

Automated Transformation of CDISC ODM to OpenClinica

Sophia GESSNER^{a,1}, Michael STORCK^a, Stefan HEGSELMANN^a, Martin DUGAS^a
and Iñaki SOTO-REY^a

^a*Institute of Medical Informatics, University of Münster, Germany*

Abstract. Due to the increasing use of electronic data capture systems for clinical research, the interest in saving resources by automatically generating and reusing case report forms in clinical studies is growing. OpenClinica, an open-source electronic data capture system enables the reuse of metadata in its own Excel import template, hampering the reuse of metadata defined in other standard formats. One of these standard formats is the Operational Data Model for metadata, administrative and clinical data in clinical studies. This work suggests a mapping from Operational Data Model to OpenClinica and describes the implementation of a converter to automatically generate OpenClinica conform case report forms based upon metadata in the Operational Data Model.

Keywords. Data Collection, Access to Information, Health Information Systems, Metadata, Surveys and Questionnaires.

1. Introduction

Nowadays, medical research data is predominantly documented in electronic data capture (EDC) systems [1]. The benefits of EDC systems when compared to paper-based data collection, such as facilitated data management and reduced data collection costs, have been discussed previously [2–4]. OpenClinica (OC), an open source EDC system, used in over 3000 clinical studies [5], represents an example for an EDC system. A crucial task, setting up an EDC based study is the generation of electronic case report forms (eCRFs). They contain the metadata for the collected data, i.e. item names, data types, measurement units, questions and answer options.

In OC, eCRFs can be defined in an Excel template to import them into the system [6]. To fill the eCRFs with clinical data, researchers may enter data directly via the user interface. Additionally, OC allows uploading existing clinical data via a REST interface, as long as it is stored in CDISC Operational Data Model (ODM), a platform-independent format for exchanging and storing metadata, administrative and clinical data [6, 7]. In order to further process or share data, OC provides various export formats, including formats used by statistical evaluation tools and ODM.

There is a discrepancy in OCs support for ODM. While clinical data can be imported as ODM files, metadata cannot. The only way to generate eCRFs is to upload metadata via an OC Excel template, hampering the reuse of pre-defined metadata in

¹ Sophia Gebner, Institute of Medical Informatics, Albert-Schweitzer-Campus 1, 48149 Münster, Germany; E-mail: Sophia.Gessner@uni-muenster.de.

standard formats. The demand for reuse of metadata in order to save time establishing eCRFs and even routine documentation forms is growing [8].

A research infrastructure, called Portal of Medical Data Models (MDM) [9], an open-access repository for medical forms, offers more than 10.000 medical forms in ODM for reuse in medical research.

The objective of this work is to provide an automated transformation of given metadata from ODM to the OC Excel import template.

2. Methods

Based upon the current version of ODM, 1.3.2, a mapping from ODM to the OC Excel spreadsheet template, v3.9, was defined.

In order to provide the possibility of an automated transformation of ODM's metadata into the Excel template, an ODMtoOpenClinica converter was developed using Java programming language. To validate the ODM files against the ODM 1.3.2 schema, a previously implemented ODM validator [10] was used and improved. To transform elements from the XML structure of ODM to Java objects, a Java library developed by the University of Münster, was used [11]. In order to adhere to OC's restrictions concerning the structure and naming of metadata, OC specific validations and adaptations were developed. The converter was integrated into the Portal of MDM. Furthermore the converter can directly be accessed via web interface or an API [12].

A test set of 10 ODM files with intentionally generated exceptional cases, such as multiple form definitions, missing ODM elements and files, covering all data types and range checks available in ODM, were converted. Additionally 30 random samples were downloaded from the Portal of MDM, converted with ODMtoOpenClinica and examined in OC. Errors leading to invalid OC CRFs were identified and corrected. For errors that could not be resolved due to incompatibility between ODM and OC CRF, a log file is generated and provided to the user.

3. Results

Elements of the ODM XML tree structure, relevant for the mapping to the corresponding columns of the OC template, are presented in [Figure 1](#). The root XML element of an ODM file is named ODM. Its child element Study contains the study's metadata information as GlobalVariables with study name and description, BasicDefinitions for the definition of measurement units and MetaDataVersion elements. Further elements serve as references, such as ItemRef elements, linking elements within and across ODM files. To simplify the presentation of the mapping, ODM, GlobalVariables and reference elements have been omitted in [Figure 1](#).

The OC template consists of five Excel sheets that specify the metadata necessary to set up a study form in OC:

- CRF, capturing information like the name and version of the CRF
- Sections, indicating the layout and organization of items
- Groups, grouping items logically together
- Items, defining all variables and their attributes
- Instructions, briefing users generating forms

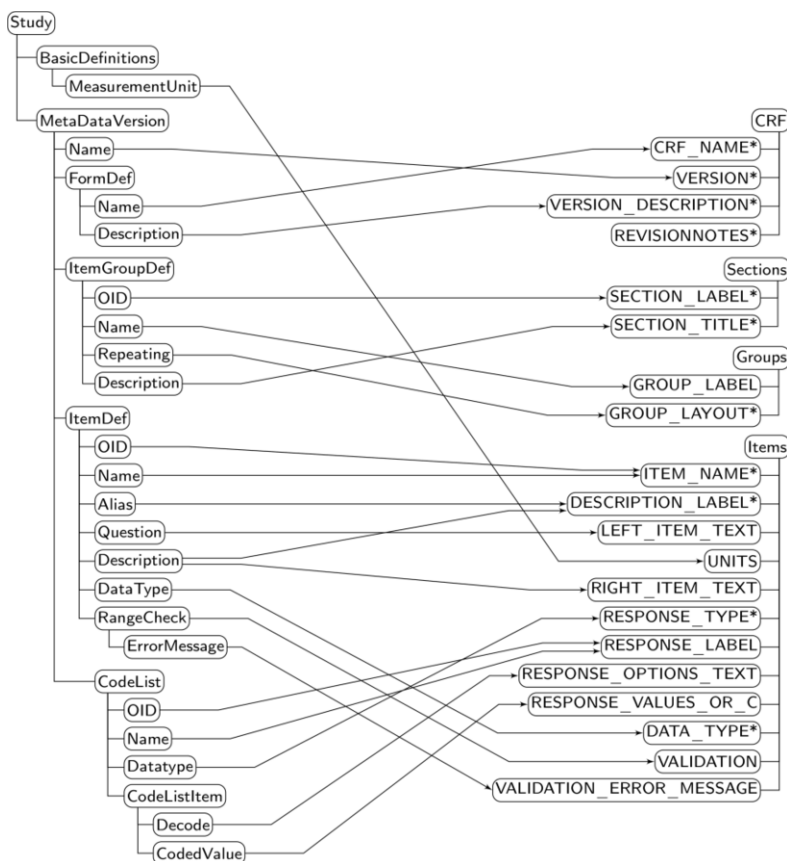


Figure 1. Mapping from ODM to the OC Excel template. (*=required fields in OC)

Figure 1 presents the excerpt of the ODM structure on the left side, linked via arrows to the corresponding elements of the OC template on the right side.

The converter implements the mapping to enable the automated transformation of ODM files into instantly applicable OC eCRFs. While ODM allows the definition of metadata for a whole study with multiple study events and various forms, the OC template is designated to represent a single form in a study. An ODM file containing multiple forms is transformed into multiple Excel files indicating FormDefNames and StudyEventDefNames as filenames.

Reference elements may indicate order numbers on items. If specified, items are sorted ascending, enumerated and transferred to QUESTION_NUMBERS, otherwise the order of items is derived from their positions in ODM. Mandatory items, also defined in ItemRefs, are mapped to the REQUIRED field in OC. The various DataTypes and their mapping are presented in Table 1.

ODM supports multilingual forms, by permitting multiple TranslatedText elements for ItemDefQuestions. Since OC only allows one language to be uploaded in one CRF, the converter checks the ODM for an indicated language code and chooses items in the language available. If no code is transmitted, English TranslatedTexts are used to generate the CRF, otherwise texts in German or the default language serve as backup.

Table 1. Mapping of ODM Item and CodeList DataTypes and their corresponding elements in OC. Further DataTypes, not presented in the table, provided by ODM are mapped to the OC DATA_TYPE “ST” and the OC RESPONSE_TYPE “Text”. (ST=any character string, INT=whole numbers, REAL=decimal numbers, PDATE=partial date, either MMM-YYYY or YYYY, Textarea=multi-line box, Text=one-line box)

ODM DataType	OC DATA TYPE	OC RESPONSE TYPE
text/string	ST	Textarea
integer	INT	Text
float	REAL	Text
date	DATE	Text
time	ST	Text
boolean	INT	Single-Select (Yes, No)
partialDate	PDATE	Text

Certain OC specific requirements demanded preprocessing of elements. The DESCRIPTION_LABEL gives further information about an item, resembling the information contained in the Description attribute of items in ODM. As the Alias attribute in ODM gives the possibility to indicate further synonyms of the items, this information is also adopted into the DESCRIPTION_LABEL field. Since the REVISION_NOTES field in OC is mandatory, we followed OCs instructions to fill it with the note “This is a brand new form.”. With every converted file, a record of all modifications is provided to the user via supplementary csv files, called “conversionNotes”. The modifications are classified by their severity level concerning the resulting OC file: NOTICE, for minor changes to generate valid OC files, WARNING for metadata alternating changes and CRITICAL, for modifications that require manual post processing of the OC template after conversion.

Some elements of the OC template, such as subtitles, instructions or layout information, could not be filled in, since the information is not available in ODM. These elements are additional information and not obligatory in OC. RESPONSE_LAYOUT is left blank, displaying items in OC vertically.

4. Discussion

This work presents a solution to the limited possibilities of reuse of eCRFs in a widely used open-source EDC system. By creating a mapping from ODM, the implementation of a converter enabling the automated transformation of ODM files to OC compliant Excel sheets, became feasible.

The integration of ODMtoOpenClinica into the Portal of MDM enables the reuse of more than 10.000 medical forms from clinical trials, routine documentation, standard instruments and common data elements, available in ODM, paving the way for multicentric standardized structured documentation. Medical concepts of items in the Portal of MDM are semantically annotated with Unified Medical Language System codes. This information is transferred to OC, leading to a better semantic interoperability of clinical data. Nevertheless, it might be a valuable feature extension for OC to enable semantic annotation of medical concepts in a separate field.

By generating one Excel file per ODM form, the converter flattens the hierarchy of the original structures. Still, an ODM file may contain further information about the study, such as study title and study events. This information could not be included in the OC template, so the manual input and editing of the study title and definition of

study events in OC is inevitable. A newer version of the OC's Excel import template should include Study information, which would enhance its interoperability with ODM.

The efficiency improvement in the set-up of studies as well as the conversion of ODM files issued from other EDC systems should be evaluated.

5. Conclusion

The proposed mapping between ODM and OC enabled the development of a converter for ODM forms into OC compliant forms. By integrating the converter into the Portal of MDM, more than 10.000 medical forms including 6.750 from clinical trials can easily be imported to OC and used for data collection.

Acknowledgements and Conflict of Interest

This work was supported by German Research Foundation (Deutsche Forschungsgemeinschaft, DFG grant DU 352/11-1). We would like to thank Maurice Heine for his contribution to this project. The authors state that they have no conflict of interests.

References

- [1] K.A. Kessel, S.E. Combs, Review of Developments in Electronic, Clinical Data Collection, and Documentation Systems over the Last Decade - Are We Ready for Big Data in Routine Health Care?, *Frontiers in oncology* **6** (2016) 75.
- [2] I. Pavlovic, T. Kern, D. Miklavcic, Comparison of paper-based and electronic data collection process in clinical trials: costs simulation study, *Contemporary clinical trials* **30** (2009) 300–316.
- [3] P.V. Staziaki, P. Kim, H.V. Vadvala, B.B. Ghoshhajra, Medical Registry Data Collection Efficiency: A Crossover Study Comparing Web-Based Electronic Data Capture and a Standard Spreadsheet, *Journal of medical Internet research* **18** (2016) e141.
- [4] I.C. Olsen, E.A. Haavardsholm, E. Moholt, T.K. Kvien, E. Lie, NOR-DMARD data management: implementation of data capture from electronic health records, *Clinical and experimental rheumatology* **32** (2014) S-158-62.
- [5] OpenClinica - Electronic data capture for clinical research, <https://www.openclinica.com/>, accessed 8 February 2017.
- [6] OpenClinica Reference Guide, <https://docs.openclinica.com/3.1/study-setup/build-study/create-case-report-forms-crfs>, accessed 8 February 2017.
- [7] CDISC - Clinical Data Interchange Standards Consortium, Operational Data Model (ODM) - XML, 2016, <https://www.cdisc.org/standards/foundational/odm>, accessed 21 November 2016.
- [8] M. Dugas, K.-H. Jöckel, T. Friede, O. Gefeller, M. Kieser, M. Marschollek, E. Ammenwerth, R. Röhrig, P. Knaup-Gregori, H.-U. Prokosch, Memorandum "Open Metadata". Open Access to Documentation Forms and Item Catalogs in Healthcare, *Methods of Information in Medicine* **54** (2015) 376–378.
- [9] Metadata Repository for Medical Forms, <https://medical-data-models.org/>, accessed 29 May 2017.
- [10] M. Storck, R. Krumm, M. Dugas, ODMSummary: A Tool for Automatic Structured Comparison of Multiple Medical Forms Based on Semantic Annotation with the Unified Medical Language System, *PLoS one* **11** (2016) e0164569.
- [11] I. Soto-Rey, M. Dugas, M. Storck, Implementation of an ODM and HL7 Compliant Electronic Patient-Reported Outcome System, *Studies in health technology and informatics* **228** (2016) 421–425.
- [12] ODMToolbox, <https://odmtoolbox.uni-muenster.de/>, accessed 29 May 2017.