Journal of Electrocardiology xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Electrocardiology

journal homepage: www.jecgonline.com

Archiving and exchange of digital ECGs: A review of existing data formats

Fabio Badilini [a,*], Brian Young [b], Barry Brown [c], Martino Vaglio [a]

[a] AMPS LLC, New York, NY, USA
[b] GE Healthcare Inc., Wauwatosa, WI, USA
[c] Welch Allyin Inc., Milwaukee, WI, USA

ARTICLE INFO

Available online xxxx

ABSTRACT

Digital ECG is today a common practice but a universal format for its storage and exchange has never been widely implemented. The reason is linked on one side to the need of the manufacturing industry to (rightly) protect intellectual propriety and technology, but on the other to an inadequate effort of the research community to sufficiently enforce the use of digital ECG data. To some degree, and at least from a practical point of view, the problem is also linked to other factors, such as the need in some instances to protect patient-sensitive information, and whether digital exchanged data should also include annotations and measurements from an algorithm or by human intervention.

As a result, after more than 30 years it is still common that the full ECG acquired information is not preserved, but only partially stored or saved as a PDF report. Paradoxically, the modern era of hospital information technology and the advent and large diffusion of electronic health record systems did not bring expected improvements: the process of digital ECG retrieval and management remains extremely complicated and cumbersome. The ultimate risk is that the ECG may end up being considered "just" an image rather than a voltage-versus-time signal as it has always been.

A critical review of the most commonly used formats for digital ECG will be given, focusing in particular to those linked with DICOM, HL7 and SCP-ECG standards, and highlighting the respective advantages and limitations with special emphasis to the needs typically encountered by the research community. The goal is to provide a snapshot of the present, and to discuss mid- and long-term potential directions and changes, emphasizing what digital ECG organizations could do to "save" ECG information and facilitate its widespread exchange.

Millions of new ECGs exams are taken every day, and in most cases, they need to be adequately archived. However, and despite modern technology which allows the ECG signal to be digitally *acquired*, the same records are most frequently not digitally *stored*, meaning that waveform source data is either lost, or only partially saved.

The reason for this apparent paradox can be conducted to a number of combined factors, the simplest of all being that a printout report, which can be produced by any ECG machine, is very easy to obtain and to be stored, either as a paper record or as a scanned image file. Even within modern hospitals, where an ECG management system empowering digital ECGs is often in place, the storage of ECG records is achieved sending a PDF report to the electronic health record, which is considered "good enough". Ultimately, whenever ECG data needs to be retrieved (such as for serial comparison or for retrospective analysis) the available data is most frequently a paper printout, or in the best scenario, a PDF report.

Is this really a problem? Or perhaps an image copy of an ECG report could be considered as a proper way of ECG archiving? Most ECG experts (and the authors of this editorial would agree) would probably have no doubts to state that a digital ECG should first of all preserve the "voltage-versus-time" nature of the signal that was described more than 100 years ago by Dr. Einthoven. But other noticeable researchers from Computer Science disciplines involved with recent innovations in the area of machine learning and artificial intelligence may argue and answer differently. Indeed, the number of publications that consider the ECG as an image just as any other medical imaging record is rapidly growing (an interesting and very recent example is given in reference [1]), and the risk that images, rather than the voltage-versus-time signals will prevail.

The problem should perhaps be rephrased, and a precise definition of a digital EGG, accurately addressing which information should be captured for adequate archiving, should be first given. If the goal is to preserve the original signal nature of an ECG, the stored information should always allow reconstruction of the voltage-versus-time waveform signal, with no loss of information, i.e. exactly as it was acquired from the body surface of the subject (the ECG leads). By this definition, most of the image-based archiving used today would be excluded, as the signal typically drawn on standard ECG printouts is 1) only a portion of the signal originally recorded (e.g. only a sub time-segment for most

* Corresponding author at: 418 Central Park West, apt 21, New York, NY, USA.
E-mail address: badilini@amps-llc.com (F. Badilini).

https://doi.org/10.1016/j.jelectrocard.2018.07.028
0022-0736/

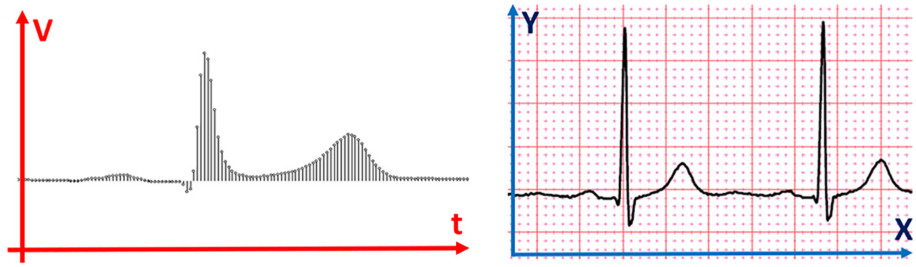

Fig. 1. Difference between a digital ECG (left panel) and an image-scanned ECG (right panel) for a single lead. For a digital ECG, the stored information is the numerical value of discrete samples over time (voltage versus time); the grid is just an optional feature that can be added when drawing the time sequence. For an image-scanned ECG the stored information is the color level of pixels within a bi-dimensional (Y vs X) space; the grid (when available) is intrinsically embedded in the picture (and so does the color and depth of the waveform).

of the leads, and 2) a lossy representation of the source data, both in terms of sampling rate and amplitude resolution. A simple example of a "voltage-versus-time" versus an image-based ECG is given in Fig. 1.

Paradoxically, there are actually multiple options to capture and store a digital ECG, either using one of many ECG formats well described in the literature or adopting one of the available standards compatible with the transmission and storage of the ECG signal [2,3].

An ECG format can be seen as an agreed structured way to store ECG data. In most cases, an ECG format is good enough for ECG data archiving and exchange, and two "popular" examples are the MIT and the ISHNE formats, which were both introduced for the need to share and make available research data [4,5] and which are in widespread use for the certification of software-based medical devices. These file formats have very simple structures and are fairly easy to implement. There are several other formats that allow storage of ECG waveforms, and for sure it is worth mentioning the European Data Format (EDF), which allows storage of multichannel biological and physical signals [6] and the company-specific formats (most of them based on XML) used by the manufacturing industry to export their proprietary data.

Data standards are more complex, and typically involve larger organizations, which work together with specific goals. Standards have a structured governance, include support and technology, and are often formally recognized by regulatory or government agencies. Lastly, whenever applicable, they embed usage of coding systems (vocabularies), for example to lead names and for annotations (see later).

There are many standards, but those that also include the storage of the ECG are DICOM, HL7 (mainly used in USA), and SCP-ECG (mainly used in Europe).

The DICOM standard (Digital Imaging and Communication in Medicine) was introduced in 1983 by the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA) with the primary goal to favor transmission, storage and retrieval of medical imaging information. Basically, it was a standard to allow interoperability of medical images throughout the Hospital meant to be implemented by the PAC systems. The possibility to extend the standard beyond images (including the ECG) came with the so-called supplement 30, which was released in 2000. Today most ECG manufactures are supporting DICOM as an essential feature of their interoperability offering spectrum.

Health Level Seven (commonly known as HL7) is another large organization (represented today by members from 50 different countries) which was founded in 1987 to define a set of standards, formats and definitions for exchanging and developing electronic health records. The ECG formal support by HL7 is also known and referred to as the aECG (annotated ECG). Despite some small overlaps (both standards can encode a digital ECG) the combined usage of DICOM and HL7 is today standard routine in any hospital using an electronic heath record platform.

SCP-ECG is the standard communication protocol for computer assisted ECGs, and it has been developed with specific aim at the ECG signal. It was first released in 1993 (mainly for resting 10-second ECGs) and it is widely used in Europe. The final draft of a revised version that will include the storage and communication of continuously-acquired ECG information (such as Holter or stress test data) had been recently submitted to European Committing of Standardization (CEN) for formal voting [7].

All ECG formats and standards can be used to store ECG source waveform data. In some case (and specifically for research purposes) one of the "simple" formats can be sufficient. However, for more complex scenarios adopting one of the existing standards could be preferable as it would allow it to benefit from all the advantages provided by the standard infrastructure, such as the usage of coding systems and the possibility to apply shared technology to validate and view the data.

From a technical point of view, the structure of the file encoding an ECG can be very different from one format to another and will require adequate software to be interpreted. Most formats the file structure is binary but can be structured (e.g. DICOM and SCP), whereas in another case the file is text-based (as with HL7, which uses XML-structured organization). In addition, the ECG data samples can be optionally compressed (SCP) or encoded (HL7).

Proper archiving of ECG records also involves components that go beyond the source data ECG waveforms. For example, most ECG machines mathematically compute a representative cardiac complex for each of the acquired leads. These special waveforms, frequently referred to as "median" beats (based on one of the methods that can be used to compute them), are extremely important, as they are used by commercial algorithms to determine the parameters that lead to ECG diagnosis. To the best of our knowledge, a comparison between representative complexes from different algorithms has never been done, but it is reasonable to expect potential discrepancies in diagnostic results, particularly in borderline examples. Because of this, a digitally stored ECGs should provide the option to include the representative beats data.

Of note, storage of ECG source data and of representative beats and have nothing to do with the way a specific software tool may actually display them (on a computer screen, or to a printer). These are rendering details that should not be confused with the data that is actually stored. For example, in one specific modality, one would prefer to display the representative beats in waterfall (superimposed) display modality, or in a specific layout organization (e.g. 3 × 4 configuration).

Differences in the acquisition front ends of electrocardiographs may also affect the resulting digital data. This is the case for example of the filtering applied to the source data ECG. While this is not specifically a concern of the data standard used to store the digital data, some formats (like HL7) would capture these differences by allowing the storage of both the raw and filtered data within the same document file. Another important component of a digitally stored ECG data relates to measurements or annotations (which, for the context of this review, can be considered synonymous). In general, an ECG annotation should be seen as

Please cite this article as: Badilini F, et al, Archiving and exchange of digital ECGs: A review of existing data formats, Journal of Electrocardiology (2018), https://doi.org/10.1016/j.jelectrocard.2018.07.028

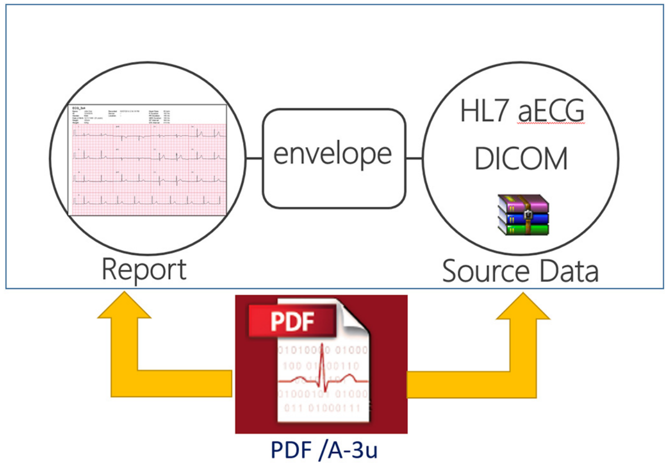

Fig. 2. PDF-ECG: a hybrid format that combines the need to save an ECG report and the source waveform data information. A PDF-ECG object (file) embeds within the same object both a graphical image (report) and the source data used to generate it, which can be in any of the existing formats/standards such as HL7 or DICOM.

an observation related to ECG (either made on source data or on representative beats) that could include one or more specification elements:

- a value (which could be numerical, like 400, or simple text, like "atrial fibrillation")
- a lead (like lead II, or all recorded leads)
- a region of interest (like the onset/offset of an interval, or the isoelectric level and sample position where an amplitude measurement could be done)

An example of an annotation that requires full specification is the QT interval, for which a numerical value (e.g. 400), a lead where it has been measured (e.g. lead II), and the region of the ECG lead where it has measured (e.g. onset at 1.124 s from start and offset at 1.524 from start) are needed. Conversely, more complex annotations such as indexes of morphology of the repolarization may only require the storage of a value, as a sort of a footprint of the ECG, without the need to specify additional information. Finally, the digital ECG container should also allow the inclusion of interpretative diagnostic statements, which could be seen as special cases of annotations or treated as a separate section.

The choice of the correct ECG format or standard is particularly sensitive when it goes down to annotations. Depending on the type of annotations (such as, for example, beat-type versus rhythm-type annotations) some formats may be better than others. Conversely, standards always provide comprehensive solutions to address all types of annotations, and with the possibility to store multiple annotations sets, from different readers or even mixing human and machine readings.

Patient demographics are technically simple to store, but need to be considered sensitive data, and proper de-identification (and in some country data encryption) may be required. Generally, subject first and last name, and the date of birth, should be avoided. Which type of demographic information should be archived depends on the purpose of data storage (what exactly do we want to exchange, and why) and is beyond the scope of this review.

In summary, the lack of adequate storage of digital ECG cannot be adducted to the lack of formats or standards. Evidently, the practical advantage to use an image-bases ECG report still overwhelms the impact and the costs associated with a more structured implementation of digital ECG archiving. And the industry, both the electronic health record providers and the ECG manufactures on the other, should be given incentive and motivation.

A noticeable effort worth mentioning is PDF-ECG, a public-domain concept that using existing technology and without the need to reinvent another ECG format solves the issue of having on one side a simple ECG visualization while assuring preservation of ECG waveforms and annotations. Briefly, PDF-ECG is a hybrid archival format that using PDF/A-3u (a PDF standard which assures long-term data preservation) brings together the advantages of a PDF report (done at whichever resolution) while allowing to embed in the same object (data file) the source data that was used to generate the image report (see Fig. 2 for more details). PDF-ECG has been previously described [8] and a proof of concept was carried out within the electronic health record of an Italian hospital [9]. The core of the experiment was to by-pass the default workflow and replace standard PDF reports with PDF-ECG that were seamlessly accepted by the health record, thus demonstrating the possibility to achieve full archiving with minimal impact.

In conclusion, the diffused spread of digital ECG archiving is achievable and should be somehow enforced, potentially involving research organizations that are sensitive to the subject. For sure, we do not need to invent new ECG formats, we should rather use better what is already available.

References

[1] Labadi RD, Munoz E, Piuri V, Sassi R, Scotti F. Deep-ECG: convolutional neural networks for ECG biometric recognition. Pattern Recogn Lett 2018. https://doi.org/10.1016/j.patrec.2018.03.028 [in press].
[2] Bond RR, Finlay DD, Nugent CD, Moore G. A review of ECG storage formats. Int J Med Inform 2011;80:681–97.
[3] Trigo JD, Alesanco A, Martinex I, Garcia J. A review on digital ECG formats and the relationship between them. IEEE Trans Inf Technol Biomed 2012;16(3):432–44.
[4] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation 2000;101:e215–20.
[5] Badilini F. The ISHNE Holter Standard Output File Format. , 3(3)A.N.E.; 1998; 263–6.
[6] Kemp B, Värri A, Rosa AC, Nielsen KD, Gade J. A simple format for exchange of digitized polygraphic recordings. Clin Neurophysiol 1992;82(5):391–3.
[7] Rubel P, Pani D, Schloegl A, Fayn J, Badilini F, Macfarlane PW, et al. SCP-ECG V3.0: an enhanced standard communication protocol for computer assisted electrocardiography. Comput Cardiol 2016;43:309–12.
[8] Sassi R, Sparagino L, Stockbrige NL, Guadiana JM, Badilini F. Proof of concept for an international long-time preservation ECG format. Comput Cardiol 2014;41:461–4.
[9] Sassi R, Bond RR, Cairns A, Finlay DD, Guldering D, Libretti G, et al. PDF-ECG in clinical pracrice: a model for long-term preservation of digital 12-lead ECG data. J Electrocardiol 2017;50:776–80.

Please cite this article as: Badilini F, et al, Archiving and exchange of digital ECGs: A review of existing data formats, Journal of Electrocardiology (2018), https://doi.org/10.1016/j.jelectrocard.2018.07.028