dHealth 2019 – From eHealth to dHealth
D. Hayn et al. (Eds.)

doi:10.3233/978-1-61499-971-3-97

Achieving an Interoperable Data Format for Neurophysiology with DICOM Waveforms

Silvia WINKLER[a,1], Martin HUBER[a] and Tilmann KLUGE[b]
[a]*Sigma Software Solutions OG, Vienna, Austria*
[b]*Austrian Institute of Technology, Vienna, Austria*

Abstract. Modern healthcare faces multiple challenges: diagnosis and treatment happens multidisciplinary and distributed. The key principle to accomplish this is interoperability. Some disciplines like radiology are well experienced in interoperable workflows and cross institution data exchange; other disciplines just realize the growing importance. In this paper we analyze the situation in neurology and give an overview of attempts made in the past to establish an interchangeable, interoperable data format for biomedical signal data, which would be suitable for neurology, too. Focusing on EEG data we will discuss how DICOM Waveforms could be used to cover many of the requirements. As a result necessary adaptions and remaining issues are identified. With DICOM Waveforms a specification is available that covers most of the interoperability requirements. With only little adjustments DICOM Waveforms could establish data interoperability in neurology.

Keywords. health information interoperability, electroencephalography

1. Introduction

Modern healthcare requires the exchange of data. Today medicine acts in an interdisciplinary and integrated way. Data have to be exchanged across departments and even across hospitals or bigger organizations like hospital associations or insurances. In the past great efforts for developing unified medical IT platforms for data exchange and storage have been made. The medical devices are more and more connected to these platforms and medical data is exchanged in a secure and standardized way making it available to other departments which did not record it. The key issue for cross department and moreover cross enterprise data exchange are standardized data formats and communication protocols. IHE is a worldwide initiative by healthcare professionals and industry to improve the interoperability of healthcare systems and to provide integration profiles for a broad range of clinical workflows. Based on medical standards like HL7 and DICOM a working interoperability between healthcare applications can be reached. Today clinical systems for administration or workflow (e.g. HIS, RIS or LIS), report repositories and imaging archives work together using these standards. A substantial example for the benefits of standardized data is the Austrian public health record ELGA, which was designed according IHE

[1] Corresponding Author: Silvia Winkler, Sigma Software Solutions OG, Markhofgasse 1-9/3/338, 1030 Vienna, Austria, E-Mail: silvia.winkler@sigmasoft.at

specifications: HL7 CDA is the standardized document format for medical reports in ELGA and DICOM is the required data format and communication protocol for medical images and image related data like signal data or evidence documents. The decision of the European Commission 2015/1302 of 28 July 2015 [1], which declares 27 IHE profiles as 'eligible in public procurement', emphasizes the legitimacy to require interoperable data formats in future public tenders.

Prior condition for interoperable systems is a standardized data format. In the past different organizations provided normative standards or noncommittal format specifications for neurophysiological time signals like EEG; a comprehensive comparison of many of them was done by A. Schloegl [2]. Until the recent past none of the formats found a broader acceptance by manufacturers of EEG devices. We still see the situation, that manufacturers use proprietary formats – in most of the cases with restricted access to the specifications. Stead et al. [3] recently discussed the urgent need to establish a common format and proposed the use of the MEF3 format.

This paper shows that the existing, well-established DICOM standard is able to achieve interoperability for neurophysiology data like EEG and gives an overview of the necessary extensions to the existing DICOM Waveforms specification.

2. Methods

Starting with a research on existing standards and noncommittal format specifications developed by industry or research institutions, which are suitable for encoding and exchanging EEG signal data, we took a close look on their capabilities, on their metadata, and supported data encoding. We summarized their advantages and disadvantages and, finally, evaluated them in terms of being integrated in today's healthcare eco systems.

In comparison to these we analyzed the capabilities of the already existing DICOM waveform specification, which is available for some different sorts of biomedical signal data but still lacks a specification for neurophysiological recordings. This is remarkable in view of the fact that many of the historical standards were taken into consideration when the DICOM waveform specification was developed in 2001; particularly are to be mentioned

- ASTM E31.16 - E1467
- HL7 V2.3
- CEN TC251 PT5-007 SCP-ECG
- CEN TC251 PT5-021 VITAL
- IEEE P1073 – Medical information BUS (MIB)

Both, ASTM E1467 respectively its successor ACNS TS1 [1] and HL7 V2.x [6], are evaluated later in this document. The two CEN standards mentioned in the DICOM waveform specification existed only as drafts released from the CEN Technical Committee 251 (WG 5 - Medical Device Communication in Integrated Healthcare). SCP-ECG reached broader usage especially in cardiology use cases, today it is part of the CEN/ISO/IEEE 11073-9x specifications. CEN VITAL and IEEE MIB became the core standards in CEN/ISO/IEEE 11073 [7], [8].

In order to achieve interoperability for neurophysiology signal data the evaluation of the different standards included not only technical parameters but focused on their ability to be included in clinical workflows and existing healthcare systems.

3. Results

3.1. ACNS TS1 (ASTM E1467)

The American Clinical Neurophysiology Society released the first version of this standard (ASTM E1467) in 1992 as a result of a joint undertaking of clinical, academic, and vendor interests. As its successor ACNS TS1 [1] was released in 2008. Both versions are based on NCCLS LIS 5-A [5], which was one of the fundamentals of HL7 v2.x, too.

ACNS TS1 claims support not only for EEG data but for a broad range of digital electro-physiologic waveform data in clinical and research environment like electromyogram (EMG), polysomnography (PSG) and evoked potentials (EP) as well.

The standard is strictly ASCII based; in the earlier version even the signal data had to be stored as 7.bit ASCII values. The current version still supports ASCII encoded sample data and recommends them still for short recordings. In addition, sample data can be provided in numeric form as well, but this requires additional data files. Annotations are supported and stored as the waveform data itself in result segments. For annotations the segment category ANA is used.

Besides the waveform data and their acquisition parameters the specification contains in-depth structured data segments, well defined data types, comprehensive lists of allowed values and codes and guidelines for message exchange. The message format is similar to HL7 V2.x messages

ACNS TS1 is a comprehensive standard for medical waveform data, which includes administrative and acquisition context. It is notable for its broad nomenclature, which contains defined terms for almost every single parameter.

The standard defines different levels of implementation (Level I – Waveforms only; Level II- Waveform or Procedure Annotations or Both; Level III – Coded Information) and suggests a “Description of Implementation”, which a system should provide in order to declare details of its compliance.

Although the message format itself is very close to HL7 V2.x, which is worldwide used, there is no known implementation of this standard.

3.2. HL7 V2.6

Health Level Seven (HL7) is an international organization which provides standards for healthcare interoperability.

HL7 V2 [6] is a message based standard with focus on administrative tasks like patient and order management, and communication of results like laboratory measures. It is broadly used all over the world and facilitates communication between hospital information systems and departments like radiology or laboratory. It is used for pharmacy tasks as well as for billing and – last but not least – in electronic health records. It is one of the base standards used in IHE integration profiles.

The standard is message-oriented: triggered by events, messages are exchanged usually in a request-reply manner. Unrequested sending is supported, too. Waveform support was introduced in V2.3 (1997), for this research we looked at the specification of V2.6 (2008).

HL7 V2 messages use 7-bit ASCII encoding as default. An alternative character set can be defined in the message header, even multi-byte character sets or Unicode are allowed. The message itself consists of segments built up in a defined order of required and optional fields separated by a fixed character used as separator.

Waveform data are supported as observation content (OBX) within observation result (ORU) messages, which have a defined structure depending on the use case. The ORU message contains segments with information about the clinical context – i.e. the patient, the order or the result. Signal data are included in observations segments (OBX) with category result type WAV. Only ASCII encoding, having a defined delimiter after each sample value is allowed. Waveform annotations are transmitted in OBX result segments with category ANO. Annotations are coded data associated with a given point in time during the waveform recording. Relationship to the channels is provided via ID.

HL7 is well-known and broadly used in the healthcare domain. It's used for messaging tasks and as such, it plays an important role in transmitting clinical observations as well. It is focused on the message and not on the persisted object – this could be the reason, why it is not used to store medical signal recordings except within databases. There is no defined file format and no possibility to store the data on portable media. None of the known vendors uses this format to store EEG recordings. Maybe there are some implementations, which send relevant signal parts enclosed in the documentation to a hospital information system.

Another reason, why there is no broad use of HL7 waveforms, might be the limitation to ASCII data. Having large sets of binary data being converted to ASCII representation brings a large overhead.

3.3. EDF and EDF+

The European Data Format EDF [9] and its successor EDF+ [10] are an open, non-proprietary guideline for ASCII based encoding of signal data. There is a large set of free available tools to handle this format.

EDF+ is widely compatible with EDF. With the more recent version, discontinuous recordings and annotations are supported. EDF and EDF+ are frequently used for sleep data (PSG), EEG, ECG, and EMG.

A standard EDF file consists of a header record with metadata for the recording followed by data records. The filename extension has to be .edf or .EDF. A data file contains recordings which were acquired with the same technique and with common amplifier settings. Different settings result in different files.

For metadata provided in the file header only a limited character set (ASCII 32-126) is allowed. A space separated CHAR-array contains the patient data. A space separated ASCII-array contains information for the recording. Per-channel properties define the physical conditions of the recording. There is no information about the recording geometry, i.e. the electrode positions.

EDF and EDF+ only support 16-bit sample values in strict chronological order. Values have to be stored as 2-complement, in little-endian byte order (low byte first). Recorded data are split into chunks with a maximum size of 61440 bytes.

The necessity to store sample values in 16-bit values leads to size overhead, if the acquisition only produces 8-bit values. On the other hand sample values with more than 16 bits are not supported and can only be stored after downsizing.

EDF does not support annotations. EDF+ supports text annotations, events and stimuli. They are stored together with the signal data as an additional channel with a defined channel label. The sample data for the channel then contains characters instead of 2 byte integers.

EDF and EDF+ are without doubt the most popular formats for biomedical signal data. They are supported by a wide range of EEG manufacturers and software vendors. In many cases interfaces for importing or exporting EDF are available.

3.4. GDF and OENORM K 2204

The General Data Format GDF [11] was defined with the aim to overcome some of the limitations of EDF+, for example missing support for more than 16 bit sampling width or missing electrode positions. For this purpose some elements of some other standards were incorporated in the specification. An open source implementation (C++ and MATLAB) including tools for conversion and a library for reading and writing GDF 2.x is available at [12]. In 2015 GDF became an Austrian Standard [13].

GDF provides a comprehensive set of metadata in fixed file header containing patient and recording related information and a variable header reserved for channel specific parameters.

Signal data is stored in the data section of the .gdf file. Data is organized in records, each containing a defined amount of samples, first block for first channel, second block for second channel, and so on. The sampling rate, number of samples and data type may differ for each channel. The signal values are stored as numeric data. The data type for each channel is defined in the variable header, the recommended data format is 32-bit integer, but there are 13 different data types defined ranging from 8-bit integer up to 128-bit float.

Annotations are stored in a table of events after the data section. Its start address within the .gdf file is calculable. Events have to follow a well-defined structure containing type, position in samples, channel, duration, and time stamps. For different types of events a code table is provided within the specification.

The General Data Format GDF overcomes some limitations of EDF+ like the missing electrode localization, which is done via coordinate positions (XYZ) with originates in the center of the head. GDF supports any kind of signal data and is not restricted to EEG. The software is used in some research projects (listed at [12]), there is no known commercial use.

3.5. DICOM Waveforms

DICOM [14] is a standard for medical imaging. It is based on a well-defined information model and contains normative specifications for the communication protocol and the data objects as well. Clinical workflow, image transmission, printing, viewing, etc. are covered in this comprehensive standard.

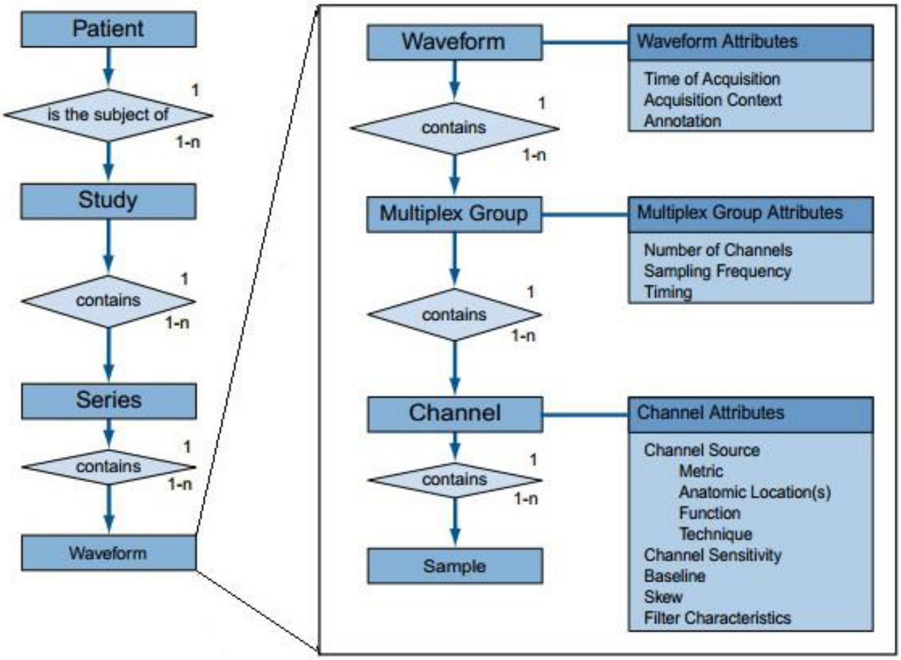

Figure 1. DICOM Waveform Information Model (from [14] PS3.17)

Support for different character sets is granted, the used character set is contained in the DICOM message itself. Unicode and multi-byte character sets are supported, too. Usage of a defined nomenclature is mandatory for many of the properties of the acquisition system; use of different code systems is supported.

DICOM contains Waveform objects since 2001. Object definitions exist for audio data, different types of ECG data, hemodynamic and respiratory signal data as well. Waveform acquisition can happen in context of an image acquisition or without. DICOM allows handling both situations: waveforms can be stored together with an imaging context or on their own, as separate information objects.

DICOM Waveform objects are structured like the well-known DICOM image objects following a well-defined information model – see **Figure 1**. The metadata provide the full clinical context ranging from patient data to data acquisition parameters.

The signal data itself can be stored in different formats, depending on the signal type and on the physical parameters of data acquisition (i.e. the bit-depth of the AD-converter). Defined terms for waveform sample types are listed in **Table 1**.

DICOM waveforms also support annotations. They are stored together with the waveform in one information object. The annotation information can be free text or a coded item. In case of a coded item this can contain a numeric measurement or a coded concept.

Furthermore the well-known and broadly supported DICOM Structured Report objects could be used to store evaluation results and neurology reports.

Table 1. DICOM Waveform Sample Formats (according [14])

Waveform Bits Allocated	Waveform Sample Interpretation	Meaning
8	SB	signed 8 Bit linear
	UB	unsigned 8 Bit linear
	MB	8 Bit µ-law
	AB	8 Bit A-law
16	SS	signed 16 Bit linear
	US	unsigned 16 Bit linear

4. Discussion

In spite of various attempts to find a common data format neurophysiology still lacks interoperability. Although EDF is supported by many device manufactures it is not supported by any healthcare platforms. Extending DICOM Waveforms to neurophysiology data would accomplish this and would result in additional advantages.

DICOM's main objective is interoperability. The standard is – besides HL7 - one of the major components of the IHE integration framework in the radiology and also in the cardiology domain.

To add support for a General EEG Waveform Storage SOP Class analogously to the already existing General ECG Waveform Storage SOP Class only some domain specific adaptions would be necessary. Most of them are easy to achieve, like deviating value ranges for recording properties such as sampling frequency or scaling factors. The main effort results from identifying adequate nomenclatures for electrode positions, coded acquisition context information and coded annotations.

For EEG recordings the anatomical position of the electrodes is an important information. Clinical routine EEG uses electrode positions on the surface of the skull according to the International 10-20 or 10-10 system [16]. ISO/IEEE 11073 – 10101 [15] provides standardized terms for these locations.

To achieve interoperability, DICOM supports annotations with coded content. Therefore ISO/IEEE 11073 – 10101 [15] provides a nomenclature and codes for neurology comprising measurement, device, and patient related events.

Moreover DICOM provides mechanisms to store and to preserve the spatial or temporal relationship of DICOM instances. Synchronization of DICOM objects is possible even if acquired on different devices. For clinical use this could be of interest especially for EEG synchronized to video, fMRI, PET or SPECT.

Especially video recordings are important in neurology, for example in epilepsy monitoring or sleep studies. As DICOM supports different video formats like MPEG2, H.264 or H.265, the videos can be stored as (separate) DICOM files. These DICOM instances can be synchronized to the waveform object – i.e. the EEG - using the mechanisms described above.

Compression remains an open issue because the DICOM standard currently does not include any time-series-specific compression algorithms. Integration of compression algorithms is possible in principle and can be done if the DICOM committee accepts their usage. Algorithms working for different types of signal data (ECG, EEG, MEG, pressure waveforms, etc.) would be preferred; patent protection could be an obstacle. As long as there are no waveform compression algorithms supported by the DICOM standard, waveforms can make use of the deflated transfer syntax (RFC 1951; i.e. the zip algorithm) which is applied to the data set as a whole.

Another open issue is the missing support of (almost) real-time online submission of waveforms. DICOM waveform objects are designed and intended to be persisted. The format does not permit continuous writing due to the structure of the data. Even though the DICOM standard defines a communication protocol (called Transfer Syntax) to stream image data, which is based on the JPEG 2000 image format, there is no such mechanism defined for waveform objects. The need for such communication mechanism is well-known. Efforts will have to be taken in future by the standardization bodies to define streaming protocols for DICOM waveforms for future implementations.

References

[1] Commission Decision (EU) 2015/1302 of 28 July 2015 on the identification of 'Integrating the Healthcare Enterprise' profiles for referencing in public procurement (Text with EEA relevance), https://eur-lex.europa.eu/eli/dec/2015/1302/oj, last access: 21.3.2019

[2] A. Schloegl, An overview on data formats for biomedical signals. *Image Processing, Biosignal Processing, Modelling and Simualtion, Biomechanics. Munich, Germany*: Word Congress on Medical Physics and Biomedical Engineering, 2009; pp 1557-1560

[3] M. Stead, J.J. Halford, Proposal for a Standard Format for Neurophysiology Data Recording and Exchange. *Clin. Neurophysiol.* **33**(5) (2016), 403-413

[4] ASTM E1467: American Clinical Neurophysiology Society Technical Standard 1 (ACNS 1) *Standard for Transferring Digital Neurophysiological Data Between Independent Computer Systems*, 2008

[5] NCCLS. *Standard Specification for Transferring Clinical Observations Between Independent Computer Systems*. NCCLS document LIS5-A [ISBN 1-56238-493-7]. NCCLS, 940 West Valley Road, Suite 1400, Wayne, Pennsylvania 19087-1898 USA, 2003

[6] HL7 Messaging Standard Version 2.6, 2007

[7] ISO/IEEE 11073 – 10201 *Health informatics Point-of-care medical device communication* Part 10201:2004 Domain information model

[8] ISO/IEEE 11073 – 30300 *Health informatics – Point-of-care medical device communication* Part 30200:2004 Transport profile – Cable connected

[9] B. Kemp, A. Värri, A.C. Rosa, K.D. Nielsen and J. Gade, A simple format for exchange of digitized polygraphic recordings. *Electroencephalogr. Clin. Neurophysiol.* **82**(5) (1992), 391-393. http://www.edfplus.info/specs/edf.html, last access: 8.2.2019

[10] B. Kemp, J. Olivan, European data format 'plus' (EDF+), an EDF alike standard format for the exchange of physiological data. *Clin. Neurophysiol.* **114**(9) (2003), 1755-61. http://www.edfplus.info/specs/edfplus.html, last access: 8.2.2019

[11] A. Schloegl, GDF - A General Data Format for Biosignals, https://arxiv.org/abs/cs/0608052, last access: 8.2.2019

[12] A. Schloegl, The BioSig Project, http://biosig.sourceforge.net/index.html, last access: 8.2.2019

[13] Austrian Standards Institute ÖNORM K 2204 *General data format for biomedical signals*, 2015

[14] NEMA PS3 / ISO 12052, *Digital Imaging and Communications in Medicine (DICOM®) Standard*, National Electrical Manufacturers Association, Rosslyn, VA, USA

[15] ISO/IEEE 11073 – 10101 *Health informatics – Point-of-care medical device communication* Part 10101:2004 Nomenclature

[16] Jasper, H.H., The ten–twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol*. **10** (1958), 371–375

[17] Klem, G.H., Luders, H.O., Jasper, H.H., Elger, C., The ten–twenty electrode system of the International Federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr. Clin. Neurophysiol.* **Suppl.**(52) (1999), 3–6