

# The Unified File Format for HL7 Electronic Documents – DUFF

Mario E. Cortés  
Fundación HL7 Colombia  
Cali, Colombia  
[mario.cortes@datasalud.net](mailto:mario.cortes@datasalud.net)

Diego M. López, PhD  
Telematics Department  
University of Cauca  
Popayán, Colombia  
[dmlopez@unicauca.edu.co](mailto:dmlopez@unicauca.edu.co)

Bernd Blobel, PhD  
eHealth Competence Center  
Regensburg, Germany  
[bernd.blobel@klinik.uni-regensburg.de](mailto:bernd.blobel@klinik.uni-regensburg.de)

**Abstract**— Interoperability through document-based meaningful exchange of Personal Health Records and Electronic Health Records has to be supported by portable document formats. For HL7 structured documents, XML offers an encoding syntax which guarantees interoperability. However, portability is not guaranteed when other files and related documents have to be attached to the structured documents, as it might be necessary to know in advance (or modify) the location (path) of the referenced objects. The objective of the paper is to propose a standard file format for electronic health documents which improves their portability, human readability, and usability. The proposed DUFF HL7 specification complements the HL7 v3 structured document standards, encapsulating the content and information attached to the electronic health documents in a single file.

**Keywords**—Portability; Encapsulation, HL7 CDA; HL7 SDA

## I. INTRODUCTION

Portability of software products is defined as their capability to be transferred from one platform to another [1]. The Extensible Markup Language (XML) is a W3C standard created to support the development of portable, open, human-legible, flexible web applications [2]. HL7 makes use of XML as a technology to represent instances of their Version 3 message specifications, as well as Clinical Document Architecture (CDA) and Structured Document Architecture (SDA) specifications.

XML provides HL7 structured documents with an encoding syntax which guarantees interoperability. However, portability and usability is not guaranteed when other files and related documents have to be attached to the structured documents, as it might be necessary to know in advance (or modify) the paths to the included objects.

Currently some storage formats for electronic documents using XML syntax do exist which allow to encapsulate in a single file different related documents, e.g., standards such as the ISO/IEC 26300 OpenDocument [3] and the ECMA-388 Open XML Paper [4]. However, specific file formats for electronic health documents do not exist.

A Special Interest Group (SIG) in HL7 Colombia identified the need for such a format and decided to work on the design and implementation of a specification named DUFF HL7 (HL7 electronic Unified Document File Format).

## II. OBJECTIVES

The objective of the paper is to propose a standard file format for electronic health documents which has the following properties:

1. Provide an electronic file format that facilitates the portability, usability and human legibility.
2. Present electronic health documents as a composite information unit. It integrates different instances of HL7 structured document specifications (CDA, CCD, SPL, Order Sets, and other structured documents) in one single unified format.
3. Include other contextual information (e.g. images, video, PDF, etc) in the encapsulated computer file.
4. Guarantee the aforementioned features, without altering the characteristics of HL7 electronic documents, or limiting their interoperability, performance attributes, and processing capabilities in software applications and devices.

## III. THE UNIFIED FILE FORMAT FOR HL7 ELECTRONIC DOCUMENTS – DUFF

The Unified File Format for HL7 electronic documents – HL7 DUFF is a standard file format for storing electronic health documents which complements the Structured Document Architecture (HL7 V3 SDA, R1) and Clinical Document Architecture (HL7 CDA R2; ISO/HL7 27932:2008) specifications which are part of HL7 v3.

DUFF HL7 stores, and supports the exchange, of electronic health documents as a single information unit which includes not only the XML instances of HL7 structured documents, but also contextual information associated to the document in other formats (DICOM, jpeg, gif, png, pdf, etc).

HL7 DUFF is a compressed ZIP file which uses the extension *.hl7* and contains a set of files and folders as described in Figure 1. Following, the main files in HL7 DUFF are described:

- **content.xml**: This file stores the main content of the structured document and uses XML syntax. Its structure supports HL7 v3 structured documents (CDA, CCD, SPL, etc).

- **meta.xml**: This file contains the metadata of the electronic document, e.g., type of template, type of structure, document ID, date and time of creation, author, target record, etc.

Nombre	Tamaño	Tipo	Fecha de modificación
Images	137,4 KiB	Carpeta	26 abril 2010, 19:57
META-INF	884 bytes	Carpeta	26 abril 2010, 21:22
Object1	116,9 KiB	Carpeta	26 abril 2010, 19:57
Styles	20,4 KiB	Carpeta	26 abril 2010, 20:28
Thumbnails	2,4 KiB	Carpeta	26 abril 2010, 19:18
content.xml	11,3 KiB	Documento XML	26 abril 2010, 21:08
meta.xml	4,6 KiB	Documento XML	26 abril 2010, 21:08
mimetype	23 bytes	Desconocido	26 abril 2010, 22:10

Figure 1. Structure of an HL7 DUFF

- **mimetype**: This file encodes information as MIME (Multipurpose Internet Mail Extensions) for sending the content by e-mail.
- **settings.xml**: This file contains properties for document processing and reading for specific software applications (optional).
- **Styles/styles.xsl**: This file contains information about the document's display style, using the W3C Extensible Stylesheet Language (XSL).
- **META-INF/manifest.xml**: It contains a list of all files in the *.hl7* file.

The included folders are dedicated to:

- **Images/**: Stores image files (jpeg, jpg, gif, png, bmp, svg, etc) associated with the document.
- **Objectn/**: Stores the related files in the main document (other structured documents, DICOM, pdf, odt, doc, docx, etc).
- **META-INF/**: Stores the manifest.xml file.
- **Styles/**: Stores the styles.xsl file, the cascading style sheets (css) and all other files necessary for displaying a document.
- **Thumbnails/**: Stores an image of a document preview (optional).

It is clear that HL7 DUFF offers a clear separation of the electronic document itself, its metadata and visualization information (style), nevertheless storing all information in a single file.

In order to demonstrate how an HL7 DUFF file could be displayed, prototype plug-ins for Firefox and Google Chrome have been developed. The plug-ins provide free and simple mechanisms for document presentation. Figure 3 shows a HL7 DUFF displayed in Firefox.

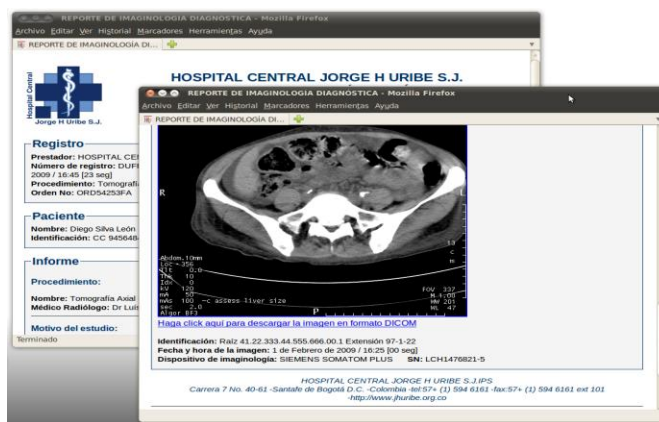


Figure 2. A CDA document in a HL7 DUFF displayed in Firefox.

#### IV. DISCUSSION AND CONCLUSION

The file format proposed in HL7 DUFF is based on widely used standard formats for electronic documents such as the OpenDocument and Open XML Paper specifications. HL7 DUFF is a complementary approach to HL7 Self Displaying CDA (DSTU), facilitating the inclusion of additional multimedia content and other related files. When exchange of structured documents between offline applications (using media such as CD, DVD, USB-memory, etc) is required, HL7 DUFF can be used in conjunction with the *ISO 9660 removable storage media R1* specification. For this occasion, it is necessary to relate the HL7 DUFF artifacts in the HL7DIR.xml file and store it in the HL7/ folder. As the information is encapsulated in the HL7-DUFF format, it is not necessary to use non-HL7 directories to store the document attachments.

In conclusion, HL7 DUFF improves portability, usability (displaying) and human readability of electronic health documents. HL7 DUFF complements the HL7 structured document standards, encapsulating the content and information attached to electronic health documents in a single file. If accepted, HL7 DUFF would be the first HL7 v3 specification proposed by HL7 community members in Latin America.

#### ACKNOWLEDGMENT

The authors are in debt to thank Diego Kaminker, HL7 Argentina, and partners from HL7 Colombia for their comments on the paper. This work was founded by the QUIPU Program under FIC/NIH grant D43TW008438, Ministerio de educación in Colombia under contract IF004-09 (ImagenMantis Project) and the University of Cauca.

#### REFERENCES

- [1] ISO/IEC 9126: Software Engineering -- Product Quality, part 1. 2001.
- [2] World Wide Web Consortium (W3C). <http://www.w3.org>
- [3] ISO/IEC 26300. Information technology--Open Document Format for Office Applications (OpenDocument) v1.0, 2006
- [4] ECMA-388. Open XML Paper Specification, European Computer Manufacturers Association, 2009
- [5] ISO/IEC DIS 9660: File System Specification, 1999