

Queries management: Interfacing SAS with in-house EDC

Michel Toussaint, Clin Data Management, Rouffach, France
Virginie Freytag, Clin Data Management, Rouffach, France

ABSTRACT

Despite the growing use of EDC by pharmaceutical companies in clinical trials and the confirmed leverage of this technology, off-line edit checks for data validation are still widely used.

SAS®, mainly due to its flexibility, is therefore an appealing alternative for their implementation. The management of generated queries is, however, usually a high hurdle to overcome, particularly in large multicenter studies. Most EDC applications on the other hand offer full user friendly possibilities for manual queries management by the monitor.

In this paper we will show how we successfully interfaced our SAS workflow for automatic queries generation with our in-house EDC application, thus enabling to combine the advantages of both sides for the efficient management of off-line queries

INTRODUCTION

In clinical trials, the collected data can be inconsistent and need to be corrected. Basically, the cleaning of data consists in first, identifying inconsistencies and then sending appropriate queries to the investigator for data correction.

From the data management side, checking data for discrepancies is done using edit checks to target missing or out-of-range values, or checking consistency between item values.

With the use of EDC technology, such checks can be applied on-line during the data collection. EDC technology also allows the efficient management of the queries process: the queries are sent and answered directly in the application, and data can be corrected on-line while the EDC system tracks all the queries and records any data modification.

However, off-line checks, which are applied by the data manager on the study database using dedicated programs, are still widely used.

In fact, even if on-line controls are cost-saving due to the limitation of errors at the time of data entry, the implementation of complex controls becomes rapidly cumbersome in EDC platforms. On the other hand, bias that could result from exhaustive on-line controls can be subject to controversy.

SAS is an interesting alternative for the implementation of edit checks, especially in the working context of our company which is specialized in providing in-house fully integrated EDC solutions. SAS offers the possibility for creating optimized and flexible programs, needed for edit checks implementation.

Therefore, we have on the one hand, an EDC solution which provides a sophisticated feature to write manual queries and on the other hand, an efficient workflow for the implementation of off-line edit checks.

Our motive for interfacing our EDC solution queries module with our SAS programs is obvious and explained further in this paper.

ON-LINE EDIT CHECKS

Most EDC solutions provide on-line edit-checks in a rather effective way to reduce data entry errors and/or data discrepancies from the investigators' sites: the error is detected at the time of the data entry and the correction can be performed directly on-line. Such processes avoid numerous human omissions or errors (missing values, inconsistency of dates for example).

Clin Data Management's EDC application (WEBCDM), which is a web-based application developed in Microsoft C#.NET, enables classical on-line edit checks during data entry as well.

However, several issues arise with the use of on-line checks. Firstly, on-line edit checks can be easily developed when they are mainly based on data already contained inside the form itself and when the algorithm behind it is not too complex.

PhUSE 2011

Otherwise, the time delay to displaying the web page of the form could be seriously affected, thus considerably reducing the advantage of on-line technology.

Moreover, even if an inconsistency is found in the data, it is not obvious to define whether the correction should be mandatory or not. Exhaustive mandatory corrections would put the investigator under unreasonable pressure to enter data, which is of course not acceptable in the context of clinical trials where reliability of collected data must be ensured.

In our company, we have defined two types of checks clearly distinguishable on-line. Mandatory checks which are carefully defined at the beginning of the study prevent the form submission in case of detected inconsistency while most other checks, are warning checks meaning that no correction is needed to submit the form.

In this second case, some inconsistencies exist in the database and off-line edit checks have to be applied for sending corresponding queries to the investigator.

OFF-LINE EDIT CHECKS USING MANUAL EDITING TOOL

Our EDC application includes a query management module used during the data cleaning process by clinical research associates (CRAs), data managers and of course by investigators.

Below is a description of the function of this module stressing its ease of use and efficiency.

WRITING QUERIES (CRA OR DATA MANAGER)

This screenshot displays the intuitive user interface developed for creating queries inside a form. The user simply ticks the inconsistent item inside the form and selects a classical message from a drop down list. The content of the list is set up for each study accordingly with specific needs and can be adapted by country and/or language within the study.

The query generated is positioned by the author in the form, so that users (including investigators) can easily identify which data entered are inconsistent.

The screenshot shows a portion of an EDC form. The 'Date of visit' section has a date field set to 05/04/2011. Below it, the 'Informed consent' section has a date field for 'Date of written informed consent'. A 'Question?' dropdown menu is open, showing a list of options: 'Please specify/document', 'Please check', 'Please check and correct', 'Please provide indication', and 'Other'. The 'Please check and correct' option is highlighted.

Figure 1: Creating a query

ANSWERING QUERY (INVESTIGATOR)

This screenshot displays how the investigators can efficiently answer and close several queries. The tool automatically highlights the item concerned by the query inside the form, so that the investigators can directly visualize the corresponding CRF section. They have classical possibilities to confirm or correct the data. In case of correction, they have direct access to the form for modifying data, in the same way they previously entered the data.

The screenshot shows the EDC form with the 'Informed consent' and 'Demography' sections. The 'Date of birth' field in the 'Demography' section is highlighted in red, indicating an inconsistency. The 'Query to be answered' panel on the left shows the query details and options to confirm, correct, or not applicable. The 'Form to be corrected. Inconsistent data highlighted in red.' section shows the 'Date of birth' field with a red border and the date 25/04/2010.

Figure 2: Answering a query

QUERIES SOLVING (CRA OR DATA MANAGER)

From the investigator's answer, the users can then close the query. In the case of an unsatisfactory answer or correction, they have the possibility to send back the query. Therefore, the query flow is similar to the one described above, except that the users can also track the full query background.

PhUSE 2011

QUERIES SUMMARY

As the number of queries can rapidly increase during the course of the study, all users have access to an adapted summary (author and date of the query, investigator sites, status of the query...), so that they can easily navigate through the study queries.

Of course, despite the customized features of this module, in large multicenter studies it rapidly becomes a challenge to manually generate all queries retrieved from the study database through off-line edit checks, all the more as a manual process is likely to introduce errors (e.g., queries sent in the wrong form or for the wrong patient). Automated queries generation is therefore required.

AUTOMATIC OFF-LINE EDIT CHECKS PROCESS

The flexibility of SAS, as already mentioned, partly justifies the SAS implementation choice for off-line edit checks. Moreover, SAS programs are processed on a clinical structured database which first presents the advantage of being independent of the EDC application, and secondly allows rather complex edit checks to be implemented.

PROPOSED ARCHITECTURE

In our company, structured data panels (Data Warehouse) are directly extracted from the eCRF database using an ETL (Extract Transform and Load) stand alone tool.

The eCRF (EDC) database is developed to manage the on-line data entry process (eCRF forms are stored in XML format), in the most efficient way, whereas the structured database is more adapted to data management and statistical processes. Both databases are implemented on Oracle.

The extraction tool from the eCRF database is launched through a task scheduler on a regular basis depending on the study needs.

Therefore, automatic edit checks controls can be applied by dedicated SAS programs on the updated structured data and subsequently, the inconsistencies can be loaded by the EDC application.

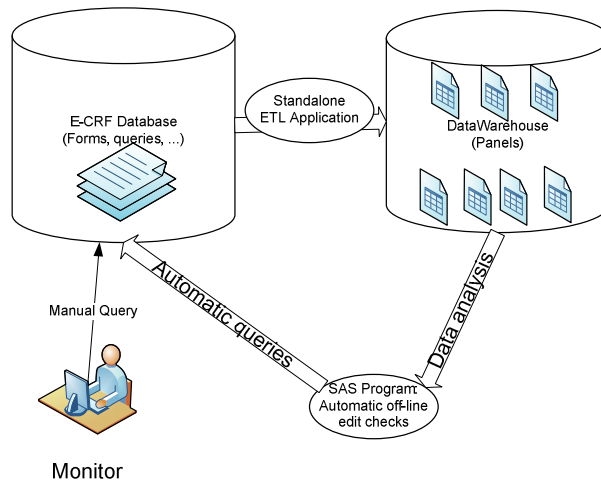


Figure 3: Global architecture

DATA FLOW AND AUTOMATIC PROCESS

At the beginning of the study, the following queries management tables are set up:

- the ecrf queries table which will contain all the queries sent within the EDC application
- the SAS queries table which will contain all the inconsistencies found in the study database
- the Check fields table which indicates for each SAS edit check, the suitable item position within the concerned eCRF form.

PhUSE 2011

STEP 1 - AUTOMATIC SENDING OF QUERIES

Inconsistencies targeted by SAS programs are associated with appropriate identification items (*i.e. check id, country, patient number, visit, ...*) and queries messages; the SAS queries table is then updated. Translation of queries messages in local languages can also be loaded during the edit check execution.

Then, an SAS program loads new identified inconsistencies and associates the corresponding eCRF field to each query. Finally, a single .csv file is exported, containing all the information needed to generate queries in the EDC application. This file is scanned by the query management module of the EDC application and the automatic queries are sent: for each SAS inconsistency, a query is sent within the EDC application for the identified patient/visit, in appropriate position within the eCRF form and with appropriate query message (in local language if applicable).

Therefore, on the EDC system side, automatic queries will be displayed exactly in the same way as if they were entered manually.

The SAS edit check programs are run automatically on a regular basis, along with the data warehouse update.

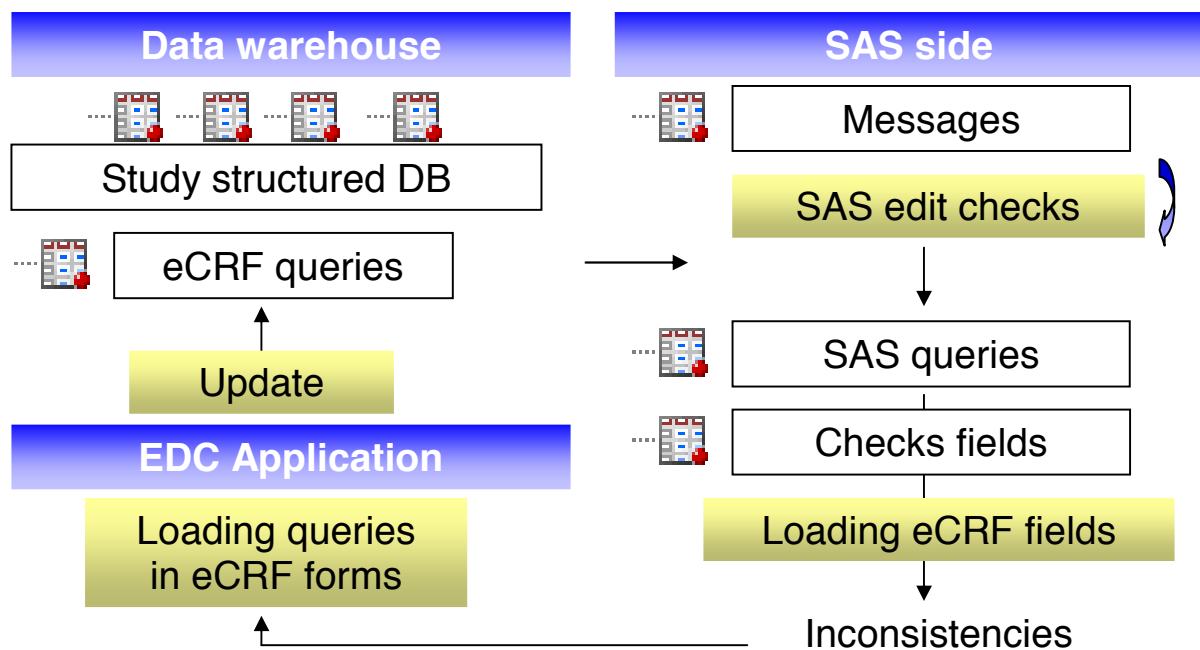


Figure 5: Data Flow diagram

STEP 2 - RECONCILIATION

Once they have been sent through EDC, off-line queries are managed manually directly from the query management module in the EDC application. A CRA or a data manager updates the query status on-line according to the investigator's answer, as described in the first section.

For completing the cleaning process, the data manager has next to perform the reconciliation of eCRF queries and SAS inconsistencies during the trial:

- closing eCRF queries for which inconsistencies are not found any more in the study database,
- checking eCRF queries that are solved with existing inconsistencies in the study database

This is not a major issue, and it can be easily managed with both SAS programs comparing the status of queries/inconsistencies between databases, and specific workflow designed for managing the SAS inconsistency status.

PhUSE 2011

CONCLUSION

The integration of an SAS off-line edit checks management workflow with our EDC application is a concrete example of how the data management processes can be efficiently adapted to new eClinical technologies. The illustrated architecture enables full use of both the advantages of technologies used with the EDC system on the one hand, and the classical SAS programs on the other hand, resulting in the improvement of the overall data cleaning process. Working with in-house developed EDC applications offers interesting possibilities for further developments, and to extend this approach to other data management processes.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the authors at:

Authors Names : Virginie Freytag and Michel Toussaint
Company : Clin Data Management
Address : Zone artisanale, Rue d'Alsace
City / Postcode: 68 250
Work Phone: +33 (0)3 89 49 56 60
Fax:: +33 (0)3 89 78 51 81
Email: v.freytag@cdm-trials.com and m.toussaint@cdm-trials.com
Web: www.cdm-trials.com

Brand and product names are trademarks of their respective companies.