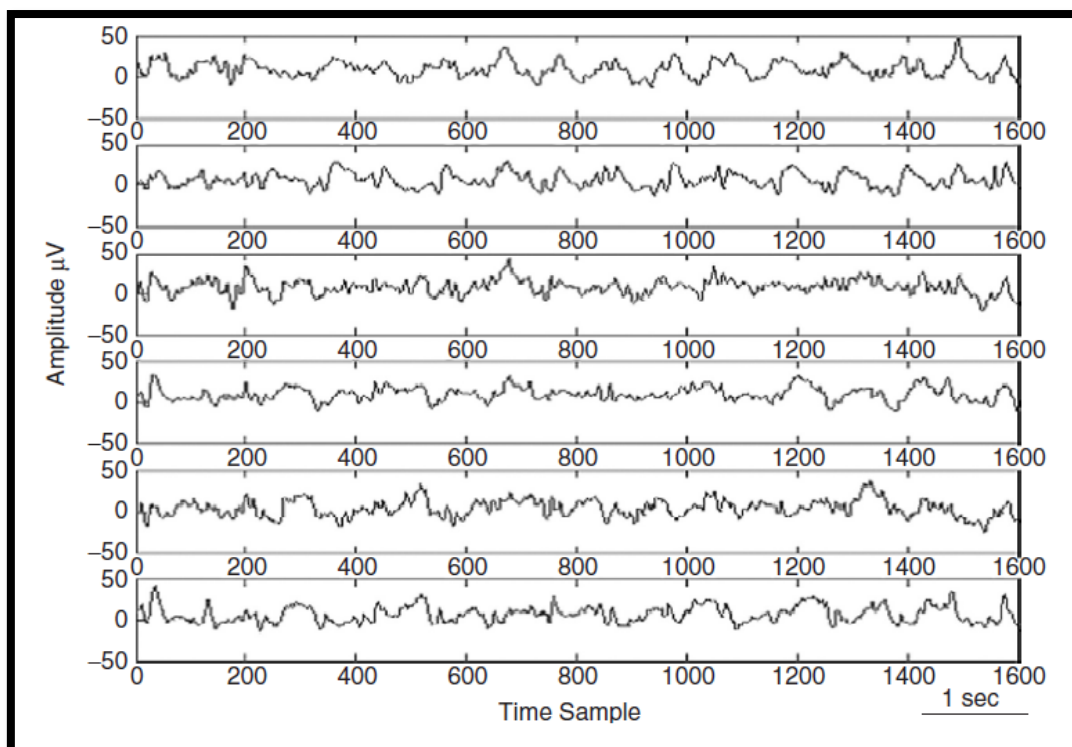
<sup>6</sup> Zdroj obrázku: Bc. Robin Horniak, Analýza EEG signálu, s. 12

## 2.2.2 EEG signál

Každá z nervových buněk v lidském mozku vytváří elektrické potenciály. Těmi jsou akční potenciál a dále postsynaptické potenciály. Akční potenciál má velmi krátkou dobu trvání (cca 1 ms), a proto se v EEG signálu neprojevuje. Postsynaptické potenciály, však trvají déle (cca 100 ms) a v EEG signálu se již projeví. A jsou to právě změny těchto potenciálů, které vytvářejí elektrické pole. Sloučením takto vzniklých elektrických polí snímaných z povrchu hlavy vzniká EEG signál.<sup>7</sup>

EEG je signálem elektrickým, a proto se v souvislosti s jeho popisem se nejčastěji diskutuje jeho amplituda a frekvence. Amplituda EEG signálu se pohybuje nejčastěji v desítkách nebo maximálně stovkách mikrovoltů ( $\mu\text{V}$ ). Mnohem větší výpovědní hodnotu má v tomto případě frekvence. Ta se mění například při inhibici očního nervu (otevření oka), či v závislosti na fázi spánku nebo naopak se zvyšující se aktivitou člověka.<sup>8</sup>

Na obrázku 2.5 je zobrazen průběh EEG signálu normální mozkové aktivity zdravého dospělého člověka v klidovém stavu s dobou trvání přibližně 7 sekund. Za povšimnutí stojí, že jsou zobrazena data pouze ze šesti kanálů (viz dále) a není měřen žádný další biologický signál.



Obrázek 2.5 - Výřez z EEG záznamu činnosti mozku zdravého pacienta pro 6 kanálů v délce cca 7s<sup>9</sup>

V souvislosti s EEG signálem je důležité uvést jeden klíčový pojem - kanál. Signál naměřený z elektrod je pro vyhodnocení lékařem nepoužitelný. Signál bývá silně ovlivněn nežádoucími artefakty, s nízkou amplitudou a hlavně je na počítači jen těžko zobrazitelný,

<sup>7</sup> Jiří Hrdlička, 2011, s. 2

<sup>8</sup> Jiří Hrdlička, 2011, s. 2

<sup>9</sup> Zdroj obrázku: Sanei, Chambers, EEG signal processing, s. 18

poněvadž se jedná o signál analogový. Proto se signál z elektrod přivede na vstup diferenciálního zesilovače, výstup z tohoto zesilovače se nazývá *kanál*. Základní jednotkou registrace bioelektrické aktivity je tzv. svod, který je tvořen dvěma elektrodami připojenými na dva vstupy diferenciálního zesilovače jednoho EEG kanálu. Podle způsobu uspořádání použitých svodů rozlišujeme následující zapojení elektrod do zesilovače:

- referenční,
- bipolární,
- zdrojové.

Výsledkem konkrétního způsobu zapojení je signál, jenž lze zobrazit pomocí počítače. Je to právě takto vzniklý signál, který lékaři zkoumají.<sup>10</sup>

### 2.2.3 Artefakty

---

Při měření EEG signálu se velmi často vyskytují tzv. artefakty. Tyto artefakty lze rozdělit do dvou základních skupin:

- 1) biologické,
- 2) technické.

Mezi biologické artefakty lze zařadit především projevy lidské aktivity během EEG vyšetření (např. mrknutí, svalová aktivita, pocení, aj.).

Do skupiny technické artefakty se řadí artefakty způsobené především měřicími prostředky (např. špatná elektroda, špatné odstínění aj.).

Artefakty jako takové značně ztěžují hodnocení EEG záznamů lékařem. Proto je nutné tyto oblasti v záznamu nalézt a odstranit, aby došlo k usnadnění klasifikace dat.<sup>11</sup>

## 2.3 EEG vyšetření

EEG vyšetření je většinou prováděno na základě doporučení lékaře pokud má u pacienta podezření na nějakou nemoc centrálního nervového systému. Toto vyšetření lékař indikuje například při mozkové příhodě, meningitidách, úrazech a otřesech mozku, ale i psychologických onemocněních a jiných krátkodobých a dlouhodobých poruchách vědomí. Zvláštní význam má EEG při diagnostice epilepsie. Výsledky tohoto vyšetření lze také využít pro terapeutickou metodu biofeedback.<sup>12</sup>

Při vyšetření je pacientovy na hlavu nasazena „EEG čepička“ s elektrodami, sloužící pro měření napětí z povrchu hlavy. Vyšetření může trvat několik minut, až několik hodin, čemuž také odpovídá délka záznamu. Kromě EEG se většinou měří i některé další biologické signály, např. činnost srdce (EKG), respirační signál (PNG), pohyby očí (EOG), činnost svalů (EMG) a další. Výsledkem tohoto vyšetření je soubor s EEG signály, který také obsahuje i data o doprovodných biologických signálech např. EKG (jsou-li měřeny).

---

<sup>10</sup> Jiří Hrdlička, 2011, s. 4

<sup>11</sup> Bc. Robin Horniak, 2010, str. 5

<sup>12</sup> Radek Polák, 2009, s. 2

## 3 Současné situace

---

### 3.1 Vybavení pracoviště

V současné době používá klinika neurologie elektroencefalograf od firmy Walter-Graphtek, série PL. Tento přístroj je připojen k počítači s operačním systémem MS-DOS, který je schopen uložit výstup z elektroencefalografu na magnetickou pásku. Vzhledem k objemu naměřených dat je možné na jednu magnetickou pásku uložit pouze několik (jednotky) záznamů. Tuto pásku je třeba vyměnit vždy, pokud si lékař přeje zobrazit data pacienta, jež nejsou uložena na aktuálně vložené pásce, nebo na pásce nezbyvá volné místo.

Kromě počítače připojeného k elektroencefalografu je pracoviště dále vybaveno pracovní stanicí s operačním systémem Windows 2000. Na této stanici je nainstalován program PL-Winsor, též od společnosti Walter-Graphtek. Stejná stanice je síťově spojena s počítačem, který zajišťuje obsluhu elektroencefalografu.

Program PL-Winsor umožňuje lékařům prohlédnout si přes síť data z měření, která jsou fyzicky uložena na vedlejší stanici na magnetickém záznamovém médiu. Tento program poskytuje grafické uživatelské rozhraní umožňující lékařům procházet data po časové ose, vybírat jednotlivé kanály či využívat funkce zoom (změna amplitudy, frekvence a doby trvání epochy signálu). Verze programu PL-Winsor je závislá na používaném elektroencefalografu. Proto tento software nepodporuje některé funkce běžné mezi moderními neurologickými programy, zejména propojení s IS nemocnice.

Nová verze programu, jež zahrnuje pokročilou funkčnost, je vázána na nový přístroj. To v konečném důsledku znamená, že pokud by chtěla klinika neurologie novou verzi programu PL-Winsor, která by zahrnovala velké množství nových funkcí (propojení s IS, segmentace), musela by pořídit nový elektroencefalograf a také většinu jeho příslušenství.

Informační systém je důležitou součástí mnoha organizací. Fakultní nemocnice v Plzni v současné době používá IS od firmy MediCalc. Tento systém, mimo jiné, umožňuje lékařům přistoupit k výsledkům všech vyšetření, která pacient absolvoval napříč všemi pracovišti FN Plzeň, přímo z jejich počítače. Na základě těchto informací pak lékař může rychle stanovit přesnou diagnózu, či navrhnout konkrétní léčbu. Mnoho z pracovišť je na systém MediCalc napojeno, což umožňuje online dostupnost vyšetření pacienta.

### 3.2 Uložení EEG dat do IS nemocnice

Problémem je forma, ve které jsou do IS MediCalc zanášeny výsledky EEG vyšetření. V současné době je do IS zanesen pouze textový popis vyšetření. Každé vyšetření je „ručně“ ohodnoceno vyšetřujícím lékařem a tento textový popis je ukládán do IS. Tento postup je vyhovující pouze, pokud si lékař přeje jednorázově přistoupit k datům pacienta. Porovnávání několika vyšetření či přesnější pohled na data není možné realizovat jen na základě textového popisu. Potřebuje-li lékař srovnat např. dva EEG záznamy, je nutné je vyhledat a zobrazit

přímo v laboratoři s programem PL-Winsor. Proto byl vznesen požadavek na uložení konkrétního výřezu (segmentu) EEG dat do IS.

### 3.3 Vymezení problému

Cílem této práce je analyzovat způsob ukládání dat z měření mozkové aktivity pacienta a navrhnout takové řešení, jež by umožnilo uložit do IS jeden, či několik segmentů konkrétního EEG vyšetření místo textové popisu. Lékaři by mělo být umožněno vyznačit v datech zajímavou oblast a tu uložit do IS. Takto vzniklý segment by poté byl přístupným všem dalším pracovištím napojených na IS MediCalc.

## 4 Analýza problému

---

### 4.1 Rozbor možných řešení

Pro vyřešení problému specifikovaného v kapitole 3.3 jsme stanovili následující možnosti řešení:

1. Naprogramovat rozšíření (zásuvný modul – anglicky: plug-in) pro program PL-Winsor, umožňující vybrat a uložit konkrétní segment.
2. Použít některý ze stávajících nekomerčních programů pro práci s výsledky EEG vyšetření jako základ pro stand-alone aplikaci umožňující výběr a uložení segmentu.

Pokud bychom se vydali cestou nastíněnou v první možnosti, dosáhli bychom hned několika pozitiv. Největším z nich je téměř nulová úprava stávajícího uživatelského rozhraní. Lékaři by mohli pracovat s programem jako doposud, jen s tím rozdílem, že by bylo možné úsek vyšetření vybrat a uložit do IS. Problémem je celková koncepce programu PL-Winsor. Pro realizaci tohoto řešení by bylo zapotřebí přistoupit ke zdrojovému kódu tohoto programu a zakomponovat do něj námi vytvořené rozšíření, což se po konzultaci s firmou Walter-Graphtek ukázalo jako problém.

Další možností bylo vytvořit specializovanou aplikaci, která lékařům umožní zobrazit data z EEG měření a dále jim nabídne možnost výběru zajímavých segmentů, které bude možno uložit do IS nemocnice. Základem této aplikace by mohl být některý z široké řady otevřených prohlížečů EEG dat. Pokud bychom zvolili tento postup, bylo by nutné přenést data z vyšetření do takového formátu, aby je bylo možné v některém z dostupných prohlížečů zobrazit. Tento postup je z hlediska realizace mnohem přijatelnější než předchozí řešení.

### 4.2 Způsob uložení dat z EEG vyšetření

Elektroencefalograf firmy Walter-Graphtek využívá pro uložení dat z jednoho měření dvou souborů. Tyto soubory mají následující přípony a význam:

1. \*.WG1
  - Soubor s touto příponou obsahuje data pro jednotlivé kanály.
  - Tento soubor je uložen v binárním formátu.
2. \*.WG2
  - Soubor s touto příponou obsahuje identifikaci pacienta.
  - Tento soubor je uložen v binárním formátu.

Důležitým faktem je, že k datům zaznamenávající mozkovou činnost pacienta jsou „přibalena“ data měřící srdeční rytmus (EKG). Soubory typu WG1 tedy obsahují jak data pro jednotlivé kanály z „EEG čepičky“, tak data sledující srdeční aktivitu.



Problémem je, že soubory s příponou WG1 i WG2 jsou binární. Jedná se tedy o proprietární formáty firmy Walter-Graphtek. To znamená, že bez znalosti toho, co který bit znamená, nejsme schopni tyto soubory načíst ani s nimi jakkoliv jinak pracovat v žádném dalším programu mimo PL-Winsor.

Proto jsme společně s doktorkou Klečkovou vznesli dotaz na české zastoupení firmy Walter-Graphtek s žádostí o poskytnutí technické dokumentace k datovým formátům WG1 a WG2. Odpověď vyzněla v tom smyslu, že formáty WG1 a WG2 se neustále vyvíjejí a najít technickou specifikaci té verze formátů, která je generována elektroencefalografem kliniky neurologie je problematické. Tímto jsme dospěli k závěru, že data nebude možné převzít přímo z výsledků vyšetření uložených ve formátech WG1 a WG2.

## 4.3 Alternativní způsob uložení dat

Ideálním způsobem pro uložení dat by byl nějaký otevřený a standardizovaný formát. V tomto ohledu bylo nutné prozkoumat možnosti exportu dat v programu PL-Winsor. Po konzultaci s personálem kliniky neurologie jsme zjistili, že lze data převést hned do několika formátů. Nejzajímavější se jevila možnost převést EEG data do formátu EDF.

### 4.3.1 European Data Format

---

Proč se nejperspektivněji jevil právě formát EDF? Především proto, že splňuje všechny naše požadavky. EDF je zkratkou od anglického sousloví European Data Format (Evropský datový formát). EDF je jednoduchý, flexibilní a otevřený formát pro výměnu a ukládání biologických signálů. Tento formát byl vyvinut několika Evropskými zdravotnickými inženýry, kteří se poprvé sešli v roce 1987 na Mezinárodním spánkovém kongresu v Kopenhagenu. S podporou profesorky Annelise Rosenflackové tito inženýři začali Evropský projekt analýzy mozkové aktivity během bdělého stavu a spánku. Tento projekt probíhal v letech 1989 až 1992. Cílem projektu bylo aplikovat jimi navržené algoritmy spánkové analýzy na data druhé skupiny a následně porovnat výsledky. A tak se jednoho březnového rána roku 1990 shodli na velmi jednoduchém společném formátu dat, ve kterém budou ukládány výsledky jejich projektu. Tento formát se později stal známý pod zkratkou EDF.

Specifikace formátu EDF byla zveřejněna v roce 1992 v časopise „Electroencephalography and Clinical Neurophysiology 82“ na stránkách 391 až 393. Od té doby se EDF stalo de facto standardem pro záznamy dat z měření EEG a PSG. Tento standard našel také podporu ze stran výrobců měřících zařízení.

V roce 2002 bylo vytvořeno rozšíření formátu EDF pojmenované EDF+ (čti EDF plus). Toto rozšíření je z valné většiny kompatibilní s předchozí specifikací formátu EDF. To znamená, že všechny doposud existující prohlížeče EDF signálů jsou schopné zobrazit i signály uložené podle standardu EDF+. Standard EDF+ však navíc nabízí možnost přerušovaného záznamu a také podporuje tvorbu anotací (textového popisu záznamu). Proto je možné, aby ve formátu EDF+ byla ukládána jakákoliv medicínská data jako např. EMG, evokované potenciály, ECG, apod. Specifikace EDF+ je také více striktní, což umožňuje

použití funkce automatické kalibrace elektrod a lokalizace. V návrhu EDF+ je také opraveno a doplněno několik „opomenutí“ oproti staršímu formátu. Za všechny jmenujme vyřešení problémů s ukládáním celých čísel ve formátu little endian, dále přibila možnost rozlišovat mezi desetinnou tečkou a desetinnou čárkou, a několik dalších.

Standard EDF+ byl poprvé publikován v roce 2003 v časopise „Clinical Neurophysiology 114“ na stranách 1755 až 1761. Od té doby je tento formát ukládání dat používán především v oblasti klinické neurofyzologie, dále pak při analýze mozkové aktivity během spánku a v kardiologii.

Za zmínku také stojí, že jak standard EDF, tak standard EDF+ jsou k dispozici zdarma a to bez jakékoliv formy poplatku za využití. Dále je možné zdarma nahlédnout do specifikací obou formátů a to buď prostřednictvím výše zmíněných publikací, nebo prostřednictvím internetových stránek. Více informací o vzniku a vývoji formátu EDF lze nalézt v [6].

## 4.4 Jaký zvolit prohlížeč?

Klíčovým krokem byla možnost převést data do otevřeného a řádně specifikovaného formátu EDF(+). Dalším krokem bylo vybrat nekomerční program, jenž by uměl data v takovémto formátu načíst a umožnil jejich procházení. Pokud by se nám podařilo vybrat takový EDF prohlížeč, který by nabízel námi požadované funkce, bylo by možné jej využít jako základ pro stand-alone aplikaci zajišťující výběr a uložení segmentu.

Požadavky na prohlížeč jsou následující:

1. Prohlížeč bude umět načíst data v otevřeném formátu EDF(+).
2. Prohlížeč bude k dispozici bez nutnosti zakoupení licence.

Po zralé úvaze jsme okruh pro výběr zúžili na tři kandidáty. Prvních z nich bylo rozšíření programu MATLAB s názvem *EEGLAB*. Druhým kandidátem se stal prohlížeč naprogramovaný v jazyce JAVA s názvem *jEDF* a třetím prohlížečem byl *EDFbrowser* implementovaný v jazyce C++.

### 4.4.1 EEGLAB

---

*EEGLAB* (viz obrázek 4.1) je interaktivní MATLAB toolbox pro zpracování nepřetržitých a událostmi vyvolaných aktivit mozku, zejména EEG, MEG a dalších elektrofyziologických dat. *EEGLAB* zahrnuje i nástroje pro analýzu artefaktů, frekvenční analýzu signálů a samozřejmě tako pro vizualizaci zaznamenaných dat.

Součástí prostředí *EEGLAB* je i interaktivní grafické uživatelské rozhraní, které umožní uživatelům flexibilně a interaktivně pracovat s jejich EEG daty (a to i s vysokou hustotou signálu). Program je také vhodný pro práci s dalšími daty, která lze získat měřením mozkové aktivity. Další podporovanou funkcí tohoto programu je možnost tvorby automatických evaluačních skriptů, které lze dávkově spustit z příkazové řádky programu MATLAB.



#### 4.4.2 jEDF

---

Program *jEDF* (viz obrázek 4.2) je dílem francouzského autora Nizara Kerkeniho. Jedná se o otevřený EDF prohlížeč naprogramovaný v programovacím jazyce JAVA. Mezi jeho hlavní funkce patří podpora vizualizace a manipulace se soubory ve formátu EDF(+), podporuje však i tvorbu hypnogramů a manipulaci s nimi.

Díky tomu, že je program implementován v programovacím jazyce JAVA, je možné jej provozovat na operačních systémech Windows i Linux. *jEDF* není vázán na žádný software třetích stran (připomeňme, že nástroj *EEGLAB* vyžaduje pro svou funkčnost platformu MATLAB, která bude nainstalovaná na téže stroji, kde má být provozován *EEGLAB*), mimo virtuálního stroje JAVA.

V porovnání s programem *EEGLAB* je však *jEDF* poněkud více specifický. *EEGLAB* se zaměřuje na komplexní analýzy prováděné nad různými daty získanými měřením mozkové aktivity (uložené v široké paletě podporovaných formátů souborů). Oproti tomu *jEDF* se zaměřuje výhradně na data uložená ve formátu EDF(+). A tomu také odpovídají poskytované nástroje.

Paleta nástrojů programu *EEGLAB* je také mnohem širší, a to jak v oblasti analytických algoritmů (podporované jsou algoritmy ICA a TFA), tak v nástrojích uživatelského rozhraní. V *jEDF* například nelze vybrat pouze určitou množinu kanálů či obraz přiblížit na určitý výřez, což jsou funkce plně podporované *EEGLABem*.

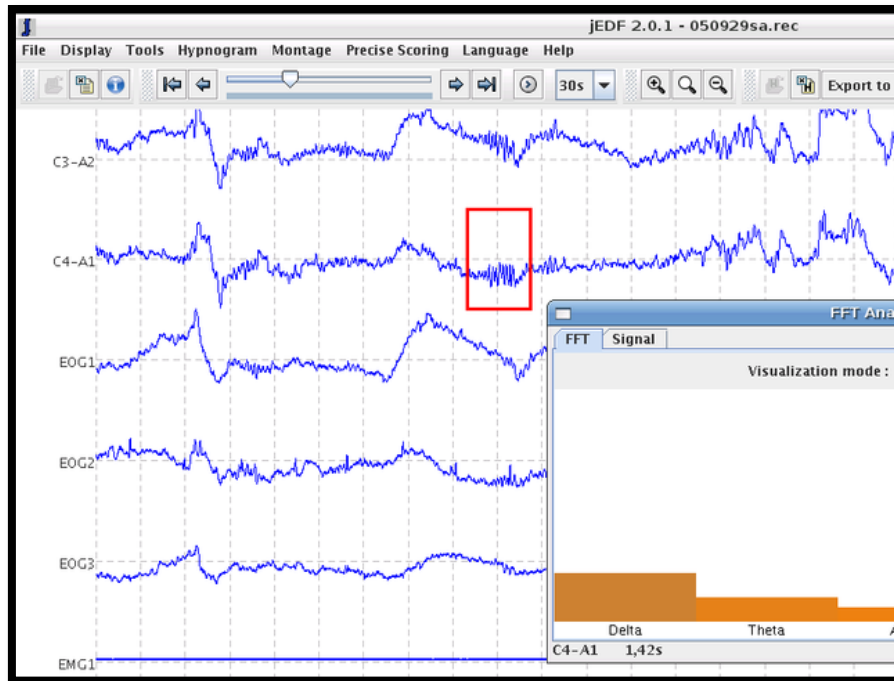
Nicméně zaměřenost programu *jEDF* na datový formát EDF není na škodu. Výhodou je také možnost převést data z formátu EDF(+) do běžného textového souboru, což je funkce, která není podporovaná toolboxem *EEGLAB*. Dalším plusem *jEDF* je fakt, že autor uvolnil jeho zdrojové kódy k nekomerčnímu použití. Je tedy možné přímo nahlédnout do zdrojového kódu, nebo jej modifikovat. Další informace týkající se tohoto prohlížeče jsou dostupné z [8].

Hlavní výhody programu *jEDF* z pohledu této práce:

- Podpora formátu EDF(+).
- Program je přenositelný mezi operačními systémy Windows a Linux
- K provozování programu stačí běžně dostupné JAVA JRE
- Možnost procházet data po časové ose
- Jednoduchá rozšiřitelnost (zdrojové kódy jsou veřejně dostupné).

Následuje přehled některých nevýhod programu *jEDF*:

- Špatné zobrazení EKG, které je „přibaleno“ k datům z EEG (viz kapitola 4.6)
- Téměř neexistující dokumentace



Obrázek 4.2 - Ukázka grafického prostředí programu jEDF<sup>14</sup>

#### 4.4.3 EDFbrowser

Program *EDFbrowser* (viz obrázek 4.3) je dílem nizozemského inženýra Teunise van Beelena. Opět se jedná o open-source EDF prohlížeč, který byl vytvořen v programovacím jazyce C++. Ačkoliv byl tento program vyvíjen pro operační systém Linux, existuje i jeho verze pro operační systém Windows. Lze tedy tvrdit, že jde o program multiplatformní.

V porovnání s výše uvedenými programy bych *EDFbrowser* zařadil funkčností mezi *EEGLAB* a *jEDF*. Program je zaměřen na datové formáty EDF, BDF, EDF+ a BDF+, což není tak komplexní jako u *EEGLABu*, ale v oblasti nástrojů nabízí mnohem širší možnosti, než *jEDF*. Mezi hlavní přednosti *EDFbrowseru* patří: možnost vybrat pouze specifické kanály, detailní nastavení parametrů příslušných jednotlivým kanálům, statistiky ke konkrétním kanálům (maximum, minimum, průměr, apod.), či široké možnosti importu (do EDF lze zkonvertovat např. formáty *Nihon Kohden*, či *Emsa*). Další informace týkající se tohoto programu jsou dostupné v [9].

Hlavní výhody programu *EDFbrowser* z pohledu této práce:

- Podpora formátu EDF(+).
- Program je přenositelný mezi operačními systémy Windows a Linux
- K provozování programu není potřeba knihoven třetích stran.
- Možnost procházet data po časové ose
- Funkce zoom.

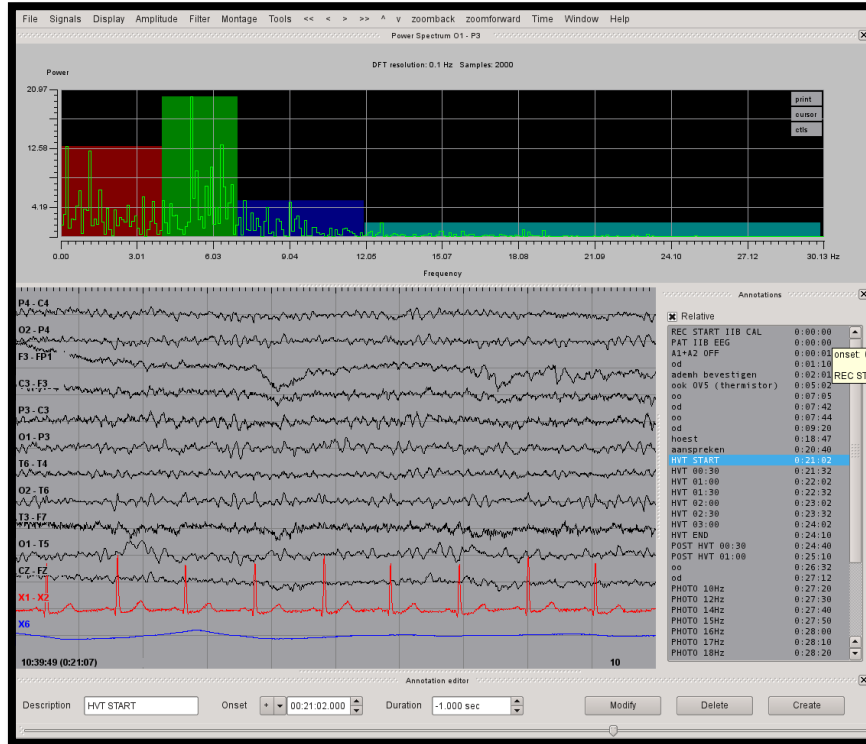
<sup>14</sup> Zdroj obrázku:

[http://freecode.com/screenshots/bd/e6/bde65720258dbefb60f1b844e002bf36\\_medium.png](http://freecode.com/screenshots/bd/e6/bde65720258dbefb60f1b844e002bf36_medium.png)

- Kvalitní dokumentace.
- Podpora anotací.

Následuje přehled některých nevýhod programu *EDFbrowser*:

- Špatné zobrazení EKG, které je „přibalené“ k datům z EEG (viz kapitola 4.6)



Obrázek 4.3 - Ukázka grafického prostředí programu *EDFbrowser*<sup>15</sup>

## 4.5 Testovací data

Klíčovou položkou pro testování zvolených EDF prohlížečů byl vzorek reálných dat ve formátu EDF. S požadavkem na několik souborů s EEG záznamy jsme se obrátili na neurologickou kliniku.

Nutno dodat, že získání těchto dat nebylo dílem okamžiku. Díky procesu anonymizace, kdy jsou z výsledků vyšetření odstraňována všechna citlivá data, zejména osobní údaje pacienta a vytížení pracoviště, došlo při získávání dat k nezanedbatelné časové ztrátě.

Vzorek několika dat nám byl nakonec poskytnut a my jsme mohli otestovat, jak si s ním poradí námi vybrané EDF prohlížeče.

Obsahem testovacích souborů byla pouze lékařská data bez jejich propojení na konkrétní osobu, jejímž vyšetřením tato data vznikla. Pro řešitele této práce není možné (a ani být nemělo) spojit obdržaná data s konkrétním pacientem. Data obsahovala kromě EEG údajů také signály zachycující srdeční rytmus (EKG).

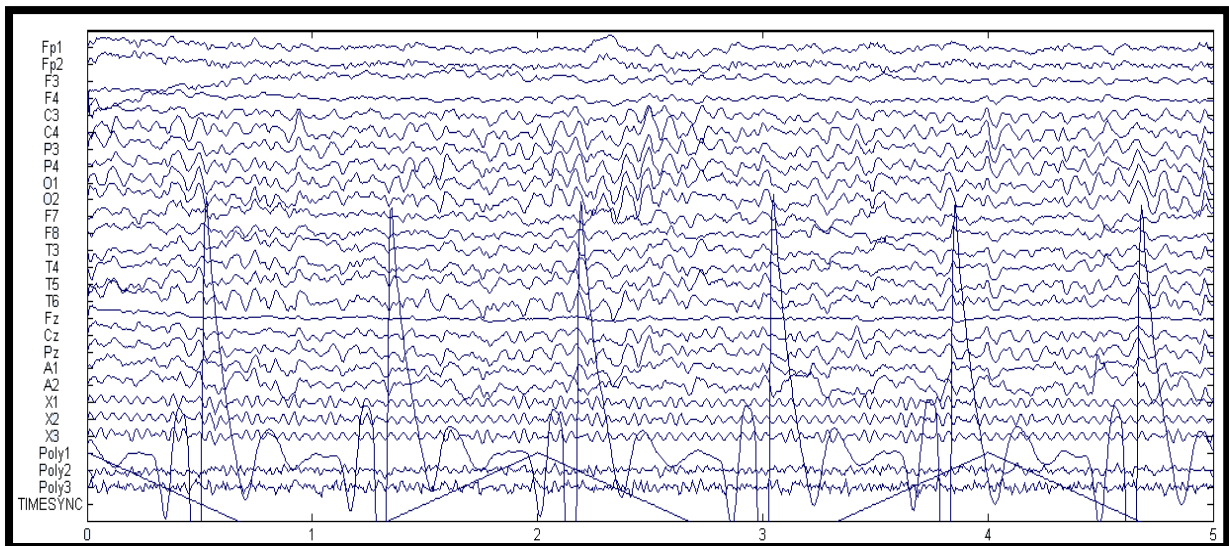
<sup>15</sup> Zdroj obrázku: [http://www.teuniz.net/edfbrowser/edfbrowser\\_2\\_big.png](http://www.teuniz.net/edfbrowser/edfbrowser_2_big.png)

## 4.6 Test zvolených EDF prohlížečů

Při zobrazování testovacích dat ve vybraných EDF prohlížečích vyšla najevo společná slabina všech těchto programů. Problémovým se ukázal být doplňkový signál zaznamenávající EKG a také signál TIMESYNC.

Přidání dalších signálů k EEG datům je v lékařství běžná praxe. V tomto případě slouží signál s EKG k detekci artefaktů a to zejména těch biologických (viz kapitola 2.2.3). Signál TIMESYNC zaznamenává hodinový impuls.

Problémem obou výše zmíněných signálů je, že mají oproti EEG signálům jiné charakteristiky, zejména pak frekvenci. Ani jeden z vybraných prohlížečů neumí správně zobrazit data z měření EEG spolu s přidávanými signály pro EKG a TIMESYNC. Toto se projevuje tak, že EEG data jsou v určitých místech překryta signály s vyšší frekvencí (viz Obrázek 4.4). Na tomto obrázku je zřetelně vidět, jak signál pro srdeční rytmus a hodiny přesahuje do oblasti jednotlivých kanálů. Signál s trojúhelníkovým průběhem je TIMESYNC, periodicky se opakující signál je pak EKG.



Obrázek 4.4 - Problém s přidáním signálů EKG a TIMESYNC při zobrazení dat v programu EEGLAB

Z výše uvedeného je patrné, že žádný z prohlížečů není schopen zobrazit naše data správně. Každý z nich bude vyžadovat úpravu zobrazovacího algoritmu tak, aby signály pro EKG a TIMESYNC byly zobrazeny s odpovídající frekvencí.

U prohlížeče *EEGLAB* bude postup úpravy pravděpodobně nejjednodušší. S výhodou bude možné využít existující funkci vytvoření výřezu, kterou budeme moci upravit tak, aby s její pomocí bylo možné vybraný segment uložit.

Program *jEDF* obsahuje velmi omezené možnosti vymezení konkrétního úseku dat a jeho úpravy by patrně zabrali největší množství času.

Prohlížeč *EDFbrowser* disponuje podobnou funkčností jako *EEGLAB*, ale nemůže se mu rovnat z hlediska využitelnosti. Okolo *EEGLABu* se za dobu jeho existence vytvořila poměrně široká komunita uživatelů, což do jisté míry zaručuje větší spolehlivost.

Z našeho testování vyšlo najevo, že nejvhodnějším základem pro budoucí aplikaci bude program *EEGLAB*, který obsahuje velké množství již implementovaných funkcí a zároveň je možné do něj snadno přidat další moduly.

## 4.7 Požadavky na budoucí program

Se znalostmi možností programu *EEGLAB* jsme mohli specifikovat konkrétní požadavky, jež bude nutné implementovat v našem programu. Výsledný program bude lékařům umožňovat:

- Zobrazení dat s výsledky měření EEG z programu PL-Winsor.
- Procházení dat po časové ose
- Funkce zoom.
- Výběr pouze některých kanálů.
- Výběr pouze určitého výřezu dat (segmentu).
- Identifikace konkrétního segmentu a její přidružení ke konkrétnímu pacientovy.
- Uložení segmentu do IS MediCalc.

Navrhované řešení nabylo obrys stand-alone aplikace, která bude schopna zpracovat výsledky vyšetření z programu PL-Winsor a dále umožní lékařům jejich procházení a výběr zajímavých částí. Takto vzniklé segmenty bude možné uložit do IS a přidružit je ke konkrétnímu pacientovi.

## 4.8 Prezentace navrženého řešení

S připravenou představou o budoucí aplikaci jsme společně s vedoucí práce prezentovali tuto zástupcům neurologické kliniky. Avšak naše schůzka odhalila závažné nedostatky. Následuje jejich bodový seznam.

1. Jako základ pro budoucí program bude použit software třetích stran.
  - Námí zvolený prohlížeč není navržen pro využití v klinické praxi.
2. Vybraný prohlížeč vyžaduje pro svou funkčnost instalaci dalšího softwarového vybavení (MATLAB runtime).
  - Zde si dovolím připomenout, že pracoviště obsahuje dvě pracovní stanice, jednu s OS MS DOS, druhou s OS Windows 2000. Obě tyto stanice jsou nutné pro správnou funkčnost elektroencefalografu. Aby byla funkčnost tohoto přístroje zajištěna, poskytuje firma Walter-Graphtek servis v režimu 24/7. Podmínky tohoto servisu však zahrnují minimální modifikace těchto dvou cílových stanic. Z tohoto důvodu není možné žádnou z nich jakkoliv programově upravovat.



3. Bylo by nutné nově proškolit personál a detailně jej seznámit s ovládáním nového programu.

- Personál neurologického oddělení je v současné době detailně obeznámen s ovládáním lékařského IS MediCalc a programu PL Winsor dodávaného firmou Walter-Graphtek. Dát lékařům k dispozici další program, který bude nutno komplexně ovládat, představuje nadbytečnou zátěž.

Z výše uvedených důvodů je patrné, že cesta skrze rozšíření již existujícího programu není schůdná a je nutné přijít s novým a lepším návrhem, který vyřeší daný problém tak, aby bylo řešení pro nemocnici přijatelné.

## 5 Návrh alternativního řešení

---

Schůzka s vedením neurologické kliniky nepřinesla jen samá negativa. Novým faktem bylo představení několika nových funkcí IS MediCalc. Tento systém slouží nejen pro ukládání výsledků z vyšetření, ale i pro jejich zobrazení prostřednictvím interního PACS prohlížeče s názvem *m4pacs*. Tento systém je tak používán, že je vnímán jako jakýsi standard.

Proto se nabízí možnost směřovat aplikaci jako doposud s tím rozdílem, že jako základ nebude použit žádný software třetích stran (jako např. *EEGLAB*), ale přímo informační systém nemocnice, konkrétně jeho PACS prohlížeč.

Výše uvedeným způsobem získáme všechny výhody, které *m4pacs* nabízí, jako např. funkce zoom, označení částí záznamu, apod., bez nutnosti softwarové úpravy pracovních stanic v laboratoři nemocnice. Program MediCalc, a jeho komponenta *m4pacs*, je již v laboratoři nainstalován.

### 5.1 PACS a jeho význam

PACS je zkratkou anglického sousloví Picture Archiving and Communication system. Cílem těchto, ve zdravotnictví používaných, systémů je poskytnout lékařům přístup k uložení a zobrazení zdravotnických dat. Zajímavé je, že tato data mohou vzniknout v několika modalitách, což znamená, že data mohou být generována různými vyšetřeními a v různých formátech. PACS tedy eliminuje nutnost manuálně přenést obraz z místa vyšetření k ošetřujícímu lékaři. Vše je vyřešeno elektronicky pomocí tohoto systému, takže lékař může přes síť přistupovat přímo k úložišti se zdravotnickými daty, kde si může vyhledat a zobrazit požadovaný záznam.

Obecně lze tvrdit, že každý PACS systém má čtyři základní komponenty:

1. Jednotlivé modalities (CT, MR, XA).
2. Zabezpečené síťové připojení pro přenos údajů o pacientovy.
3. Pracovní stanice pro vyhledání a zobrazení dat.
4. Servery pro ukládání dat.

Aby bylo možné dosáhnout takové úrovně modality, je nutné data uložit v přesně daném formátu. Toho je dosaženo díky implementaci univerzálního DIOCM standardu.

Pro zobrazení dat v prostředí systému PACS je však nutné data uložit v DICOM formátu. Naše data jsou však ve formátu EDF. Abychom tedy mohli zakomponovat funkci výběru a uložení segmentu do *m4pacs* je nutné nejdříve data převést z formátu EDF do formátu DICOM.

## 6 Převod formátu EDF na DICOM

### 6.1 Specifikace formátu EDF(+)

Okolnosti vzniku tohoto formátu pro uchování dat z měření mozkové aktivity jsou stručně shrnuty v kapitole 4.3.1. Následuje přesný popis vnitřní struktury tohoto binárního formátu.

EDF soubor se skládá z hlavičky (anglicky *header*) za kterou následují datové záznamy (anglicky *data record*). Obsahem hlavičky, která má proměnnou délku, jsou především data identifikující pacienta. Hlavička dále obsahuje informace o technických podmínkách, ve kterých bylo prováděno měření (přístroj, jeho typ, apod.) a také charakteristiky měřených signálů. Datová část obsahuje po sobě jdoucí epochy, které tvoří samotný záznam. Každá z epoch má pevnou dobu trvání.

Prvních 256 bajtů hlavičky specifikuje verzi EDF formátu, ve kterém byl záznam pořízen, lokální identifikaci pacienta a záznamu jako takového, časové informace o záznamu, počet naměřených vzorků dat a konečně počet signálů (dále NS). Dále pro každý signál následuje dalších 256 bajtů hlavičky, které udávají, o jaký typ signálu se jedná (EEG, EKG, tělesná teplota, apod.), amplitudu signálu, počet vzorků v každém datovém záznamu (epoše). Známe-li počet vzorků v každém datovém záznamu a jeho dobu trvání, můžeme odvodit vzorkovací frekvenci. Tímto způsobem je dosaženo toho, že každý signál může být vzorkován s jinou frekvencí. Hlavička celkem obsahuje bajtů. Tabulka 6.1 zachycuje detailní pohled na jednotlivá pole EDF hlavičky.

Hlavička EDF souboru (EDF Header)			
Velikost	Popis	Příklad	Poznámka
8 ASCII	Verze použitého formátu EDF	0	-
80 ASCII	Lokální identifikace pacienta.	-	Doporučuje se dodržet krok 3 z kapitoly 6.1.1.
80 ASCII	Lokální identifikace záznamu.	-	Doporučuje se dodržet krok 4 z kapitoly 6.1.1.
8 ASCII	Datum začátku záznamu.	dd.mm.yy	Doporučuje se dodržet krok 2 z kapitoly 6.1.1.
8 ASCII	Čas začátku záznamu.	hh.mm.ss	-
8 ASCII	Počet bajtů hlavičky.	-	-

Hlavička EDF souboru (EDF Header)			
Velikost	Popis	Příklad	Poznámka
44 ASCII	Rezervováno.	-	-
8 ASCII	Počet datových záznamů.	-1 pokud není přesně známo	Doporučuje se dodržet krok 10 z kapitoly 6.1.1.
8 ASCII	Doba trvání datového záznamu v sekundách.	-	-
4 ASCII	Počet signálů (NS) v datovém záznamu.	-	-
NS · 16 ASCII	Popisek každého kanálu.	EEG Fpz-Cz, tělesná teplota, apod.	Doporučuje se dodržet krok 9 z kapitoly 6.1.1.
NS · 80 ASCII	Typ převodníku pro každý kanál.	AgAgCl elektroda	-
NS · 8 ASCII	Fyzikální dimenze (jednotky) každého ze signálů.	μV, °C, apod.	-
NS · 8 ASCII	<i>Physical Minimum</i> každého signálu.	-500 nebo 34	-
NS · 8 ASCII	<i>Physical Maximum</i> každého signálu.	500 nebo 40	-
NS · 8 ASCII	<i>Digital Minimum</i> každého signálu.	-2048	-
NS · 8 ASCII	<i>Digital Maximum</i> každého signálu.	2048	-
NS · 80 ASCII	Hodnota <i>Prefiltering</i> každého signálu.	HP: 0.1 Hz LP: 75 Hz	-
NS · 8 ASCII	NS · počet vzorků v každém datovém záznamu	-	-
NS · 8 ASCII	NS · rezervováno	-	-

Tabulka 6.1 - Přehled datových polí v hlavičce EDF souboru

Textové informace obsažené v hlavičce ve formě ASCII řetězců musí být zarovnány vlevo a vyplněny mezerami tak, aby bylo vždy dosaženo jejich plné délky. Dejme tomu, že v poli pro identifikaci pacienta (které má velikost 80 bajtů) jsme vyplnili pouze „Miroslav Vacek“ (tj. řetězec o velikosti 14 bajtů). Velikost tohoto pole je však 80 bajtů, a proto musí zbývajících 66 bajtů vyplněno znakem mezera.

Půlnoc reprezentuje čas „00:00:00“. Dále se také doporučuje, aby každý datový záznam trval celý počet vteřin a jeho velikost (počet bajtů) nepřesáhla 61 440.

Hodnota pole hlavičky „*Digital Minimum*“ a „*Digital Maximum*“ každého signálu by měla zahrnovat extrémní hodnoty, které se mohou vyskytnout během záznamu. Toto jsou často extrémní výstupní hodnoty použitého A/D převodníku. Pole „*Physical Minimum*“ a „*Physical Maximum*“ každého signálu by mělo odpovídat výše uvedeným hodnotám polí *Digital Min/Max* pro konkrétní kanál a udávat fyzickou (a často také fyziologickou) dimenzi signálu. Tyto čtyři extrémní hodnoty udávají offset a zesílení signálu.

Datovou část EDF souboru tvoří po sobě jdoucí datové záznamy obsahující „*Doba trvání datového záznamu*“ sekund každého z NS signálů, přičemž každý signál je reprezentován specifickým (zadaným v hlavičce) počtem vzorků. Aby se omezila velikost takto uložených dat a bylo dosaženo lepší kompatibility s běžně užívanými programy pro zobrazení a zpracování EEG dat, je každá hodnota uložena jako dvoubajtové celé číslo (*integer*) v druhém doplňku. Tabulka 6.2 blíže ukazuje formát každého z datových záznamů.

Datové záznamy (EDF Data Record)	
Velikost	Popis
Počet_vzorků[1] · integer	První signál v datovém záznamu.
Počet_vzorků[2] · integer	Druhý signál.
...	
Počet_vzorků[NS] · integer	Poslední signál.

Tabulka 6.2 - Struktura samotných dat v EDF souboru

Další informace o formátech hlavičky a datových záznamů lze nalézt v [10].

### 6.1.1 Dodatečná specifikace EDF+

Jak je uvedeno ve stručné historii vzniku formátu EDF v kapitole 4.3.1, formát EDF byl po čase doplněn specifikací s označením EDF+. EDF+ si klade za cíl rozšířit možnosti použití formátu EDF, přičemž se snaží být zpětně kompatibilní s původním formátem EDF. Proto uveřejnili pánové Bob Kemp a Jesus Olivan několik doporučení, po jejichž splnění bude zajištěna kompatibilita formátů EDF a EDF+. Stručný popis těchto doporučení následuje.

1. V hlavičce EDF souboru by měli být použity pouze tisknutelné znaky US-ASCII s ASCII hodnotami mezi 32-126.
2. Pole hlavičky „*Datum začátku záznamu*“ a „*Čas začátku záznamu*“ by měli obsahovat pouze číslice 0-9 se znakem „tečka“ (.) použitým jak oddělovač.
  - Např.: „02.05.51“ jako hodnota pole „*Datum začátku záznamu*“ využívá rok 1985 k „oříznutí“, čímž se zabrání problému Y2K. To znamená, že roky 1985-1999 musí být reprezentovány jako yy = 85-99 a roky 2000-2084 jsou reprezentovány hodnotami yy = 00-84. Všechna

data následující po roce 2084 musí být uvedena ve tvaru, který je blíže popsán v kroku 4).

3. Pole hlavičky „*Lokální identifikace pacienta*“ musí začínat takzvaným polem „*subfields*“. Toto pole je vlastně skupina ASCII znaků, která neobsahuje mezery (ale používá je jako oddělovače záznamů) a obsahuje:

- Identifikační kód pacienta, pod kterým je tento veden v nemocničním informačním systému.
- Pohlaví (zde jsou přípustné pouze znaky M a F, z anglického Male a Female, což v překladu znamená muž, resp. žena).
- Datum narození pacienta ve formátu **dd-*MMM*-*yyyy***. Jako *MMM* je možné doplňovat pouze anglické třípísmenné zkratky názvů měsíců, které musí být uvedeny velkými písmeny. Např. **02-AUG-1951** je validní reprezentace data narození, která udává, že pacient se narodil 2. 8. 1951, zatímco **2-AUG-51** již není formátem EDF+ přípustná.
- Jméno pacienta.
- Pokud některé z výše uvedených údajů nejsou známy, či musí zůstat anonymní, je nutné je nahradit následujícím znakem: **X**.
- Pole „*subfields*“ může obsahovat i další položky mimo těch zmíněných v tomto kroku, avšak tyto musí následovat až po vyplnění identifikačního kódu, pohlaví, data narození a jména pacienta.
- Jakákoliv mezera v identifikačním čísle pacienta, které používá konkrétní nemocnice, či v jeho jménu musí být nahrazena jiným znakem, např. podtržítkem.

Pole „*Lokální identifikace pacienta*“ může tedy začínat následujícím řetězcem „*subfields*“: **MCH-0234567 F 02-MAY-1951 Haagse\_Harry**, kde jsou známy všechny požadované údaje.

4. Pole hlavičky „*Lokální identifikace záznamu*“ musí začínat takzvaným polem „*subfields*“. Toto pole je vlastně skupina ASCII znaků, která neobsahuje mezery (ale používá je jako oddělovače záznamů) a obsahuje:

- Následující text: „**Startdate**“.
- Samotné datum začátku záznamu ve formátu **dd-*MMM*-*yyyy***. Jako *MMM* je možné doplňovat anglické třípísmenné zkratky názvů měsíců, které musí být uvedeny velkými písmeny.
- Nemocniční číslo, které jednoznačně identifikuje tento záznam.
- Kód pracovníka zodpovědného za provedení vyšetření.

- Kód specifikující konkrétní přístroj, na kterém bylo vyšetření prováděno.
- Pokud některé z výše uvedených údajů nejsou známy, či musí zůstat anonymní, je nutné je nahradit následujícím znakem: **X**.
- Jakákoliv mezera ve výše uvedených údajích musí být nahrazena jiným znakem, např. podtržítkem.
- Pole „*subfields*” může obsahovat i další položky mimo těch zmíněných v tomto kroku, avšak tyto musí následovat až po vyplnění všech povinných údajů.

Pole „*Lokální identifikace záznamu*“ může tedy začínat následujícím řetězcem „*subfields*“: **Startdate 02-MAR-2002 PSG-1234/2002 NN Telemetry03**, kde jsou známy všechny požadované údaje. Pokud nejsou známy žádné údaje identifikující záznam, bude hodnota pole „*Lokální identifikace záznamu*“ následující: **Startdate X X X X**.

5. Hodnota pole hlavičky „*Digital Maximum*“ musí být větší než hodnota pole „*Digital Minimum*“. V případě, že je zisk zesilovače záporný, musí být odpovídající hodnota pole „*Physical Maximum*“ menší než hodnota pole „*Physical Minimum*“. Hodnota „*Physical Maximum*“ se také musí lišit od hodnoty „*Physical Minimum*“.
6. Nikdy nepoužívat oddělovač tisíců u číselných hodnot. U desetinných čísel je nutné použít desetinnou tečku, místo desetinné čárky.
7. Hodnoty signálů (celočíselné hodnoty o velikost dva bajty uložené v druhém doplňku) musí být uloženy ve formátu „*Little Endian*“ (tzn. LSB bit je uložen jako první).
8. Hodnota pole „*Čas začátku záznamu*“ by měla být uveden v časovém pásmu odpovídajícím místu měření.
9. Na následující adrese jsou dodatečné texty sloužící jako doplňkové informace k určení polaritě zesilovače či vysvětlení některých neurologických pojmů: <http://www.edfplus.info/specs/edftxts.html>.
10. Pole „*Počet datových záznamů*“ může nabývat hodnoty -1 pouze v průběhu nahrávání záznamu. Jakmile je záznam ukončen, je již známo správné číslo udávající přesný počet datových záznamů a musí být proto uloženo právě do pole „*Počet datových záznamů*“.
11. Pokud byly na signál v průběhu jeho pořízení použity filtry (např. horní propust' - HP, či dolní propust' - LP), a to nejlépe automaticky, je nutné jejich použití specifikovat v poli hlavičky „*Prefiltering*“ např. jako „HP: 0.1 Hz, LP: 75 Hz“. Pokud EDF soubor se záznamem obsahuje výsledky analýzy,

měly by být v poli „*Prefiltering*“ obsaženy odpovídající parametry se kterými byla analýza prováděna.

12. Pole hlavičky „*Typ převodníku*“ by mělo obsahovat popis použitého senzoru jako např. AgAgCl elektroda nebo termistor.

Detailní přehled všech vlastností formátu EDF+ je k nahlédnutí v [11].

### 6.1.2 Příklad souboru uloženého v EDF formátu

Dále následuje příklad hlavičky souboru, ve kterém je zaznamenáno 24 hodin záznamu mozkové aktivity pacienta spolu s informacemi o jeho teplotě. Signál je vzorkován s frekvencí 500 Hz (EEG), resp. 0,1 Hz (teplota). Záznam začíná 16. září 1951 ve 20:35 a trvá 1440 minut (2880 · 30s). Za povšimnutí také stojí, že všechna pole jsou vyplněna ve formátu, který je přípustný jak formátem EDF, tak formátem EDF+, s výjimkou pole „*Rezervováno*“, které je podporováno pouze formátem EDF. Dále si také můžeme všimnout, že pokud není pole bezezbytku zaplněno, je zbytek jeho povolené délky vyplněn tzv. „bílymi znaky“, v tomto případě mezerami. Všechny údaje v následujícím příkladu (kód 6.1) jsou ponechány v anglickém znění.

```
0      MCH-0234567 F 16-SEP-1987 Haagse_Harry
Startdate 16-SEP-1987 PSG-1234/1987 NN Telemetry03
16.09.8720.35.00768      Reserved field of 44 characters
2880   30       2   EEG Fpz-Cz      Temp rectal
AgAgCL cup electrodes
Rectal thermistor

uV      degC   -440   34.4   510   40.2   -2048
-2048   2047   2047   HP:0.1Hz LP:75Hz N:50Hz
LP:0.1Hz (first order)
```

Kód 6.1 - Příklad hlavičky EDF souboru<sup>16</sup>

## 6.2 DICOM

Nyní již známe všechny podstatné informace o datovém formátu, ve kterém jsou exportována data z nemocničního programu PL Winsor. Je vhodné přejít k popisu datového formátu, do kterého budou tato ukládána.

### 6.2.1 Co je to vlastně DICOM?

DICOM je třetí verzi standardu vyvinutého *American College of Radiology* (ACR) a *National Electrical Manufacturers Association* (NEMA). Počátek tohoto standardu se datuje

<sup>16</sup> Převzato z: <http://www.edfplus.info/specs/edf.html>



od roku 1980, kdy bylo téměř nemožné, aby kdokoliv, kromě výrobců specifických lékařských přístrojů (např. přístroje pro magnetickou rezonanci – MR), byl schopen dekódovat výsledky vyšetření (většinou v obrazové formě – např. právě výsledky z MR), prováděných na těchto přístrojích. Zhruba v této době byl vznesen požadavek na otevřený datový formát, jenž by umožnil pracovištím po celém území USA, výměnu výsledků lékařských vyšetření bez nutnosti vlastnit naprosto stejný měřicí přístroj.

V reakci na tento požadavek spojili ACR a NEMA své síly a vytvořili v roce 1983 společnou standardizační komisi. Jejich první standard, ARC/NEMA 300 byl zveřejněn o dva roky později, v roce 1985. Avšak velice brzy po jeho vydání začalo být jasné, že bude zapotřebí mnoha vylepšení. Mezi hlavní výtky patřila vágnost textu, který navíc obsahoval hned několik protichůdných doporučení.

Rok 1988 proto přinesl druhou, přepracovanou, verzi výše zmíněného standardu. Nutno dodat, že této upravené verzi se dostalo mnohem větší míry přijetí než jejímu předchůdci, a to především mezi výrobci měřících zařízení. První demonstrace standardu ACR/NEMA V2.0 proběhla na univerzitě Georgetown mezi 21. – 23. květnem roku 1990. Cílem této demonstrace bylo předvést propojitelnost naprosto rozdílných přístrojů. Přítomno bylo celkem šest velkých výrobců měřící techniky. Jmenovitě se jednalo o DeJarnette Research Systems, General Electric Medical Systems, Merge Technologies, Siemens Medical Systems, Vortech a 3M. Komerční podpora standardu ACR/NEMA V2.0 byla oznámena na výročním sjezdu skupiny RSNA (*Radiological Society of North America*) koncem roku 1990 právě výše zmíněnými firmami.

I přese všechny jeho přednosti bylo brzy jasné, že i standard ACR/NEMA V2.0 bude potřebovat určité úpravy. Tyto vznikly ve formě rozšíření, mezi které patří např. Papyrus vyvinutý na *University Hospital of Geneva* ve Švýcarsku, či SPI (*Standard Product Interconnect*), navržený firmami Siemens Medical Systems a Philips Medical Systems.

První větší nasazení technologie ACR/NEMA bylo uskutečněno v roce 1992 USAAF (*US Army and Air Force*) jako součást programu MDIS (*Medical Diagnostic Imaging Support*). Společnosti Loral Aerospace a Siemens Medical systems stály v čele konsorcia několika společností, které vybavily většinu lékařských zařízení USAAF přístroji pro teleradiologii. Zde si dovoluji zmínit, že teleradiologie je přenos rentgenových snímků z jednoho místa do druhého, aby mohl být snímek následně vyhodnocen radiologem. V návaznosti na úspěch programu MDIS byly podobně vybaveny i kliniky USN (*US Navy*) a mnohá další pracoviště.

V průběhu roku 1993 byla vydána třetí verzi tohoto standardu. Také došlo ke změně názvu standardu na DICOM (*Digital Imaging and Communications in Medicine*), aby se zvýšila potenciální míra přijetí standardu. Ve verzi tři bylo definováno mnoho nových modulů (např. síťová podpora, či pravidla konformity).

Třetí verzi je oficiálně také verzi poslední, nicméně je tato verze od svého vzniku neustále doplňována a rozšiřována. Proto se na přesnou verzi standardu DICOM neodkazuje přesným číslem verze, ale rokem vydání této verze (např. DICOM verze 2007). Zatím posledního vylepšení se DICOM dočkal v roce 2011.

Ačkoliv se DICOMu dostalo téměř univerzálnímu přijetí a to jak mezi výrobci klinických přístrojů, tak nemocnicemi, má svá omezení. DICOM je především zacílen na otázky ohledně interoperability v oblasti pořizování lékařských snímků. Jako takový by v žádném případě neměl být chápán jako architektura pro modelování workflow v konkrétní nemocnici.

Z výše uvedeného je patrné, že DICOM není pouze formátem souboru, ale komplexním standardem, který obsahuje především:

- Definici síťového protokolu pro přenos lékařských snímků.
- Syntaxi a sémantiku příkazů, které mnohou být zadávány v rámci těchto protokolů-
- Formát datových souborů, ve kterém jsou lékařské snímky ukládány a přenášeny.

Pro naše další zkoumání je zajímavá hlavně poslední zmíněná položka DICOM standardu, tedy přesný formát uložení lékařských dat do souboru.

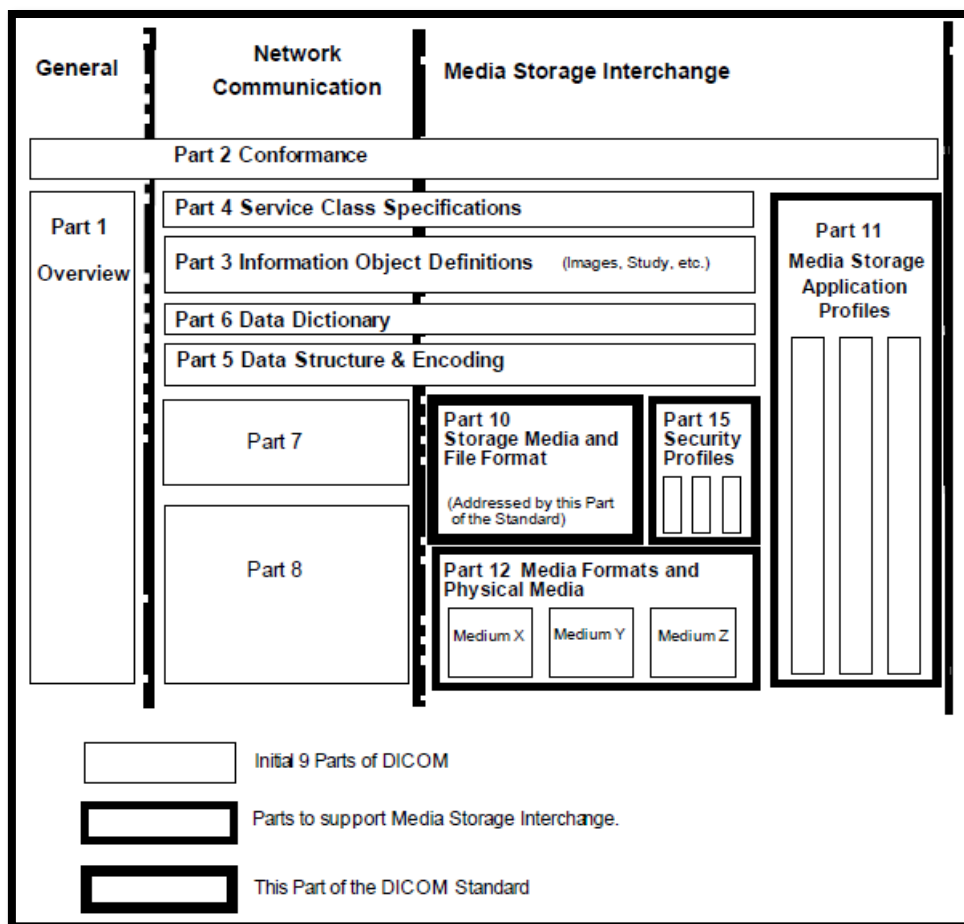
Bližší informace o vzniku a vývoji standardu DICOM lze nalézt v [12] a [13].

## 6.2.2 Datový formát DICOM

---

Na rozdíl od EDF(+) je DICOM komplexním standardem. DICOM je také primárně zaměřen na lékařské snímky (např. výsledky rentgenových vyšetření, či magnetické rezonance), ale EDF(+) bylo od samého počátku vytvořeno jako datový formát pro uchování signálových dat (např. EEG, či EKG). Proto je velice důležité pochopit, jak přesně jsou strukturovány DICOM soubory.

Pro začátek bych rád ukázal, jakou roli hraje specifikace DICOM souborů v rámci celého standardu DICOM verze 2011 (viz obrázek 6.1).



Obrázek 6.1 - Postavení formátu DICOM souborů (tučný rámeček) v rámci celého DICOM standardu verze 2011<sup>17</sup>

Nyní bych se rád věnoval přesnému členění DICOM souborů. Každý soubor musí obsahovat část s metainformacemi o uložených datech a následně samotná data (*Data Set*). Lze tedy vysledovat jistou paralelu s formátem EDF(+), který také používá členění souboru na dvě poloviny - na hlavičku a datové záznamy. Rámcovou strukturu DICOM souborů ukazuje obrázek 6.2.



Obrázek 6.2 - Rámcová struktura DICOM souboru<sup>18</sup>

Zaměříme se nyní podrobněji na strukturu metainformací, které tvoří první část DICOM souboru. Tato hlavička začíná 128 bajty preamble. Za touto preambulí následují 4 bajty DICOM prefixu a jednotlivé pole specifikující obsah souboru (seznam přípustných polí je uveden v tabulce 6.3). Hlavička je povinnou součástí jakéhokoliv DICOM souboru.

<sup>17</sup> Převzato z: [http://medical.nema.org/Dicom/2011/11\\_10pu.pdf](http://medical.nema.org/Dicom/2011/11_10pu.pdf), strana 18

<sup>18</sup> Upraveno podle: [http://medical.nema.org/Dicom/2011/11\\_10pu.pdf](http://medical.nema.org/Dicom/2011/11_10pu.pdf), strana 19

Zajímavé je, že ačkoliv preambule nevyžaduje, aby u ní byl specifikován DICOM tag a délka, je přípustné, aby v jejích 128 bajtech byla uložena data. DICOM standard tedy umožňuje do preambule vložit jakákoliv data, avšak jejich velikost nesmí za žádných okolností přesáhnout 128 bajtů. K tomu to kroku se NEMA uchýlila kvůli zvýšení kompatibility mezi různými datovými formáty, ve kterých jsou ukládány výsledky vyšetření.

DICOM prefix by měl obsahovat řetězec „**DICM**“ (bez uvozovek) zapsaný velkými písmeny v kódování podle normy ISO 8859 G0. Tento čtyřbajtový prefix opět není doplněn o DICOM tag a délku.

Jako poslední obsahuje hlavička následující meta elementy s dopovídajícími DICOM tagy a délkou (v bajtech). Jen pro úplnost dodávám, že pole tag slouží pro jednoznačnou identifikaci daného datového pole v rámci celého DICOM standardu. Jeho hodnota je uvedena ve tvaru (*gggg, eeee*), kde *gggg* specifikuje číslo skupiny, do které tag spadá a *eeee* udává přesný identifikátor elementu uvnitř dané skupiny. Přehled přípustných elementů následuje (tabulka 6.3).

Metainformace o DICOM souboru			
Název Atributu	Tag	Typ	Popis
Preamble	-	1	Pevná, 128 bajtů dlouhá, hlavička, která může být využita ke specifickému použití. Pokud není použita k žádnému speciálnímu účelu, měli by všechny bajty mít hodnotu <b>00</b> (hexadecimálně). Žádný program využívající DICOM soubory (např. čtečka), by neměl využívat tuto preambuli k určení, zda se opravdu jedná o DICOM soubor.
DICOM prefix	-	1	Čtyři bajty obsahující textový řetězec „ <b>DICM</b> “. Právě toto pole by mělo být použito při testu, zda se jedná o DICOM soubor.
File Meta Information Group Length	(0002, 0000)	1	Počet bajtů následujících od tohoto pole do posledního (včetně) pole skupiny 0002.

Metainformace o DICOM souboru			
Název Atributu	Tag	Typ	Popis
File Meta Information Version	(0002, 0001)	1	Toto je dvoubajtové pole, kde každý z bitu specifikuje verzi pole <i>File Meta Information</i> použité v aktuálním souboru. Ve verzi 1 bude první bajt nabývat hodnoty <b>00</b> (hexadecimálně), přičemž druhý bajt bude nastaven na hodnotu <b>01</b> (hexadecimálně).  Při kontrole tohoto pole by měl být čten pouze bit 0 (LSB – <i>Least Significant Bit</i> ) z druhého bajtu. Žádný další bit tohoto pole by neměl být kontrolován.
Media Storage SOP Class UID	(0002, 0002)	1	Jedinečný identifikátor SOP třídy asociované s množinou dat ( <i>Data Set</i> ). Identifikátory SOP tříd přípustné pro soubory s výsledky vyšetření jsou specifikovány v části 3.4 DICOM standardu <sup>19</sup> .
Media Storage SOP Instance UID	(0002, 0003)	1	Jedinečný identifikátor instance SOP třídy asociované s množinou dat-
Transfer Syntax UID	(0002, 0010)	1	Unikátní identifikátor „Transfer Syntax“ použité k zakódování množiny dat. Tato syntaxe se vztahuje pouze na data, nikoliv na metainformace obsažené v hlavičce.
Implementation Class UID	(0002, 0012)	1	Jednoznačně identifikuje implementaci, které vytvořila tento soubor a jeho obsah. Toto je vhodné využít, pokud se při přenosu tohoto souboru vyskytnou problémy. Přesné zásady použití tohoto pole definuje oddíl 3.7 DICOM standardu <sup>20</sup> .
Implementation Version Name	(0002, 0013)	3	Udává verzi předchozího pole (0002, 0012) ve formě až šestnácti znaků. Přípustné jsou velká písmena anglické abecedy ( <b>A-Z</b> ), arabské číslice ( <b>0-9</b> ) a znak podtržítka ( <b>_</b> ).

<sup>19</sup> Viz [http://medical.nema.org/Dicom/2011/11\\_04pu.pdf](http://medical.nema.org/Dicom/2011/11_04pu.pdf)

<sup>20</sup> Viz [http://medical.nema.org/Dicom/2011/11\\_07pu.pdf](http://medical.nema.org/Dicom/2011/11_07pu.pdf)

Metainformace o DICOM souboru			
Název Atributu	Tag	Typ	Popis
Source Application Entity Title	(0002, 0016)	3	Specifikuje tzv. <i>DICOM Application Entity</i> (AE), tj. organizaci, která vytvořila obsah stávajícího souboru (nebo jej naposledy upravovala). Pokud je toto pole použito umožňuje najít případný zdroj problémů v rámci přenosu mezi organizacemi.
Private Information Creator UID	(0002, 0100)	3	Jednoznačný identifikátor tvůrce položky s tagem (0002, 0102).
Private Information	(0002, 0102)	1C	Obsahuje privátní informace umístěné mezi metainformacemi o souboru. Toto pole je povinné, pokud je specifikován jeho tvůrce v předchozím poli s tagem (0002, 0100).

Tabulka 6.3 - Přehled přípustných metainformací o DICOM souboru

Ve všech polích mimo preambule, která má délku 128 bajtů, a čtyřbajtového DICOM prefixu, by mělo být použito kódování „Little Endian“. Každé pole by také mělo být v případě nutnosti vyplněno bílými znaky, aby bylo dosaženo jeho plné délky. Přesný formát bílých znaků, které jsou použity jako „výplň“ definuje sekce 3.5 DICOM standardu<sup>21</sup>. Pro nás je však důležité, že se ve většině použitých polí hlavičky používá znak „mezera“. Oddíl 3.5 DICOM standardu také specifikuje hodnotu pro údaje, jež nejsou známy (*Unknown* - UN). Tato hodnota by však neměla být použita u žádného pole obsahujícího metainformace o souboru. Z toho důvodu by měl být jakýkoliv tag, mimo ty uvedené v tabulce 6.3, ignorován a neměl by na něj brát ohled během parsování DICOM souborů. Toto opatření bylo zavedeno kvůli kompatibilitě s budoucími verzemi standardu DICOM.

Každý soubor také musí obsahovat jednu množinu dat (Data Set) reprezentující jednu instanci SOP třídy. Tato třída musí odpovídat údajům uvedeným v poli s tagem (0002, 0002).

Transfer Syntax použitá k zakódování množiny dat musí odpovídat hodnotě pole s tagem (0002, 0010). Tuto syntaxi také nelze přímo v datech měnit, tj. element „*Transfer Syntax UID*“ se nesmí objevit v samotných datech.

Za zmínku také stojí, že délka množiny dat není definována. To znamená, že nikde není uvedeno pole, které říká, jak velká jsou data (na rozdíl od EDF, jež pro tuto hodnotu používá v hlavičce pole s názvem „*Počet datových záznamů*“). Konec množiny dat v DICOM souboru je možné detekovat pouze na základě znaku „konec souboru“, jenž je umístěn vždy na konci dat.

<sup>21</sup> Viz [http://medical.nema.org/Dicom/2011/11\\_05pu.pdf](http://medical.nema.org/Dicom/2011/11_05pu.pdf)

Posledním elementem, který se může v množině dat vyskytovat, je element s tagem (FFFC, FFFC). Tento element je přípustný pouze tehdy, pokud je třeba všechny elementy množiny dat zarovnat na stejnou délku. Hodnota tohoto elementu nemá žádný význam a měla by být ignorována.

Detailní popis dalších vlastností DICOM souborů je dostupný v [14].

### 6.2.3 DICOM objekty

---

V předcházejícím textu bylo použito několik termínů, které si zaslouží blíže vysvětlit. Konkrétně se jedná o termíny *Service Object Pair Class* (SOP třída) a *Transfer Syntax*.

*Service Object Pair Class* je jedním z termínů, jež má tendenci odradit všechny ty, kteří se o DICOM standard zajímají. Počátek tohoto termínu je v objektovém návrhu DICOM standardu. SOP však slouží pouze k jednoduché identifikaci rolí (*capabilities*), které se vztahují k danému objektu. Pomocí SOP je tedy možné identifikovat, že role „úložiště CT obrazů“ se vztahuje jak na přidružený objekt (CT obraz), tak na přidruženou službu (úložiště). Jen pro úplnost si dovoluji doplnit, že zkratka CT skrývá termín *počítačová tomografie* (anglicky *Computed Tomography*), což je metoda umožňující zobrazení vnitřních orgánů člověka. Všechny SOP třídy jsou jednoznačně určeny přiřazeným UID a názvem. V praxi lze rozlišit dva druhy SOP tříd:

- perzistentní objektové typy (např. objekty CT či MR),
- a tzv. „*Negotiable services*“ (např. dotazování nad perzistentními objekty, tzv C-FIND).

Přesný výčet a popis všech SOP tříd, spolu s jejich jednoznačnými identifikátory (UID) a jmény je dostupný v části 3.4 DICOM standardu.

Termín *Transfer Syntax* (TS) je v jazyku DICOM standardu použit k popisu struktury DICOM souboru a také ke specifikaci síťového přenosu. Toto pole je dáno následujícími třemi hodnotami

- VR: implicitní / explicitní,
- Endianismus: little / big endian,
- Použitá kompresní metoda.

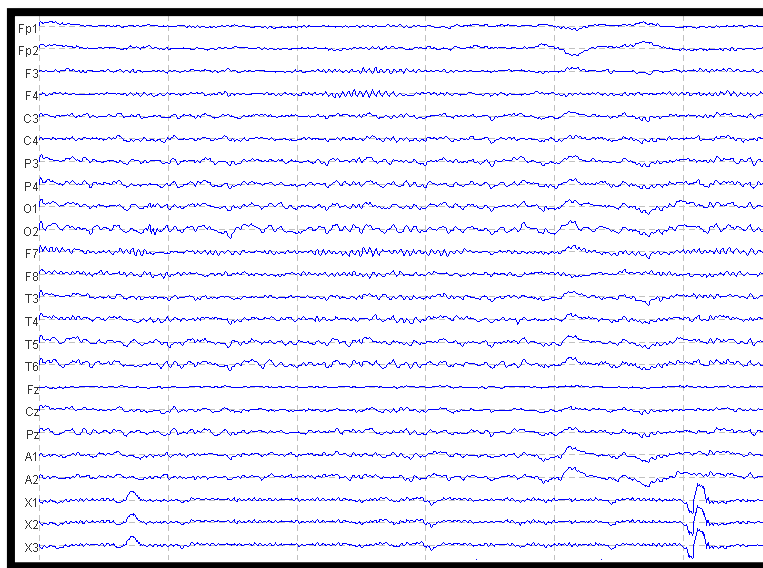
Následuje stručný popis nejpoužívanějších TS v rámci DICOM standardu:

- Nekomprimované
  - Implicitní VR Little-Endian
  - Explicitní VR Little-Endian
  - Explicitní VR Big-Endian
- Bezeztrátově komprimované
  - JPEG (bezeztrátový)
  - RLE (bezeztrátové)
  - JPEG 2000 (bezeztrátový)
- Ztrátově komprimované
  - JPEG Baseline

- JPEG Extended
- JPEG 2000 (ztrátový)

### 6.3 Porovnání EDF(+) a DICOM

Nyní již známe strukturu obou formátů, které se účastní konverze, což přináší následující poznatky. Tím zásadním je, že formát EDF(+) je určen pouze pro signálová data, zatímco DICOM formát souborů pracuje výhradně s formátem obrazovým. Pro lepší zdůraznění rozdílů si dovoluji uvést následující dva obrázky. Na obrázku 6.3 je signálový záznam z měření mozkové aktivity pacienta. Data jsou uložena ve formátu EDF. Obrázek 6.4 pak ukazuje, jak vypadá běžný CT snímek oblasti hlavy, který je uložen v DICOM formátu. Za poskytnutí dat zobrazených na obrázcích bych rád poděkoval neurologické klinice FN Lochoťín.



Obrázek 6.3 - Příklad signálových dat



Obrázek 6.4 – Příklad obrazových dat

Další zajímavostí je, že oba formáty umožňují uložení jednoznačné identifikace pacienta. Tento údaj je nesmírně důležitý, neboť na jeho základě bude možné uložit výřez z konkrétního souboru do IS MediCalc.



## 6.4 Popis dostupných knihoven

Dalším krokem algoritmu pro konverzi z formátu EDF do formátu DICOM je implementace obou výše zmíněných formátů. V zásadě se nabízejí dvě možnosti. První z nich je, že si knihovny pro práci s oběma formáty napíšeme sami. Druhou je pak možnost využít některé z existujících nekomerčních řešení.

Byť by se mohlo zdát, že první možnost bude praktičtější, není tomu tak a to hned z několika důvodů. Jedním z nich je časová náročnost na samotnou implementaci obou formátů. Formát EDF je sice veskrze jednoduchý, a jeho naprogramování by bylo otázkou dnů, ale je nutné si uvědomit, že samotnou implementací práce nekončí. Protože se jedná o základní komponentu, na které se bude stavět, musíme si být naprosto jistí, že bude fungovat správně. To v praxi znamená zaměřit se důrazně na testování. Nelze připustit, aby došlo ke špatnému přečtení vstupních dat z měření EEG pacienta. Tato chyba by dále mohla „probublat“ do formátu DICOM a způsobit tak, že lékaři budou vycházet z neúplných či dokonce chybných dat. A právě časová náročnost testování algoritmů implementujících standardy EDF a DICOM je hlavní důvodem pro zamítnutí vytvoření vlastních knihoven pro práci s těmito formáty.

Proto jsme se rozhodli prozkoumat pole s již existujícími implementacemi obou formátů a vybrat z dostupných zdrojů vhodný nástroj. Výhodou tohoto přístupu bude úspora času, který by jinak bylo nutné strávit programováním a zejména pak testováním vlastních knihoven. Protože oba použité formáty mají poměrně široké zastoupení, existuje pro každý hned několik knihoven, které nabízejí zapouzdření pro základní operace se soubory EDF a DICOM. Nutno také dodat, že většina těchto knihoven obsahuje prohlášení, jež zbavuje autory jakékoliv odpovědnosti za škodu způsobenou špatnou implementací EDF a DICOM knihoven. Tyto knihovny mají přesto dostatek uživatelů, aby bylo možné zjistit, zda ta která knihovna obsahuje závažné chyby, či nikoliv.

### 6.4.1 EDFlib

---

První knihovnou, o které bych se rád zmínil, je EDFlib od nizozemského autora Teunise van Beelena. EDFlib vznikla jako knihovna zaměřená na čtení a zápis EDF+ a BDF+ souborů, přičemž si „poradí“ i se staršími soubory ve formátu EDF. Tato knihovna je naprogramována v programovacím jazyce C a je určena pro použití v programech, jež jsou vyvíjeny v programovacích jazycích C a C++. Pro úplnost si dovoluji zmínit, že formát BDF je 24 bitovou nadstavbou formátu EDF.

Autor dále tvrdí, že knihovna byla testována pouze pod operačním systémem Linux s překladačem GCC, ale naše testování odhalilo, že ji lze úspěšně přeložit i v prostředí Microsoft Visual Studio 2010 pod operačním systémem Windows.

O jejích kvalitách je možné se přesvědčit např. v rámci iniciativy *Scilab*<sup>22</sup>, jež používá právě knihovnu *EDFLib* k tomu, aby umožnila uživatelům pracovat s EDF+ a BFD+ soubory. Toto zajišťuje velice vysokou míru otestování zdrojového kódu zmíněné knihovny.

Tato knihovna implementuje formát EDF v podobě, v jaké je uvedeno v kapitole 6.1 včetně všech zde uvedených doporučení (viz kapitola 6.1.1). Lze tedy přistupovat jak k datům z hlavičky, tak ke konkrétním signálům.

Konkrétní přehled funkcí této knihovny je dostupný z [15].

#### 6.4.2 DCMTK

DCMTK je sadou knihoven a aplikací, které implementují většinu částí DICOM standardu. Zahrnuje software pro testování, konstrukci a konverzi DICOM souborů. Umožňuje také přenášení DICOM souborů přes síť, či vytváření repositářů pro DICOM data.

Tato sada nástrojů je z části vytvořena v programovacím jazyce ANSI C a z části v jazyce C++. Původně byla vyvíjena pod operačním systémem Linux, ale v současné době existuje distribuce, kterou lze přeložit pro OS Windows pomocí překladače MS Visual Studia 2010. DCMTK je dostupný jako „open-source“. DCMTK se skládá z modulů, jež jsou uvedeny v tabulce 6.4.

Popis modulů toolkitu DCMTK	
Název Modulu	Stručný popis modulu
<i>config</i>	Konfigurační nástroje pro samotné DCMTK.
<i>dcmdata</i>	Knihovny a aplikace pro zakódování a dekodování DICOM dat.
<i>dcmimage</i>	Přidává podporu pro barevné obrázky do modulu „ <i>dmimgle</i> “.
<i>dcmimgle</i>	Podpůrné nástroje a knihovna pro zpracování DICOM obrazů.
<i>dcmjpeg</i>	Podpůrné nástroje a knihovna pro komprese a dekompresi obrázků.
<i>dcmjpls</i>	Podpůrné aplikace a knihovna pro komprese a dekompresi obrázků.
<i>dcmnet</i>	Síťová knihovna pro implementující část 8 DICOM standardu.
<i>dcmpstat</i>	Knihovna a nástroje pro zobrazení DICOM dat.
<i>dcmsign</i>	Knihovna a nástroje pro digitální podpis.
<i>dcmsr</i>	Knihovna a nástroje pro strukturované reporty.
<i>dcmtls</i>	Bezpečnostní rozšíření síťové knihovny „ <i>dcmnet</i> “.
<i>dcmwlm</i>	Databázový server pro <i>DICOM Modality Worklist Management Service</i> .
<i>dcmqrdb</i>	Databázové server pro ukládání DICOM dat.

<sup>22</sup> Viz: <http://www.scilab.org/>

Popis modulů toolkitu DCMTK	
Název Modulu	Stručný popis modulu
<i>oflog</i>	Knihovna pro logování založená na <i>log4cplus</i> .
<i>ofstd</i>	Knihovna pro třídy, které nejsou součástí DIOCM standardu, ale jsou využívány některými částmi DCMTK.

Tabulka 6.4 - Součásti DCMTK

Vzhledem k tomu, že DCMTK je spíše kolekcí nástrojů než jen knihovnou, je pro jeho bezproblémové používání nutné mít nainstalované některé z knihoven třetích stran. Konkrétní názvy a verze knihoven, jež jsou požadovány k sestavení DCMTK ukazuje tabulka 6.5.

Název knihovny	Minimální verze
<i>zlib</i>	1.2.5
<i>tiff</i>	3.9.4
<i>libpng</i>	1.4.3
<i>libxml</i>	2.7.7
<i>libiconv</i>	1.13.1

Tabulka 6.5 - Knihovny potřebné pro sestavení DCMTK

Z hlediska algoritmu konverze je zajímavý nástroj **img2dcm** z modulu „*dcmimgle*“, který umožní ze standardních obrázků (např. JPEG, či BMP) vytvořit DICOM soubor.

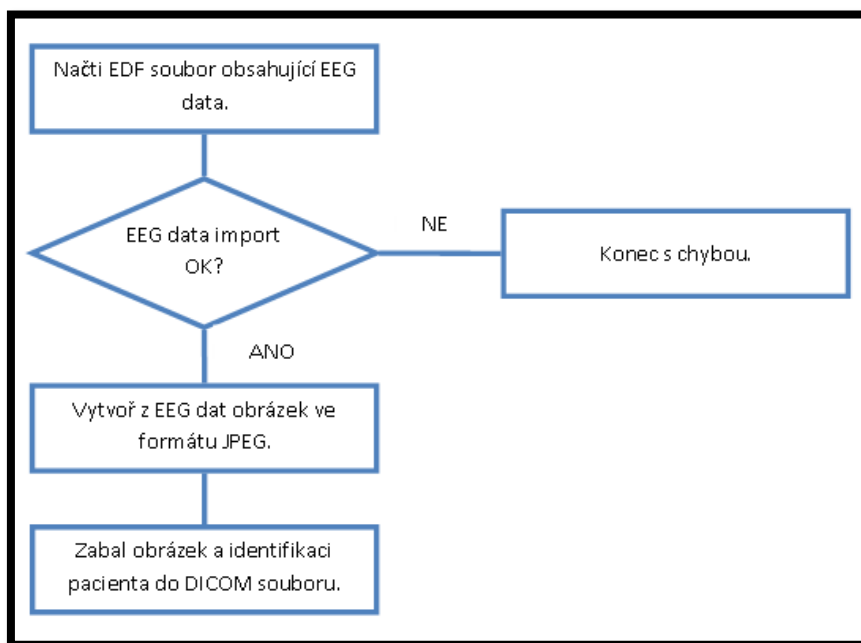
Další informace k DCMTK jsou dostupné v [16].

## 7 Popis algoritmu konverze

---

V této fázi je nutné rozhodnout, v jakém programovacím jazyce budeme program realizovat. Vzhledem k tomu, že IS MediCalc je implementován v jazyce C++, bude jeho rozšíření vytvořeno právě v tomto jazyku.

Na obrázku 7.1 je zachycen popis algoritmu konverze v podobně vývojového diagramu.



Obrázek 7.1 - Vývojový diagram zachycující funkčnost konverzního algoritmu

### 7.1.1 Načtení EEG souboru

---

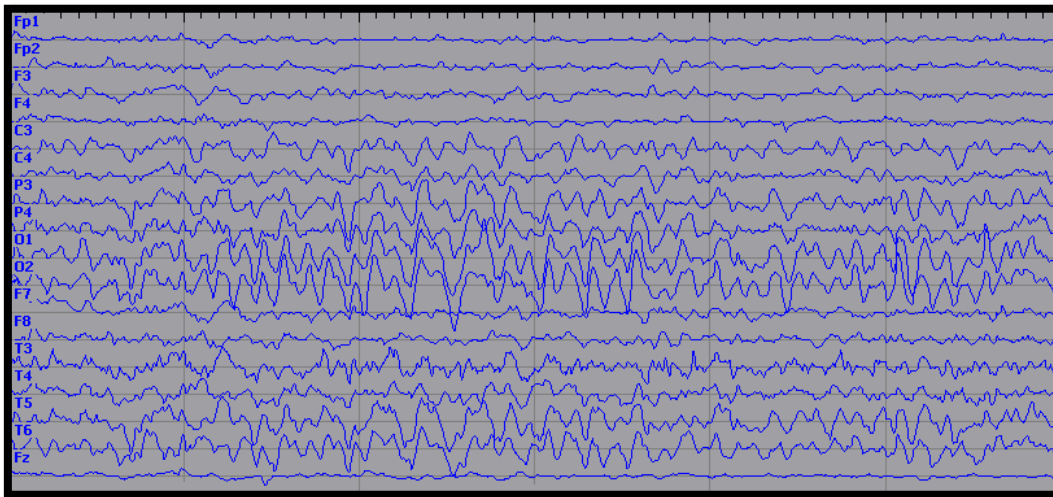
V první fázi dojde k načtení dat z měření mozkové aktivity z EDF souboru. K tomuto účelu bude využita knihovna EDFlib, která nabízí mnoho užitečných funkcí. V této fázi je nutné zkontrolovat, zda jsou v datech uloženy i identifikační údaje pacienta a záznamu. Tyto údaje budou klíčové při realizaci ukládání segmentu do IS:

Kontrolu zmíněných polí je nutné provést, protože se může stát, že program PL-Winsor, ze kterého se exportují data z měření, nemusí vždy vyplnit všechna pole z hlavičky EDF souboru. Někdy se může stát, že pole bez problému vytvoří, ale neuvede v nich žádnou validní hodnotu, což vede k chybovému stavu.

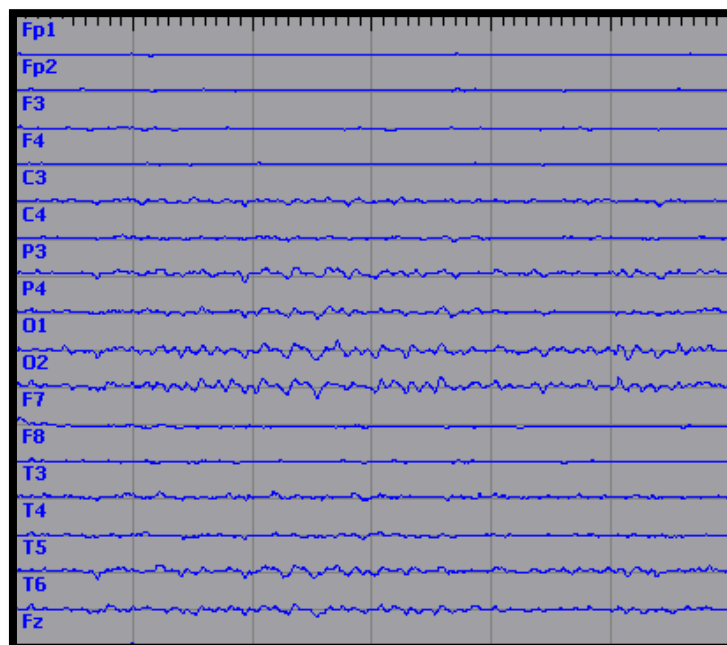
Obsahuje-li EDF hlavička všechna potřebná data, lze přistoupit k dalšímu kroku. Pokud ne, je nutné program ukončit a informovat uživatele.

### 7.1.2 Vytvoření obrázku

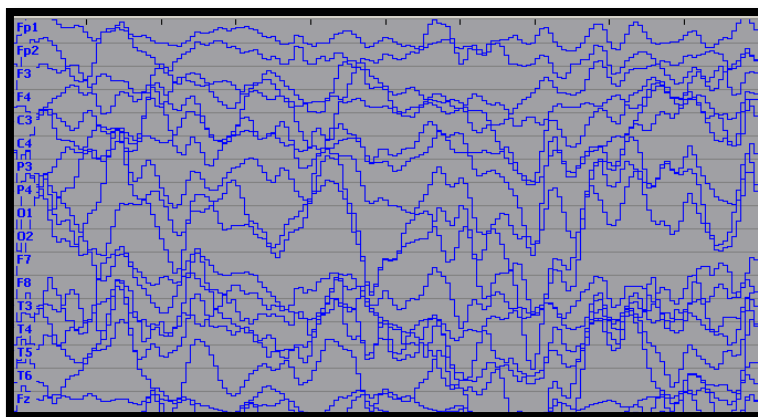
Nyní je nutné převést signálová data z EDF souboru na obrázek, aby jej bylo možno uložit do DICOM souboru. V rámci tohoto kroku je nutné správně zvolit amplitudu signálu a délku epochy. Tyto údaje by měly být zvoleny tak, aby mohl lékař při prohlížení DICOM souboru (obrázku) poznat všechny důležité momenty. Rozdíl mezi příliš velkou a příliš malou amplitudou a epochou ukazují následující obrázky. Obrázek 7.2 ukazuje EEG data s amplitudou 100 (50  $\mu\text{V}/5\text{ mm}$ ) a délkou epochy 10 s. Pokud bude amplituda a epocha větší než v tomto případě, dochází k situaci obrázku 7.3. Bude-li nižší, nastane případ z obrázku 7.4. Z uvedených obrázků má pro lékaře největší výpovědní hodnotu jen obrázek 7.2.



Obrázek 7.2 - Správně zvolená amplituda a velikost epochy (100, resp. 10 s)



Obrázek 7.3 - Vyšší amplituda a délka epochy (200, resp. 20 s)



Obrázek 7.4 - Nižší amplituda a délka epochy (20, resp. 2 s)

Chceme-li, aby došlo ke správnému vykreslení (správnému měřítku) dat, musí uživatel zadat velikost amplitudy a délku epochy. Ačkoli by se to mohlo zdát poněkud neohrabané, je důležité si uvědomit, že lékař tyto hodnoty zná. Je možné je zjistit z pohledu na data v programu PL-Winsor, ve kterém dochází k exportu do EDF. Tímto bude zajištěno, že vygenerovaný obrázek bude mít vždy dostatečnou přesnost.

K samotnému vytvoření obrázku je možné použít jakékoliv obrázkové knihovny implementované v jazyce C++ (např. *libpng*).

### 7.1.3 Vytvoření DICOM souboru

Zde přichází na řadu toolkit DCMTK, konkrétně nástroje z modulu *dcmimgle*. Pomocí těchto prostředků můžeme vzít data extrahovaná z hlavičky EDF souboru (především identifikační údaje pacienta a záznamu) a námi vytvořený obrázek s měřením a „zabalit“ je do DICOM souboru.

## 8 Závěr

---

Smyslem této práce bylo prozkoumat formát EEG dat v prostředí neurologické kliniky FN Plzeň, zjistit přesný způsob jejich uložení a navrhnout řešení, které tato data umožní segmentovat, identifikovat a uložit informačního systému.

Během této práce jsem navrhl několik řešení, jež se po konzultaci s představiteli nemocnice ukázala jako nevyhovující. Bylo tedy nutné přijít s řešením, které nejen řeší daný problém, ale splňuje i omezující podmínky kladené prostředím, v nichž bude provozováno. Mezi tyto podmínky patří zejména výpočetní prostředí laboratoře neurologické kliniky, do něhož není možné přímo zasahovat, či časová vytíženost tohoto pracoviště, jež neumožňuje operativně nasadit a testovat program v reálném prostředí. Proto nebylo možné celou aplikaci realizovat jako samostatný program, který by byl nasazen v prostředí nemocnice.

Práce nabyla směru rozšíření informačního systému, používaného nemocnicí, resp. jeho komponenty PACS. Tímto způsobem můžeme zajistit velmi jednoduché ovládání programu, bez nutnosti přímo zasahovat do programového vybavení laboratoře.

Toto řešení však s sebou přineslo nutnost převést EEG data do takového formátu, který by umožnil jejich zobrazení právě v systému PACS. Navrhl jsem proto konverzní algoritmus, řešící danou problematiku. Nebylo však možné jej v praxi implementovat. Toto vyžaduje navázání spolupráce s firmou MediCalc Software s.r.o., jež stojí za vývojem IS MediCalc. Tento postup by umožnil začlenit navržené rozšíření, včetně konverzní složky, do stávající programové struktury této aplikace.

Alternativou by bylo, aby nemocnice umožnila přístup k datům z EEG v rámci projektů, které jsou vedeny skupinou MISRG, působící v rámci katedry informatiky a výpočetní techniky Západočeské univerzity v Plzni. Tento způsob by umožnil mnohem pružnější vývoj a testování bez nutnosti zasahovat do programového vybavení laboratoří neurologické kliniky.

# Seznam zkratek

---

CT	Computed Tomography
ECG	Elektrokardiogram
EDF	European Data Format
EEG	Elektroencefalografie
EKG	Elektrokardiografie
EMG	Elektromyografie
ICA	Independent Component Analysis
IS	Informační systém
MR	Magnetic Resonance
NS	Number of Signals
PACS	Picture archiving and communication system
PSG	Polysomnografie
SOP	Service Object Pair Class
TFA	Time-Frequency Analysis
XA	X-Ray Angiography



# Použité zdroje

---

- [1] AMBLER, Zdeněk. *Základy neurologie*. šesté, přepracované a doplněné vydání. Praha: Galén, 2006. ISBN 80-7262-433-4.
- [2] SANEI, Saeid a J. A. CBAMBERS. *EEG Signal Processing*. Chichester (Velká Británie): John Wiley & Sons Ltd, 2007. ISBN 978-0-470-02581-9.
- [3] HORNIÁK, Robin. *Analýza EEG signálu*. Praha, 2010. Dostupné z: <http://cyber.felk.cvut.cz/research/theses/papers/92.pdf>. Diplomová práce. České Vysoké Učení Technické, Fakulta elektrotechnická, Katedra kybernetiky. Vedoucí práce Ing. Václav Gerla.
- [4] POLÁK, Radek. *Měření EEG signálu pomocí digitalizační měřicí karty*. Brno, 2009. Dostupné z: [http://www.vutbr.cz/www\\_base/zav\\_prace\\_soubor\\_verejne.php?file\\_id=18803](http://www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=18803). Bakalářská práce. Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav radioelektroniky. Vedoucí práce Ing. Radim Kolář, Ph.D.
- [5] HRDLIČKA, Jiří. *Interaktivní simulátor EEG*. Kladno, 2012. Dostupné z: [http://webzam.fbmi.cvut.cz/hozman/TPR/20112012/Z\\_Hrdlicka.pdf](http://webzam.fbmi.cvut.cz/hozman/TPR/20112012/Z_Hrdlicka.pdf). Týmový projekt. České vysoké učení technické v Praze, Fakulta biomedicínského inženýrství, Katedra biomedicínské techniky. Vedoucí práce Ing. Jan Suchomel.
- [6] European Data Format - Overview. KEMP, Bob. *European Data Format (EDF)* [online]. 2003 [cit. 2012-05-12]. Dostupné z: <http://www.edfplus.info/>
- [7] EEGLAB - Open Source Matlab Toolbox for Electrophysiological Research. EEGLAB [online]. 2004 [cit. 2012-05-12]. Dostupné z: <http://sccn.ucsd.edu/eeglab/>
- [8] KERKENI, Nizar. *JEDF. Freecode* [online]. 2006 [cit. 2012-05-12]. Dostupné z: <http://freecode.com/projects/jedf>
- [9] VAN BEELEN, Teunis. *EDFbrowser: Overview*. EDFbrowser [online]. 2011 [cit. 2012-05-12]. Dostupné z: <http://www.teuniz.net/edfbrowser/>

- [10] KEMP, Bob, Alpo VÄRRI, Agostinho ROSA, NIELSEN a John GADE. European Data Format: Specification. European Data Format (EDF) [online]. 1992 [cit. 2012-05-12]. Dostupné z: <http://www.edfplus.info/specs/edf.html>
- [11] KEMP, Bob a Jesus OLIVAN. Full specification of EDF+. <Http://www.edfplus.info/specs/edfplus.html> [online]. 2003 [cit. 2012-05-12]. Dostupné z: <http://www.edfplus.info/specs/edfplus.html>
- [12] NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION (NEMA). DICOM Brochure. Rosslyn, Virginia, USA, 2011. Dostupné z: <http://medical.nema.org/dicom/geninfo/Brochure.pdf>
- [13] DICOM verze 2011. Digital Imaging and Communications in Medicine (DICOM): Part 1: Introduction and Overview. Rosslyn, Virginia, USA: National Electrical Manufacturers Association, 2011. Dostupné z: [http://medical.nema.org/Dicom/2011/11\\_01pu.pdf](http://medical.nema.org/Dicom/2011/11_01pu.pdf)
- [14] DICOM verze 2011. Digital Imaging and Communications in Medicine (DICOM): Part 10: Media Storage and File Format for Media Interchange. Rosslyn, Virginia, USA: National Electrical Manufacturers Association, 2011. Dostupné z: [http://medical.nema.org/Dicom/2011/11\\_10pu.pdf](http://medical.nema.org/Dicom/2011/11_10pu.pdf)
- [15] VAN BEELEN, Teunis. EDFlib: Overview. EDFlib [online]. 2010, 17. 6. 2011 [cit. 2012-06-16]. Dostupné z: <http://www.teuniz.net/edflib/index.html>
- [16] DCMTK - DICOM Toolkit. DICOM@Offis [online]. 2011 [cit. 2012-05-12]. Dostupné z: <http://dicom.offis.de/dcmtdk.php.en>

## **Přílohy**

---

# Seznam obrázků

---

Obrázek 2.1 - Schéma centrální nerovné soustavy .....	2
Obrázek 2.2 - Grafické znázornění mozkový vln .....	4
Obrázek 2.3 - Rozložení elektrod v „EEG čepičce“ podle systému 10/20 s 21 elektrodami .....	5
Obrázek 2.4 - Elektroencefalograf a jeho komponenty (EEG čepička a výkonná jednotka) .....	5
Obrázek 2.5 - Výřez z EEG záznamu činnosti mozku zdravého pacienta pro 6 kanálů v délce cca 7s .....	6
Obrázek 4.1 - Ukázka grafického prostředí programu EEGLAB .....	13
Obrázek 4.2 - Ukázka grafického prostředí programu jEDF .....	15
Obrázek 4.3 - Ukázka grafického prostředí programu EDFbrowser.....	16
Obrázek 4.4 - Problém s přidáním signály EKG a TIMESYNC při zobrazení dat v programu EEGLAB.....	17
Obrázek 6.1 - Postavení formátu DICOM souborů (tučný rámeček) v rámci celého DICOM standardu verze 2011 .....	29
Obrázek 6.2 - Rámcová struktura DIOCM souboru.....	29
Obrázek 6.3 - Příklad signálových dat.....	34
Obrázek 6.4 - Příklad obrazových dat .....	34
Obrázek 7.1 - Vývojový diagram zachycující funkčnost konverzního algoritmu .....	38
Obrázek 7.2 - Správně zvolená amplituda a velikost epochy (100, resp. 10 s) .....	39
Obrázek 7.3 - Vyšší amplituda a délka epochy (200, resp. 20 s).....	39
Obrázek 7.4 - Nižší amplituda a délka epochy (20, resp. 2 s) .....	40

# Seznam tabulek

---

Tabulka 2.1 – Přehled frekvencí mozkových vln.....	4
Tabulka 6.1 - Přehled datových polí v hlavičce EDF souboru .....	22
Tabulka 6.2 – Struktura samotných dat v EDF souboru.....	23
Tabulka 6.3 - Přehled přípustných metainformací o DICOM souboru .....	32
Tabulka 6.4 - Součásti DCMTK .....	37
Tabulka 6.5 - Knihovny potřebné pro sestavení DCMTK.....	37