PhUSE 2010 - Paper P007
Data Migration from Oracle Clinical to .XML using SAS
Matteo Ferrari, CROS NT, Verona, Italy - Valeria Visonà, CROS NT, Verona, Italy

ABSTRACT
The main strength of XML data structure consists in allowing the user greater flexibility in managing data in several platforms. The work performed and presented in this poster was motivated by a Sponsor’s request to export data contained in an Oracle Clinical® database into an .XML file. The Sponsor needed to own a data structure that could be used in different systems without changing the raw source data. We successfully managed to satisfy this request using SAS® as a connection between the two types of data. The aim of this poster is to present the processes followed, showing the SAS programs through which the final goal was attained. Furthermore, we would like to present the SAS program developed which allows you to automatically send the .XML file previously created, through the company mail server to the final user.

INTRODUCTION
Every platform for managing data used in the daily work has its own peculiarity. For a more flexible way to work it is useful to convert the data into a general data structure that can be easily used with several systems. In our experience to satisfy the requirements of the client, XML (eXtensible Markup Language) data structure enabled us to reach our goal. This standard language was needed to be used by the study monitors in the Clinical Trial Management System (CTMS) to maintain and manage the planning, performance, monitoring and reporting of the clinical trial; in this way monitors can easily use the data for different aims, such as scheduling of visits, assessments, etc. In the flux diagram presented below, the process followed starting from OC raw data to the delivery of the final product is summarized.

FIRST STEP: FROM ORACLE TO SAS
The first step consists in importing data into SAS format starting from Oracle Clinical data structure. This operation can be achieved using the SAS PROC SQL syntax to create a connection between the OC and the SAS structure. In the following, we restrict ourselves to presenting this process limited to certain information collected in the clinical database. In particular, in the next rows of SAS code, some variables regarding the subjects’ demography (sex, age, date of birth) are loaded from Oracle and saved into DM.sas7bdat dataset.

SECOND STEP: FROM SAS TO XML
The second step of the process focuses on converting the SAS dataset into an .XML file; to do this, the macro %sas2xml was developed. This macro needs three input parameters:
• ds => name of the SAS dataset to be converted;
• tags => specify the names of the tags separated by a ‘/’.
• path => path in which the output files will be saved.

The code presented above (lines 2-38), let the user build a support file (DM.map). This file is created to maintain the structure of the input dataset (DM.sas7bdat) in the .XML output. A block of this .MAP file is presented below. Between <COLUMN> and </COLUMN>, each variable of the input dataset is described and its own characteristics defined:

<table>
<thead>
<tr>
<th>COLUMN</th>
<th>TAGS</th>
<th>DATA TYPE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>PATIENT</td>
<td>&lt;root tag&gt;</td>
<td>STRING</td>
<td>patient’s name</td>
</tr>
<tr>
<td>AGE</td>
<td>&lt;CHILD tag&gt;</td>
<td>NUMBER</td>
<td>age of the patient</td>
</tr>
<tr>
<td>DATEOFBIRTH</td>
<td>&lt;CHILD tag&gt;</td>
<td>DATE</td>
<td>date of birth</td>
</tr>
</tbody>
</table>

The resulting SAS dataset is shown below. It contains all the variables specified in the PROC SQL, formatted as defined by the user.

<table>
<thead>
<tr>
<th>OBSERVATION</th>
<th>PATIENT</th>
<th>SEX</th>
<th>AGE</th>
<th>DATEOFBIRTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>John Doe</td>
<td>M</td>
<td>35</td>
<td>01/01/1987</td>
</tr>
<tr>
<td>2</td>
<td>Jane Doe</td>
<td>F</td>
<td>40</td>
<td>02/02/1982</td>
</tr>
<tr>
<td>3</td>
<td>Michael Smith</td>
<td>M</td>
<td>30</td>
<td>10/10/1992</td>
</tr>
</tbody>
</table>

Through the statements from line 40 to 46 of the %sas2xml macro, the .XML file is created.

CONCLUSION
A sponsor’s request sometimes can represent a starting point for the development of new processes. In this case we created a macro system that allowed us to export data and send by e-mail to the sponsor only running a SAS program. Furthermore, by scheduling the program the user is not required to run the program themselves!

THIRD STEP: AUTOMATIC SENDING OF THE .XML FILE THROUGH SAS
Finally the .XML file is ready for the delivery. In order to save time, avoiding repetitive operations such as the preparation of the e-mail for the sponsor, including attachment of the file, specification of the addresses and the text, further SAS steps were implemented.

SCHEDULING OF SAS PROCEDURES
The Sponsor specifically request the production and the delivery of the .XML file daily. To achieve this, the daily scheduling of the programs presented in this poster on a SAS server was implemented; saving time and avoiding the daily manually running of the programs.

CONCLUSION
Your comments and questions are valued and encouraged. Contact the authors at Matteo Ferrari & Valeria Visonà
CROS NT S.r.l.
Via Germania 2, Verona 37136
Office Phone: 0039 0458202666 Fax: 0039 0458205875
Emails: matteo.ferrari@cros.it valeria.visona@cros.it
Web: www.cros.it
Brand and product names are trademarks of their respective companies.