Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-317

Automated Import of Clinical Data from HL7 Messages into OpenClinica and tranSMART Using Mirth Connect

Juan Carlos CAMACHO RODRIGUEZ[1], Sebastian STÄUBERT and Matthias LÖBE
Institute for Medical Informatics, Statistics and Epidemiology (IMISE)
Leipzig University, Germany

Abstract. Electronic data capture (EDC) tools are designed to simplify data acquisition, improving data quality and managing clinical data electronically. Some data are collected from the laboratory information management system (LIMS), which is an important data source for a study. OpenClinica is an open source clinical data management system (CDMS) for web-based electronic data capture (EDC), which is used widely in academic clinical research. TranSMART is also an open source web-based platform used for the management and analysis of different data types common in clinical and translational research. Many LIMS use the Health Level 7 standard – Version 2.x (HL7) as a message exchange protocol. In this paper, we implement Mirth Connect as a Communication Server (CS) to convert these HL7 messages either to Operational Data Model (ODM) data for the automatic import in OpenClinica or tabular-delimited text format files, whose data is uploaded in tranSMART using the tMDataLoader tool.

Keywords. HL7 message exchange, Web services, CDISC ODM, Mirth Connect, OpenClinica, tranSMART, tMDataLoader.

1. Introduction

In medicine, electronic data capture (EDC) of clinical data of trial subjects is a state-of-the-art technique to get empirical data for clinical trials. The data flow process between clinical source application systems (e.g. laboratory information management system (LIMS)) and research application systems (e.g. OpenClinica) has to be considered as of particular importance. This applies for use cases like a) an embedded clinical laboratory is conducted to analyze samples of trial patients or b) existing primary care laboratory data is to be added to complement research data (secondary use), since the data are the basis for research and therewith a decisive factor of success in the expected results. OpenClinica[2] is a professionally and commercially supported open-source clinical data management system (CDMS), which offers a web-based infrastructure for EDC used in clinical research [1]. TranSMART[3] is an important open-source, web-

[1] Corresponding Author: IMISE, Leipzig University, Härtelstr. 16-18, 04107 Leipzig, Germany; E-mail: juancasax@gmail.com.
[2] http://www.openclinica.com
[3] http://www.transmartfondation.org

based application, based on i2b2[4], and used for the analysis of both clinical and molecular data [6].

An important provider for clinical data for primary care and clinical research are LIMS [7]. Usually, the data is obtained from laboratory tests as well as administrative patient data. Frequently, the data from LIMS is sent back to the research site either in hard copy (e.g. a report by fax) or in digital form (e.g. a CSV file by email, remote access to LIMS application). Neither one is directly interpretable by OpenClinica or tranSMART (since the data is not structured in the right format to be imported). Consequently all data must be re-entered manually by a data entry person, which could be subject to typing errors. This kind of errors leads to errors in the analysis and final results of the research. Together with other possible complications (e.g. availability and costs) that could occur in this process, this produces obstacles that affect and prolong the runtime of the study.

In this work we aim to develop a method, which imports clinical data electronically from LIMS into tranSMART and OpenClinica by using Mirth Connect[5] as a communication server (CS) in order to facilitate the transfer of clinical data with a minimum of human intervention.

2. Methods and materials

2.1. Communication standards

To create a link between LIMS, tranSMART and OpenClinica we consider the respective document types, which are supported in Mirth Connect, tMDataLoader and OpenClinica Web Services.

- *Health Level 7 – Version 2.x (HL7)* is a communication standard that allows to exchange, share and integrate clinical data in form of a text format message. It packages clinical data, defines how it will be structured and how it will be sent between application components.
- *Operational Data Model (ODM)* is a standard introduced by CDISC[6] to represent clinical trials data to support data interchange and archives, specifying the clinical trials data in a XML application.
- *Tabular data files* are simple text format files, whose values are tab-separated in order to create a tabular structure. This format is supported by different systems, which define the role of each value in the file (the use and definition of this type of files is explained in section 2.2.3).

2.2. Communication server

In this work, we use Mirth Connect v3 as CS. It is a health care integration engine [2]. It is used to transmit and translate laboratory data from LIMS either to OpenClinica or tranSMART. The LIMS is connected to the CS via HL7. CS and OpenClinica are connected using Web Services [3]. CS and tranSMART are connected to control the

[4] http://www.i2b2.org
[5] http://www.mirth.com
[6] http://www.cdisc.org

tMDataLoader tool (see section 2.3) to import the clinical data into tranSMART. This process is shown in Figure 1. Mirth Connect uses *channels* for connecting health information systems (HIS), which filter, transform and route messages based on user defined rules using connectors (as inputs and outputs).

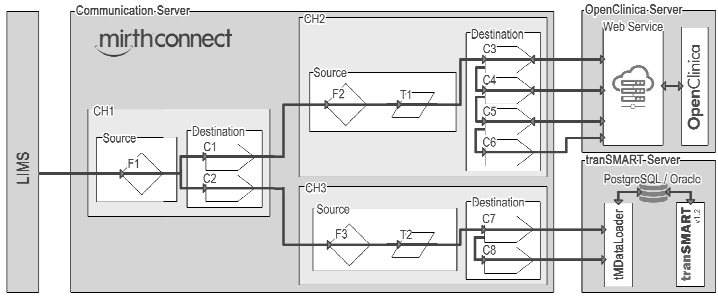

Figure 1. HL7 message processing to transfer clinical data.

2.2.1. Channel 1 (CH1: LIMS → CS)

- **Source:** The source system (in our case LIMS) sends HL7 messages to the CS. This channel listens continuously to a previously defined Transmission Control Protocol (TCP) port. Once Mirth Connect receives the message, all required values for the clinical studies are mapped using the *Filter Validation* F_1*:* The received message has to be parsed. It validates if the receiving application corresponds to OpenClinica or tranSMART.
- **Destination:** The destination contains two connectors of the type channel writer. They send all mapped data to the connection channels, which are in charge of building the files and establishing a communication with OpenClinica Web Services (channel 2 (*CH2*) using connector C_1) or tranSMART server (channel 3 (*CH3*) using connector C_2).

2.2.2. Channel 2 (CH2: CS → OpenClinica)

- **Source:** the data is received from from *CH1*:
  - *Filter Validation* F_2*:* The message is filtrated applying a set of rules to validate the entire HL7 message. This procedure checks that all patient data required for a study in OpenClinica is available and complete.
  - *Transformation Mapping* T_1*:* The connection channel maps the message through a developed JavaScript procedure, which transforms the laboratory data into single variables required for the CRF. These variables are read by the destination connectors inside the connection channel.
- **Destination:** Its connectors communicate with OpenClinica in order to transfer a Clinical Data Interchange Standard Consortium (CDISC) ODM file with medical and administrative data, as well as to manage responses received from OpenClinica. This process is realized via Web Services. To match the information acquired from LIMS correctly with a study in OpenClinica, the channel uses connectors, which must be able to create a study subject in OpenClinica (C_3), schedule an event (C_4), request the subject's ID (C_5) and send the information back. This is realized by mapping the XML response

obtained from OpenClinica. Once all the requirements are completed, the CH2 is ready to create a CDISC ODM file in the last connector C_6, implementing the mapped values received from LIMS and OpenClinica. It creates a new ODM file with all data required in the electronic Case Report Form (eCRF). The CS sends this file via Web Service to the destination (OpenClinica). The Web Service collects the file, submits and imports the laboratory data into the target eCRF of a defined study event CRF.

2.2.3. Channel 3 (CH3: CS → tranSMART)

- **Source:** Similar to *CH2*, the data received from *CH1* is read and filtered:
  - *Filter Validation* F_3: The message is filtrated like F_2 checking the availability of the patient data necessary for a clinical study defined in tranSMART.
  - *Transformation Mapping* T_2: The message is mapped using a JavaScript procedure, which transforms the laboratory data into single variables required to be imported into tranSMART for a subsequent clinical study.
- **Destination:** It counts with two connectors of the type File Writer: the connector C_7 creates two sub-folders in the *public studies* folder (or a defined private studies folder) in the tranSMART server. A folder is created with the *study_ID*, which contains the second folder called *ClinicalDataToUpload*. The tMDataLoader tool requires this folder structure in order to load the clinical data into tranSMART. Afterwards, a mapping file (tabular-delimited text format) is built (named *study_ID_Mapping_File.txt*), which contains the definition and structure of the data. The second connector C_8 creates the data file (same format as the last file), which includes all clinical trial data values. These files are now ready to be imported into tranSMART.

2.3. tMDataLoader tool

This tool runs in the tranSMART server side and uses a tranSMART ETL tool supported by Thomson Reuters (tm_etl.jar) [5]. It allows to import the clinical and molecular data into a tranSMART platform. The files generated by the CH3 are placed in its import folder. Once the tool has been executed, the clinical data is inserted and made available in the tranSMART platform.

3. Results

An efficient and completely automatized direct communication technique between LIMS, OpenClinica and tranSMART has been realized successfully. This allows the transfer of clinical data via a HL7 message coming from LIMS, parse it by performing filtering and transformation methods. Subsequently, this data will be exported as ODM messages to the Web Service, which imports the values directly into an eCRF in OpenClinica, or as tabular-delimited text file format in order to import the clinical data in tranSMART through the tMDataLoader tool. In this project, the human involvement has been reduced to installation, configuration and maintenance of the implicated

application systems. Due to this and the implementation of open source products, resources are optimized. Furthermore, by establishing electronic communication the data flow is constant, up to date and quick. This means that the operating time is reduced significantly and the process will only take a few seconds. In this way, studies that demand laboratory data will be annotated with first-hand clinical data, without typing errors that jeopardize data quality.

4. Discussion

This method offers efficiency, operability and security to send and enter laboratory data between systems, while at the same time minimizing costs and ensuring quality of the data collected in OpenClinica and tranSMART. Secondary use of clinical data has some issues [4]. These issues can be faced by using rules and filters, which select laboratory results of a specific patient for a certain study before the import process. These filters can also associate the data to a specific patient in a clinical study using an updated database, which contains mapping tables that determine relations between patient ID, study, receiving application and required data. The source system pushes HL7 messages to be evaluated in the CS. These messages are processed and delivered, when the receiver requests the transmission and the mapping tables have been upgraded. In case that an exception in a filter occurs, produced by a lack or inconsistency of data or problems during the data transfer, the issue can be protocolled and the user can be notified by email. Consequently, the failed data can be stored in form of a text file (HL7 message) in a folder established for this purpose.

The realization of the communication process is technically possible through data transfer standards, networking and a Web Services API offered by OpenClinica. These are standards, which Mirth Connect is able to process and which allow to control and transfer the data between LIMS and OpenClinica. On the other hand, the tMDataLoader has been applied too, in order to import the data sent from Mirth Connect into tranSMART.

References

[1] B. Baumann. *Overview of OpenClinica | OpenClinica Reference Guide @ONLINE.* http://docs.openclinica.com/3.1/openclinica-user-guide/overview-openclinica, 2014. Docs.openclinica.com. Accessed March 2nd, 2015.

[2] J. Brauer. *Getting Started Guide – Mirth Connect – Confluence @ONLINE. http://www.mirthcorp.com/community/wiki/display/mirth/Getting+Started+Guide,* Nov. 2013. Mirthcorp.com. Accessed February 10th, 2015.

[3] C. Collings. *Using OpenClinica Web Services* | OpenClinica Reference Guide @ONLINE. *https://docs.openclinica.com/3.1/technical-documents/openclinica-web-services-guide/using-openclinica-web-services*, 2014. Docs.openclinica.com. Accessed March 2nd, 2015.

[4] K. Pongpirul, D. G. Walker, P. J. Winch, and C. Robinson. A qualitative study of DRG coding practice in hospitals under the Thai universal coverage scheme. *BMC health services research,* 11(1):1, 2011.

[5] E. Rakhmatulin and S. Bureeva. *tMDataLoader – tranSMART ETL AND DATA FILES MANUAL,* May 2015. Wiki.transmartfundation.org. Accessed March 8th, 2016.

[6] A. Schumacher, T. Rujan, and J. Hoefkens. A collaborative approach to develop a multiomics data analytics platform for the translational research. *Applied & Translational Genomics*. 3(4): 105-108, 2014.

[7] A. Winter, R. Haux, E. Ammenwert, B. Brigl, N. Hellrung, and F. Jahn. *Health information systems – Architectures and Strategies.* 2nd ed. Springer, 2011.