

I - 4.13.3 Datenformat

THOMAS PENZEL, Marburg

Zusammenfassung

Computergestützte Polygraphiesysteme erlauben das papierlose Archivieren von Polysomnographien. Ein Datenaustausch - oder auch ein Datenrückgriff nach 10 Jahren (wenn das ursprüngliche Aufnahmesystem nicht mehr existiert) erfordert ein einheitliches und standardisiertes Datenformat. Alle Systeme zeichnen heute in verschiedenen firmenspezifischen Formaten auf. Es gibt zahlreiche Vorschläge für Datenformate von denen zur Zeit keines verpflichtend ist. Eine Kenntnis der gängigen Formate und der gängigen Bezeichnungen ist von großem Vorteil. Die DGSM empfiehlt die Verwendung des "European Data Format (EDF)" zum Austausch digitaler Polysomnographien.

Die computergestützten Polysomnographiesysteme speichern ihre Daten in Dateien auf Festplatte oder anderen Speichermedien. Die Daten können entweder in einer Datei oder in mehreren Dateien (aufgeteilt nach Signalkanälen oder Zeitabschnitten) gespeichert werden. Das Speichern in einer Datei hat den Vorteil, daß immer alle Daten unverändert zusammen vorliegen, und den Nachteil, daß, sofern ein Dateifehler auftritt, häufig alle Daten der Aufzeichnung verloren sind. Das Speichern in Dateien erfolgt entweder binär (sehr platzsparend) oder in der Form von ASCII-Zeichen (leicht lesbar, siehe ASTM weiter unten). Beim Speichern der Daten ist zwischen den reinen Signaldaten (auch Rohdaten genannt) und den Auswertungen und Kommentaren zu unterscheiden. Gerade bei den letzteren wird eine systematische Verschlüsselung sehr wichtig. Zur Verschlüsselung der Diagnosen wird der ICD bzw. der ICSD eingesetzt. Zur Verschlüsselung weiterer Begriffe werden sogenannte medizinische Nomenklaturen benutzt (SNOMED o.a.) nach denen jedem medizinischen Begriff z.B. Biosignale (EKG) oder Ableiteorte (Musculus tibialis anterior = T-14760) je ein Schlüssel zugeordnet werden kann.

Hier werden Datenformate für Biosignale behandelt, d.h. die Art, wie die Biosignale in einer Datei abgelegt werden. Unabhängig davon, ob die Daten

binär oder als ASCII-Zeichen gespeichert werden, und unabhängig davon ob die Daten in einer oder mehreren Dateien gespeichert werden, können drei verschiedene Ordnungsprinzipien unterschieden werden:

1. Signalordnung: jeder Datenwert eines Signals wird für die gesamte Meßzeit hintereinander gespeichert. Dann folgen alle Datenwerte des zweiten Signals. Der Vorteil liegt darin, daß die Abtastraten der Signale voneinander unabhängig sind.
2. Sampleordnung: auf den ersten Datenwert des ersten Signals folgt der erste Datenwert des zweiten Signals und der erste Datenwert des dritten Signals. Der zweite Datenwert des ersten Signals folgt erst, wenn alle Signale abgearbeitet sind. Diese Ordnung entspricht direkt der A/D-Wandlung (s. I-3.2.1) und ist sehr gut für ein Monitoring geeignet. Der Nachteil liegt darin, daß alle Signale die gleiche Abtastrate haben.
3. Blockordnung: für eine bestimmte Epoche, z.B. eine Sekunde werden alle Datenwerte des ersten Signals hintereinander gespeichert. Dann folgen alle Datenwerte des zweiten Signals für die gleiche Epoche, bis alle Signale gespeichert sind. Als nächstes wird die folgende Epoche in der gleichen Art gespeichert. Diese Ordnung stellt einen Kompromiß aus den beiden anderen Formaten dar.

In den Vereinigten Staaten hat eine Kommission mit Firmenvertretern und Vertretern aus neurophysiologischen Labors im Auftrag der allgemeinen Standardisierungskomitees ein Aufzeichnungs- und Dokumentationsformat entwickelt. Es heißt ASTM E 1467 und speichert alle Daten als sogenannte ASCII-Dateien. Die Formatbeschreibung beinhaltet ein sehr umfassendes Verschlüsselungsschema für neurophysiologische Untersuchungen. Ein Teil der Verschlüsselung wurde aus der SNOMED-Nomenklatur übernommen. Die Auswahl der Schemata richtet sich weitgehend nach den amerikanischen Abrechnungsmodalitäten. Bis jetzt erscheint dieser ausführliche Vorschlag zum Kodieren und Speichern von Befunden als sehr gut geeignet, jedoch zum Speichern von polygraphischen Signalen als we-

nig geeignet, da nur ASCII-Zeichen vorgesehen sind und das platzsparende Speichern und Übertragen binärer Daten nicht möglich ist.

HL7, Health Language Level 7 entwickelt sich zu einem Standard zur Archivierung von allen im Krankenhaus anfallenden Daten. Es soll der Standard für die digitale Patientenakte werden. HL7 weist einen hohen Verbreitungsgrad auf. Die Aufnahme von Langzeit-EEG und bei der Polysomnographie anfallenden Datenmengen ist bisher in diesem Standard nicht berücksichtigt. Die Integration von binären EKG-Aufnahmen ist aber prinzipiell vorgesehen. HL7 ist ein Verschlüsselschema und keine direkte Beschreibung für ein Datenformat.

Das Europäische Datenformat (EDF) wurde zum Austausch von Polysomnographien für wissenschaftliche Zwecke entwickelt. Es ist sehr einfach und beinhaltet nur die polygraphisch aufgezeichneten Signale jeweils einer Untersuchung in einer Datei. Die Daten werden binär unter Einsatz einer Blockordnung gespeichert. Viele computergestützte Polygraphiesysteme unterstützen heute dieses Datenformat. Weil es so einfach ist, kann es leicht ohne große Programmierkenntnisse implementiert werden. Es beinhaltet jedoch keinerlei Verschlüsselschemata.

haltet jedoch keinerlei Verschlüsselschemata.

Das Erlanger BiosignalfORMAT (EBS) wurde zur Aufnahme und Auswertung von EEG-Signalen für Epilepsieuntersuchungen entwickelt. Die Messung wird zusammen mit der Auswertung und den klinischen Befunden in einer Datei gespeichert. Es können Signale nach allen drei Ordnungsprinzipien gespeichert werden. Als besonderer Vorteil kann angesehen werden, daß nachträglich weitere Auswertungen angefügt werden können. Dadurch hat das Format eine besonders hohe Flexibilität, ist aber auch relativ kompliziert zu programmieren.

DICOM und MEDICOM sind Datenformate, die für den digitalen Bildaustausch entwickelt wurden und hauptsächlich in der Radiologie Anwendung finden. Heute gibt es Bestrebungen diese Formate zu erweitern, damit auch Biosignale gespeichert werden können. Der Bedarf dazu besteht bei angiologischen Untersuchungen.

Eine Standardisierungskommission der Europäischen Gemeinschaft CEN (Committee European de Normalisation) TC 251 (Medical Informatics) versucht die positiven Elemente der verschiedenen Formate zusammenzuführen und ein neues Format zu

Tabelle 1: Das Europäische Datenformat (EDF)

BLOCK	BEDEUTUNG	DATEN	
allgemeiner Header	Versionsnummer des Formats (0) Patienten Identifikation Aufnahme Identifikation Aufnahme Startdatum (DD.MM.YY) Aufnahme Startzeit (HH.MM.SS) Größe des Headers Kommentare 44 ASCII Anzahl der Datenblöcke Dauer eines Datenblocks (in Sekunden) Anzahl der Biosignale (N)	8 ASCII 80 ASCII 80 ASCII 8 ASCII 8 ASCII 8 ASCII 44 ASCII 8 ASCII 8 ASCII 4 ASCII	041 set 0 8 88 168 176 184 228 236 244 252
Signalheader	N * Kanalbezeichnung (EEG C3-A2) N * Elektrode oder Sensor N * Physikalische Dimension (uVolt) N * Physikalisches Minimum (-500) N * Physikalisches Maximum (500) N * Digitales Minimum (-2048) N * Digitales Maximum (2047) N * Filtereinstellungen N * Anzahl Datenwerte pro Signal <i>Block</i> N * Kommentare	16 ASCII 80 ASCII 8 ASCII 8 ASCII 8 ASCII 8 ASCII 8 ASCII 80 ASCII 8 ASCII 32 ASCII	0 0 16 10 96 60 104 68 112 70 120 78 128 80 136 88 216 88 224 80
Datenwerte	Daten 1. Biosignal, 1. Block Daten 2. Biosignal, 1. Block Daten N. Biosignal, 1. Block Daten 1. Biosignal, 2. Block	binär	

entwickeln. Das neue Format soll eine hohe Kompatibilität mit den bisher existierenden Formaten aufweisen. Es soll weiterhin Verschlüsselungsschemata enthalten und es soll verpflichtend für zukünftige Polygraphiesysteme in Europa werden.

Jedes einzelne Datenformat besitzt einen allgemeinen "HEADER" in dem Information über den Patienten, die Umstände der Aufzeichnung sowie die Kennzeichnung und Ordnung der Daten enthalten sind. Im speziellen Signalheader sind dann Informationen über die einzelnen Biosignale enthalten. Alle Angaben sind linksbündig als ASCII-Zeichenketten gespeichert. Darauf folgen die Datenwerte in der festgelegten Ordnung. Die Datenwerte sind als 16-bit Integer-Zahlen mit Vorzeichen gespeichert. Die Beschreibung ist in *Tab. 1* für das EDF-Format auf-

geführt. Auf den allgemeinen Header folgt der Signalheader mit so vielen Informationen (N), wie unterschiedliche Biosignale gespeichert werden. Dann folgen in Blockordnung die binären Datenwerte.

Literatur

- American Society for Testing and Materials (ASTM), Subcommittee E31.16. Standard Specification for Transferring Digital Neurophysiological Data Between Independent Computer Systems. Philadelphia, 1991.
- KEMP B, VARRI A, ROSA AC, NIELSEN KD, GADE J.: A simple format for exchange of digitized polygraphic recordings. *Electroenceph. Clin Neurophysiol* 82: 391-393, 1992
- KUHN M: Specification of the EBS file format for the processing of biosignals. Institut für Physiologie und Biokybernetik, Universität Erlangen, Nürnberg, 1993