AUTOMATED SDTM CREATION AND DISCREPANCY DETECTION JOBS: THE NUMBERS TELL THE TALE

Joris De Bondt
PhUSE Conference 2014
12-15 Oct 2014
OUTLINE

- Setting the scene
  - SGS’ Conversion Center
  - SDTM data sources
  - Timing for CDMS data refresh

- SGS’ Conversion Center
  - Glossary
  - Technologies used
  - What is it?
  - Code snippets
  - Performance tuning
A central tool to manage the end-to-end flow of refreshing the Data Management study databases

- **REQUEST:** sends an automatic data request via:
  - Web browser, on a variety of vendor specific platforms
  - SAS Connect
to obtain a new extract of the data
  (unless scheduled transfer is available)

- **DOWNLOAD:** downloads new data extract from:
  - Web browser, on a variety of vendor specific platforms
  - SAS Connect
  - sFTP server
• **INSTALL**: loads new data extract (Oracle dmp, CSV, SAS, excel, etc.) into the Data Management database

• **CONVERT**: converts all current study data into SDTM format

• **METADATA**: generates new metadata files for the SDTM data

• **REPORTS**: generates reports for external distribution

• **PAGES2LOCK**: generates overview of eCRF pages that meet locking criteria

• **A2O**: refreshes output of checks and listings

• **SAS**: creates SAS datasets from the SDTM data

• **CLINTRIAL**: starts the synchronisation of the SDTM database to Clintrial
SETTING THE SCENE: SDTM DATA SOURCES

eCRF

Paper CRF

Site specific systems

eSource

ePRO

Lab, imaging, etc.
• Refreshing
  • the SDTM database
  • the output of checks and listings

impacts the work of Data Management (DM), so
  • the best time is **overnight**, when DM staff is not at work
  • but also DM Programmers are not around

meaning that the entire end-to-end flow should run **unattended**
Windows Powershell

- Microsoft's task automation and configuration management framework
- Consists of a command-line shell and associated scripting language built on .NET Framework
- Administrative tasks are generally performed by cmdlets (pronounced command-lets), which are specialized .NET classes implementing a particular operation
- Sets of cmdlets may be combined into scripts, executables (which are standalone applications).

jQuery

- cross-platform JavaScript library designed to simplify the client-side scripting of HTML
- free, open source software, licensed under the MIT License
- jQuery's syntax is designed to make it easier to navigate a document, select DOM elements, create animations, handle events, and develop Ajax applications.

VMWare

- provides cloud and virtualization software and services
TECHNOLOGIES USED

- VMWare, Powershell
- VMWare, Powershell
- VMWare, Powershell, Oracle PL/SQL
- VMWare, Powershell, Oracle PL/SQL
- VMWare, Powershell, Oracle PL/SQL
- VMWare, Powershell, Oracle PL/SQL, Crystal Reports
- VMWare, Powershell, Oracle PL/SQL
- VMWare, Powershell, Oracle PL/SQL
- VMWare, Powershell, Oracle PL/SQL, SAS
- VMWare, Powershell, Oracle PL/SQL
- VMWare, Powershell, Oracle PL/SQL
## SGS’ CONVERSION CENTER

**WHAT IS IT?**

- All end-to-end study process information one-click-away in a single overview

### EDC CONVERSION OVERVIEW DMPROD

<table>
<thead>
<tr>
<th>PROTOCOL</th>
<th>TRIALID</th>
<th>SOURCE TYPE</th>
<th>STATUS</th>
<th>PCONVSPRC</th>
<th>LIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1328745</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>LSHU</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1328379</td>
<td>LAMPAS</td>
<td>FINISHED</td>
<td>HKOU</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1328372</td>
<td>LIQIWAY</td>
<td>FINAL_FINISHED</td>
<td>SYRO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102145</td>
<td>LAMPAS</td>
<td>FINISHED</td>
<td>SYRO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102148</td>
<td>LAMPAS</td>
<td>FINISHED</td>
<td>SYRO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102010</td>
<td>LAMPAS</td>
<td>FINISHED</td>
<td>LSHU</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102426</td>
<td>LAMPAS</td>
<td>FINISHED</td>
<td>JSRO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102058</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>RRAI</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102114</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>RRAI</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1331079</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>KAFT</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102763</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>LSHU</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102764</td>
<td>LAMPAS</td>
<td>FINISHED</td>
<td>KLOO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102130</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>SYRO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102071</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>SYRO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102864</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>JAVR</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102865</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>KLOO</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102260</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>JAVR</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td>1102549</td>
<td>LAMPAS</td>
<td>FINAL_FINISHED</td>
<td>LSHU</td>
<td>Y</td>
</tr>
</tbody>
</table>
SGS' CONVERSION CENTER
WHAT IS IT?

**History information**

<table>
<thead>
<tr>
<th>CONVID</th>
<th>START_DATE</th>
<th>DAY</th>
<th>DUR</th>
<th>AZO_SYNC</th>
<th>PACK</th>
<th>DUMPPILE</th>
<th>COMM</th>
<th>ERROR</th>
<th>FLOW</th>
<th>STATUS</th>
<th>INSERTS</th>
<th>UPDATES</th>
<th>REPO</th>
</tr>
</thead>
<tbody>
<tr>
<td>90144</td>
<td>24-AUG-14</td>
<td>SUN</td>
<td>05:18</td>
<td>25-AUG-14 04:44</td>
<td>V4.9 MCD</td>
<td>20140824</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339750</td>
<td>10668575</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>90096</td>
<td>23-AUG-14</td>
<td>SAT</td>
<td>05:27</td>
<td>25-AUG-14 04:44</td>
<td>V4.9 MCD</td>
<td>20140824</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339768</td>
<td>10668664</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>90052</td>
<td>22-AUG-14</td>
<td>FRI</td>
<td>06:44</td>
<td>25-AUG-14 04:44</td>
<td>V4.9 MCD</td>
<td>20140824</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339768</td>
<td>10668664</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>89703</td>
<td>17-AUG-14</td>
<td>SUN</td>
<td>05:38</td>
<td>18-AUG-14 05:04</td>
<td>V4.9 MCD</td>
<td>20140817</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339768</td>
<td>10668664</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>89298</td>
<td>10-AUG-14</td>
<td>SUN</td>
<td>05:27</td>
<td>11-AUG-14 04:53</td>
<td>V4.9 MCD</td>
<td>20140810</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339766</td>
<td>10668558</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>88922</td>
<td>03-AUG-14</td>
<td>SUN</td>
<td>05:16</td>
<td>04-AUG-14 04:42</td>
<td>V4.9 MCD</td>
<td>20140803</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339766</td>
<td>10668558</td>
<td>V3 RRAD</td>
</tr>
</tbody>
</table>

**Error information**

<table>
<thead>
<tr>
<th>CONVID</th>
<th>START_DATE</th>
<th>DAY</th>
<th>DUR</th>
<th>AZO_SYNC</th>
<th>PACK</th>
<th>DUMPPILE</th>
<th>COMM</th>
<th>ERROR</th>
<th>FLOW</th>
<th>STATUS</th>
<th>INSERTS</th>
<th>UPDATES</th>
<th>REPO</th>
</tr>
</thead>
<tbody>
<tr>
<td>90144</td>
<td>24-AUG-14</td>
<td>SUN</td>
<td>05:18</td>
<td>25-AUG-14 04:44</td>
<td>V4.9 MCD</td>
<td>20140824</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339750</td>
<td>10668575</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>90096</td>
<td>23-AUG-14</td>
<td>SAT</td>
<td>05:27</td>
<td>25-AUG-14 04:44</td>
<td>V4.9 MCD</td>
<td>20140824</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339768</td>
<td>10668664</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>90052</td>
<td>22-AUG-14</td>
<td>FRI</td>
<td>06:44</td>
<td>25-AUG-14 04:44</td>
<td>V4.9 MCD</td>
<td>20140824</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339768</td>
<td>10668664</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>89703</td>
<td>17-AUG-14</td>
<td>SUN</td>
<td>05:38</td>
<td>18-AUG-14 05:04</td>
<td>V4.9 MCD</td>
<td>20140817</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339768</td>
<td>10668664</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>89298</td>
<td>10-AUG-14</td>
<td>SUN</td>
<td>05:27</td>
<td>11-AUG-14 04:53</td>
<td>V4.9 MCD</td>
<td>20140810</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339766</td>
<td>10668558</td>
<td>V3 RRAD</td>
</tr>
<tr>
<td>88922</td>
<td>03-AUG-14</td>
<td>SUN</td>
<td>05:16</td>
<td>04-AUG-14 04:42</td>
<td>V4.9 MCD</td>
<td>20140803</td>
<td>1</td>
<td></td>
<td></td>
<td>FINAL_FINISHED</td>
<td>339766</td>
<td>10668558</td>
<td>V3 RRAD</td>
</tr>
</tbody>
</table>
SGS’ CONVERSION CENTER
WHAT IS IT?

- Performance statistics, for detecting problems and possible improvements

![Graph showing performance statistics over time](image.png)

Performance tuning
Automatic ‘On demand’ request of Rave SAS datasets

- Via “InternetExplorer.Application” object
- Setting the URL
- Adding username and password
Customized SGS tooltip enablement to allow for a web 2.0 experience

```javascript
$(document).ready(function() {
    getInfoTrials('<?php echo $vInstance; ?>');
    getSelections();
    convspecMode();
    checkRemote();
    createColumnFilter();

    $(function()
        .tipTip({maxWidth: "600px"});
    $(".flow").tipTip({edgeOffset: 20});
    $(".colorfilter").tipTip({edgeOffset: 20});
});
```

CONVERSION HISTORY FOR _SDS_
Processing time for small phase I trials ⇔ large phase III trials

- From minutes to hours
- When to start tuning efforts?

Phase I

| CONVERSION | start_date: 2014-08-26 05:07 | end_date: 2014-08-26 05:12 | duration: 00:05:15 | status: FINISHED | error: |

Phase IIB

| CONVERSION | start_date: 2014-08-26 03:48 | end_date: 2014-08-26 05:25 | duration: 01:36:08 | status: FINISHED | error: |

Phase III

| CONVERSION | start_date: 2014-08-25 00:58 | end_date: 2014-08-25 04:29 | duration: 03:31:13 | status: FINISHED | error: |
SGS’ CONVERSION CENTER
PERFORMANCE TUNING

Based on overall cross-study comparison

REQUEST | DOWN LOAD | INSTALL | CONVERT | META DATA | REPORTS | PAGES 2 LOCK | A2O | SAS | CLIN TRIAL

- REQUEST
- DOWN LOAD
- INSTALL
- CONVERT
- META DATA
- REPORTS
- PAGES 2 LOCK
- A2O
- SAS
- CLIN TRIAL

Duration

- 14hrs
- 12hrs
- 10hrs
- 8hrs
- 6hrs
- 4hrs
- 2hrs
- 0hrs

Duration

- 14hrs
- 12hrs
- 10hrs
- 8hrs
- 6hrs
- 4hrs
- 2hrs
- 0hrs

Data

SGS Conference, 12-15 Oct 2014
Based on overall timing evolution on study level
SGS’ CONVERSION CENTER

PERFORMANCE TUNING

- Based on **task-timing** evolution on study level

  ➢ “Expected Time of Arrival”

<table>
<thead>
<tr>
<th>DOMAIN</th>
<th>START_DATE</th>
<th>END_DATE</th>
<th>DURATION</th>
<th>STATUS</th>
<th>PROCESS</th>
<th>EXEC_ERROR</th>
<th>CATEGORY</th>
<th>ETA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNTRIAL</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:03:21</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>CNTRIAL</td>
<td>+00:00:20</td>
</tr>
<tr>
<td>A2O_SYNC</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:04:48</td>
<td>FINISHED</td>
<td>A2O</td>
<td></td>
<td>+00:00:53</td>
<td></td>
</tr>
<tr>
<td>REPORTS_PAGES2LOCK</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:00</td>
<td>FINISHED</td>
<td>REPORTS</td>
<td></td>
<td>-00:06:14</td>
<td></td>
</tr>
<tr>
<td>CHECKS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:01:19</td>
<td>FINISHED</td>
<td>CHECKS</td>
<td></td>
<td>+00:00:22</td>
<td></td>
</tr>
<tr>
<td>REPORTS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:09:12</td>
<td>FINISHED</td>
<td>REPORTS</td>
<td></td>
<td>+00:00:09</td>
<td></td>
</tr>
<tr>
<td>METADATA</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:00</td>
<td>FINISHED</td>
<td>METADATA</td>
<td></td>
<td>-00:06:01</td>
<td></td>
</tr>
<tr>
<td>AFTERCONV</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:02:50</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>+00:00:13</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:01:54</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-03:09:14</td>
<td></td>
</tr>
<tr>
<td>RELREC</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:10:45</td>
<td>FINISHED</td>
<td>RELREC</td>
<td></td>
<td>-00:34:26</td>
<td></td>
</tr>
<tr>
<td>DELETENRAPTOMSUBJECTS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:01:04</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:07</td>
<td></td>
</tr>
<tr>
<td>TRIALDESIGN</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:02</td>
<td>FINISHED</td>
<td>TRIALDESIGN</td>
<td></td>
<td>-00:06:03</td>
<td></td>
</tr>
<tr>
<td>DV</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:38</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:07:59</td>
<td></td>
</tr>
<tr>
<td>DELETEFANOMVISITS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:01:06</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:01</td>
<td></td>
</tr>
<tr>
<td>REPEATVISITS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:05:26</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>+00:00:22</td>
<td></td>
</tr>
<tr>
<td>RASHIVISITS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:17</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>+00:00:54</td>
<td></td>
</tr>
<tr>
<td>DROPOUTVISITS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:00</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:00</td>
<td></td>
</tr>
<tr>
<td>UNSVISITS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:22</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>+00:00:04</td>
<td></td>
</tr>
<tr>
<td>CE</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:29</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:02</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:22</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>+00:00:01</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:13</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:03</td>
<td></td>
</tr>
<tr>
<td>DV</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:04</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:08:34</td>
<td></td>
</tr>
<tr>
<td>DV</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:14</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:08:24</td>
<td></td>
</tr>
<tr>
<td>XS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:19</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:00</td>
<td></td>
</tr>
<tr>
<td>XS</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:16</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:02</td>
<td></td>
</tr>
<tr>
<td>XA</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:01:34</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>-00:06:04</td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>2014-08-25</td>
<td>2014-08-25</td>
<td>00:00:03</td>
<td>FINISHED</td>
<td></td>
<td></td>
<td>+00:00:43</td>
<td></td>
</tr>
</tbody>
</table>
SGS’ CONVERSION CENTER
PERFORMANCE TUNING

- Allowing for targeted performance tuning
  - Data download
    - DOWNLOAD
      start_date: 2014-04-06 23:30
      end_date: 2014-04-07 00:27
      duration: 00:57:45
      status: FINISHED
      error:
    - DOWNLOAD
      start_date: 2014-04-15 23:30
      end_date: 2014-04-15 23:56
      duration: 00:26:46
      status: FINISHED
      error:
  - SDTM conversion
    - CONVERSION
      start_date: 2014-04-07 01:40
      end_date: 2014-04-07 06:08
      duration: 04:28:08
      status: FINISHED
      error:
    - CONVERSION
      start_date: 2014-04-25 12:25
      end_date: 2014-04-25 16:03
      duration: 03:37:27
      status: FINISHED
      error:
TAKE HOME MESSAGES

- Data Management centers
  - face wide variety of
    - source data
    - data provider platforms
  - process small phase I to large phase III trials
  - require minimal impact of data refresh for DM staff

- IT solutions can be put in place to run a sequence of DM actions unattended overnight
  - allowing for DM Programmers to
    - Track unexpected errors
    - Tune the process performance
    - Manage all trials with minimal oversight effort