FDA XML Data Format Requirements Specification

Barry Brown, Product Integration Manager, Mortara Instrument
Mark Kohls, Engineering Director, GE Medical Systems-Information Technologies
Norman Stockbridge, M.D., Ph. D., Medical Team Leader, Food and Drug Administration
Others Welcome….
Send us your information….

1 Introduction

1.1 Purpose

This specification document exists for the purpose of defining the specific requirements of the FDA XML Data Format (FDADF) and ensuring that all interested individuals and organizations involved with the project have the same understanding of the data format requirements. This document is not a policy document on methodologies to perform clinical trials or how to measure ECG's or data in any other modality.

1.2 Scope of Specification

This document covers the requirements for the data format as well as relevant submission information. Areas addressed by this document include identifying requirements for the waveform and fiducial mark representations, as well as other information, e.g. measurements, that may be required. Applicability and interaction with to other standards bodies and data definitions, e.g. CDISC [1] and HL-7 [2], as well as current practice with SAS submission data sets is also discussed. Requirements for the structure of the data on electronic media are also covered.

2 Overview

2.1 Background

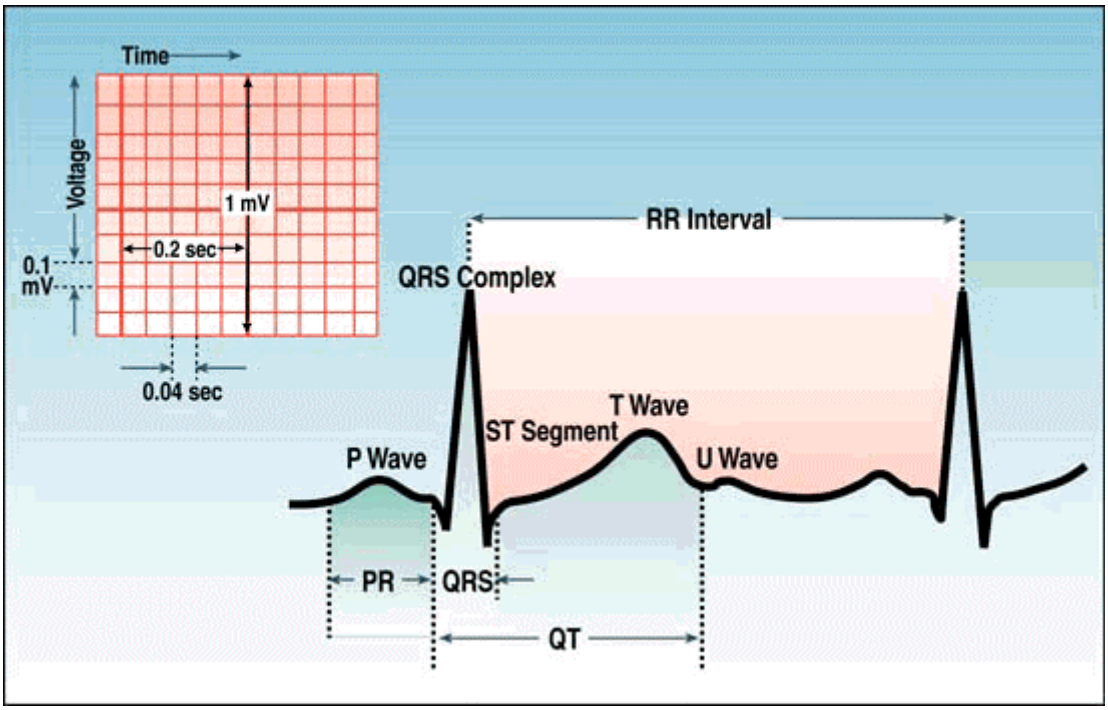

Figure 1: ECG Intervals

There is great interest by the FDA ("the Agency") in receiving digital time-series data, particularly ECG data, that can be reviewed electronically in the course of evaluating drug approval applications [3,4].

In order to facilitate this process it is desirable to establish a single standard for representing ECG and other data. Such a standard would facilitate the creation of a viewer and/or other tools for examining and manipulating the data. Of particular current interest to the Agency is the analysis of QT interval in the ECG. Figure 1 demonstrates the typical interval measurements of an ECG.

2.2 Format

2.2.1 XML

The FDA desires to use XML as the underlying technology for the specification. This matches the Agency's strategic direction for data submissions. It is also aligned with other industry initiatives such as the Operational Data Model (ODM) defined by CDISC [1].

2.2.2 Interfaces to Existing Data Sets

The proposed standard does not create a standalone data set. Additional context about the recordings will be furnished in files that are external to the XML data files. The data will be linked by the creation of a unique identifier that will unambiguously identify a recording in a study.

2.3 Data Submissions for Existing and Future Studies

As previously stated, this document is not a policy document on methodologies to perform clinical trials or how to measure ECG's or data in any other modality. Obviously, studies have been proceeding prior to the establishment of this standard or the November 19th , 2001 meeting [4], and data will have been accumulated for these studies that cannot be represented in the new format. Sponsors should contact their FDA reviewer to determine the proper course of action for existing data and studies.

Three classes of data have been identified as part of the requirements review. They are:

1) *Paper Recordings*, that is, recordings that were recorded on paper for which no digital record exists. Analysis and measurement of the data were conducted offline (or online) in an environment such that it is not possible to export the recording or annotation data in the new format. These paper recordings could be electronically scanned and represented an image. This format is not intended to support an imaged recordings. These existing paper records would not be represented in this new format.

2) *Digital Recording's Without Electronic Annotations*, that is, recordings stored digitally, but the analysis and measurements were conducted offline (or online) in an environment such that it is not possible to export the annotation data in the new format. In this case, it may be possible to send the original recording data in the new format.

3) *Digital Recordings, Electronically Annotated*. These represent the best case and future scenario. The recordings are stored digitally. Further, annotation data and analysis information are captured electronically online and exported to the Agency in the new format.

2.4 Future Considerations

2.4.1 Viewer Requirements

This document focuses on the data set, however, the data must allow support for the following viewer functions:

1) Onscreen display the submitted waveform data

2) Support onscreen calipers in both X and Y axis

3) Display the annotation information both at the individual waveform level and at the group level.

2.4.2 Ownership

The FDA is interested in turning over the long term development and maintenance of this standard to a standards organization such as CDISC.

3 Design Requirements

3.1 Data Format

Requirements

Functional Requirement:

A data representation format should be defined in XML that allows time-series and other data to be transferred to the FDA for review. It should have the following characteristics:

Performance:

1. The specification should support submission of continuous, evenly sampled, multi-channel data. Further, the base definition ought to make it no more difficult to submit ECG or electroencephalogram (EEG) data or time-series data from an implanted glucometer.
2. The specification should support general two dimension data sets that may or may not be evenly sampled. This would support submission of data for frequency–domain analysis data or support submission of vector-cardiography plots.
3. Specializations for a standard 12-lead ECG or Ambulatory ECG, e.g. common channel names, annotation types, units, should exist at a higher level of abstraction. This higher level might be external documentation. For example, if a specification for "12-lead resting ECG" is defined, it must be derived from the general "Time Series" definition, but constrained in the names for channels, annotation types, and units.
4. The standard format should explicitly capture the data structure of a recording session, epochs comprising data acquired over the same interval, and representations of the data obtained during an epoch.
5. There must be a mechanism to annotate points in time and intervals in time corresponding to single channels, representations, and, perhaps, sessions.
6. If data compression is desirable, there should be one data compression algorithm. That algorithm must be in the public domain and it must produce loss-less compression. Data compression should be at the file level, i.e. not integral to a particular XML data element or be performed across multiple files {See also section 3.3.3 below}.
7. Element and attribute naming and usage should correspond to good practices. Recommendations of The *Federal XML Developers Guide* [5] should be followed where applicable.
8. The standard should support a version identifier.
9. A single binary-to-ASCII encoding mechanism should be defined for representation of the sample data.
10. The waveform data should be stored in simplex format, i.e. data sets should not contain interleaved channels on a sample-by-sample basis.
11. A single unique identifier should be created for each recording. This will enable the creation of an unambiguous link between each recording and other data submitted with the study. The identifier only needs to be unique to a study, not across all possible studies.

Design Verification

See *FDA XML Data Format Design Requirements Specification* Section 5 [9].

3.2 Compliance with Existing Standards

Requirements

Functional Requirement:

The format shall reuse concepts, nomenclature, data structures, etc., where appropriate from the following existing standards. The design should document where and how these standards were applied or why they were not applicable.

Performance:

1. HL-7 v3.0 [2]. (HL-7 v3.0 has not defined a waveform standard at this time.)
2. CDISC [1]: This standard will map tags to CDISC core variables where appropriate. Also, SAS Data Submission Sets will have much of the contextual data for the recording.
3. DICOM [7]: Has a waveform standard, but DICOM is not XML based. May have useful concepts to borrow in the area(s) of domain specializations, waveform encoding details, etc.
4. International Conference on Harmonization (ICH) [8]: May have applicable concepts to incorporate regarding common terminology or other relevant areas.

Design Verification

See *FDA XML Data Format Design Requirements Specification* Section 5 [9].

3.3 Submission Format and Organization

Requirements

Functional Requirement:

Standards should be defined for the organization of data in the study.

Performance:

1. The organization of data into a logical unit should be specified e.g. one ECG (EPOCH or Session) per file.
2. A directory structure should be defined.
3. Data compression will be on a per file basis. Data compression will not be applied across directory structures or multiple files.
4. The acceptable forms of media should be defined. The FDA has existing standards defined see [6].

Design Verification

See *FDA XML Data Format Design Requirements Specification* Section 5 [9].

4 Glossary

Electrocardiogram (ECG, EKG):

Electroencephalogram (EEG):

Fiducial Point:

Interval:

5 References

[1] CDISC Information, http://www.cdisc.org

[2] HL-7 Information, http://www.hl7.org

[3] Stockbridge, N., *FDA Proposed ECG Interchange Standards*, November 19th, 2001, http://www.fda.gov/cder/regulatory/ersr/ECGdata.htm

[4] Lipicky, R., Stockbridge, N., et. al., *Proposed Presentations for the Meeting on 19 November 2001*, http://www.fda.gov/cder/regulatory/ersr/ECGpresentations.htm

[5] U.S. Federal CIO Council, Enterprise Interoperability and Emerging Information Technology Committee, XML Working Group, *Federal XML Developer's Guide*, December 2001, Draft Specification.

[6] FDA, *Guidance for Industry: Providing Regulatory Submissions in Electronic Format- General Considerations*, http://www.fda.gov/cder/guidance/2867fnl.pdf, See also additional comments at: http://www.fda.gov/cder/regulatory/ersr/#ECG

[7] DICOM Information, http://medical.nema.org

[8] International Conference on Harmonization, http://www.ifpma.org/ich1.html

[9] Brown, B., Kohls, M., Stockbridge, N., et. al., *FDA XML Data Format Design Specification* Rev TBD.