Leveraging XML Transformations to Get Data into SAS, R, CSV, and Other Formats

Linas Silva
Centre for Biostatistics and Clinical Trials
Peter MacCallum Cancer Centre
Goals

- Demonstrate XML transformations are a viable way to export OpenClinica data to any system
- Encourage further development
  - Support for other formats/systems (e.g. SPSS, STATA, MS SQL)
  - More seamless integration
Agenda

- Background on how the XML transformations were developed
- Describe how OpenClinica data is structured
- 3 powerful XML transformation commands
- 4 main transformation steps common to the 3 export formats
- Results for CSV, R and SAS
Background

- Peter MacCallum Cancer Centre, Australia’s only public hospital solely dedicated to cancer treatment, research and education

- Centre for Biostatistics and Clinical Trials
  - Paper based clinical trials and MS Access
  - May 2010 - began implementing OpenClinica
  - XML transformations to MS Access
  - XML transformations to SAS, R and CSV
SAS, R and CSV xsl Transformations

Some excellent xsl transform files were made available by Linas Silva on this developers list thread (also available here). These transformations can be run on full ODM 1.3 xml extracts, with the accompanying powershell scripts. The CSV powershell script is probably the most important as it chops up the output csv into into a series of csvs.

Alternatively, the transforms can be run without powershell, instead using the copy of the saxon library that comes with OpenClinica. Example syntax (all on one line):

```
"path_to_java.exe" -cp "path_to_tomcat\webapps\OpenClinica3141\WEB-INF\lib\saxon-9.7.jar" net.sf.saxon.Transform -o "output_file_path.ext" "input_xml_file.xml" "path_to_xsl.xsl"
```

Each of the xsl files refer to renaming map stylesheet called 'xml_convert_dynamic_lookup' which lists the CRF and Item Group name combinations and what they should be renamed to in the output. If this is not used, the dataset names default to the Item Group OID.
Reference – Issue Tracker
Transform what?

Format to transform
Sample CDISC ODM XML file

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
     xmlns:OpenClinica="http://www.openclinica.org/ns/odm_ext_v130/v3.1"
     xmlns:OpenClinicaRules="http://www.openclinica.org/ns/rules/v3.1"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 OpenClinica-ODM1.3-0-OC2-0.xsd">
  <Study OID="S_OC14">
    <GlobalVariables>
      <StudyName>OC14</StudyName>
      <StudyDescription>OC14</StudyDescription>
      <ProtocolName>OC14</ProtocolName>
    </GlobalVariables>
    <BasicDefinitions>
      <MeasurementUnit OID="MU_G/L" Name="g/L">
        <Symbol>
          <TranslatedText>g/L</TranslatedText>
        </Symbol>
      </MeasurementUnit>
      <MeasurementUnit OID="MU_MM_5672423" Name="mm">
        <Symbol>
          <TranslatedText>mm</TranslatedText>
        </Symbol>
      </MeasurementUnit>
      <MeasurementUnit OID="MU_X10^9/L" Name="x10^9/L">
        <Symbol>
          <TranslatedText>x10^9/L</TranslatedText>
        </Symbol>
      </MeasurementUnit>
    </BasicDefinitions>
    <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
      <Protocol>
        <StudyEventRef StudyEventOID="SE_BASELINE_6245" OrderNumber="1" Mandatory="Yes" />
        <StudyEventRef StudyEventOID="SE_CYCLE1_9588" OrderNumber="2" Mandatory="Yes" />
        <StudyEventRef StudyEventOID="SE_CYCLE2_9987" OrderNumber="3" Mandatory="Yes" />
        <StudyEventRef StudyEventOID="SE_OFFSTUDY_2598" OrderNumber="4" Mandatory="Yes" />
      </Protocol>
    </MetaDataVersion>
  </Study>
</ODM>
```
Sample CRF – how many data tables?

OC14 Tumour Response 0.1

CRF Header Info
Click the flag icon next to an input to enter/view discrepancy notes. Please note that you can only save the notes if CRF data entry has already started.

<table>
<thead>
<tr>
<th>Tumour ... (10/11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title: Tumour Response</td>
</tr>
</tbody>
</table>

**Target Lesions**

<table>
<thead>
<tr>
<th>Site/Organ</th>
<th>Longest Diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neck</td>
<td>12 (mm)</td>
</tr>
<tr>
<td>Shoulder UL</td>
<td>10 (mm)</td>
</tr>
</tbody>
</table>

**Non-Target Lesions**

<table>
<thead>
<tr>
<th>Site/Organ</th>
<th>Number of Lesions per organ</th>
<th>Multiple Lesions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

**Date of Assessment**

24-Mar-2014

**Target Lesions Response**

- CR

**Non-Target Lesions Response**

- not CR and not PD

**Overall Response**

- PR
Sample CRF – 3 data tables

**OC14 Tumour Response 0.1**

**CRF Header Info**

Click the flag icon next to an input to enter/view discrepancy notes. Please note that you can only save the notes if CRF data entry has already started.

### Target Lesion Data Table

<table>
<thead>
<tr>
<th>Site/Organ</th>
<th>Longest Diameter (mm)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Neck</td>
<td>12</td>
<td>X</td>
</tr>
<tr>
<td>Shoulder UL</td>
<td>10</td>
<td>X</td>
</tr>
</tbody>
</table>

### Non Target Lesion Data Table

<table>
<thead>
<tr>
<th>Site/Organ</th>
<th>Number of Lesions per Organ</th>
<th>Multiple Lesions</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Lung</td>
<td>12</td>
<td>✗</td>
<td></td>
</tr>
</tbody>
</table>

### Response Data Table

- **Date of Assessment**: 24-Mar-2014
- **Target Lesions Response**: CR
- **Non-Target Lesions Response**: not CR and not PD
- **Overall Response**: PR
Study Events (Screen)

<table>
<thead>
<tr>
<th>Study Subject ID</th>
<th>Baseline</th>
<th>Cycle 1</th>
<th>Cycle 2</th>
<th>Off Study</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANT</td>
<td></td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Actions" /></td>
</tr>
<tr>
<td>BAT</td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Status" /></td>
<td><img src="image.png" alt="Actions" /></td>
</tr>
</tbody>
</table>

4 Study Events: Baseline, Cycle 1, Cycle 2 and Off Study
Study Events and Forms (XML)

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
     xmlns:OpenClinica="http://www.openclinica.org/ns/odm_ext_v130/v3.1"
     xmlns:OpenClinicaRules="http://www.openclinica.org/ns/rules/v3.1"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     FileOID="OC14_Full_ExtractD20140509150942+1000" Description="OC14 Full Extract" CreationDateTime="2014-05-09T15:09:42+10:00" FileType="Snapshot" ODMVersion="1.3"
     xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 OpenClinica-ODM1-3-0-OC2-0.xsd">
  <Study OID="S_OC14">
    + <GlobalVariables/>
    + <BasicDefinitions>
      - <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
        + <Protocol>
          + <StudyEventDef OID="SE_BASELINE_6245" Name="Baseline" Repeating="No" Type="Scheduled">
          + <StudyEventDef OID="SE_CYCLE1_9588" Name="Cycle 1" Repeating="No" Type="Scheduled">
          + <StudyEventDef OID="SE_CYCLE2_9987" Name="Cycle 2" Repeating="No" Type="Scheduled">
          + <StudyEventDef OID="SE_OFFSTUDY_2598" Name="Off Study" Repeating="No" Type="Scheduled">
            + <FormDef OID="F_OC14HAEMATOL_01" Name="OC14 Haematology - 0.1" Repeating="No">
              + <FormDef OID="F_OC14TUMOURRE_9904_01" Name="OC14 Tumour Response - 0.1" Repeating="No">
                + <FormDef OID="F_OC14OFFSTUDY_01" Name="OC14 Off Study - 0.1" Repeating="No">
```

4 Study Events

3 Forms
Cycle 1 Study Event uses Haematology and Tumour Response Forms only
Study Event Forms (XML)

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
     xmlns:OpenClinica="http://www.openclinica.org/ns/odm_ext_v130/v3.1"
     xmlns:OpenClinicaRules="http://www.openclinica.org/ns/rules/v3.1"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     FileOID="OC14_Full_ExtractD20140509150942+1000" Description="OC14 Full Extract" CreationDateTime="2014-05-09T15:09:42+10:00" FileType="Snapshot" ODMVersion="1.3"
     xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.3 OpenClinica-ODM1-3-0-OC2-0.xsd">
  <Study OID="S_OC14">
    <GlobalVariables/>
    <BasicDefinitions/>
    <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0"/>
    <Protocol>
      <StudyEventDef OID="SE_BASELINE_6245" Name="Baseline" Repeating="No" Type="Scheduled">
        <FormRef FormOID="F_OC14HAEMATOL_01" Mandatory="Yes"/>
        <FormRef FormOID="F_OC14TUMOURRE_9904_01" Mandatory="Yes"/>
      </StudyEventDef>
      <StudyEventDef OID="SE_CYCLE1_9588" Name="Cycle 1" Repeating="No" Type="Scheduled">
        <FormRef FormOID="F_OC14HAEMATOL_01" Mandatory="Yes"/>
        <FormRef FormOID="F_OC14TUMOURRE_9904_01" Mandatory="Yes"/>
      </StudyEventDef>
      <StudyEventDef OID="SE_CYCLE2_9987" Name="Cycle 2" Repeating="No" Type="Scheduled">
        <FormDef OID="F_OC14HAEMATOL_01" Name="OC14 Haematology - 0.1" Repeating="No"/>
        <FormDef OID="F_OC14TUMOURRE_9904_01" Name="OC14 Tumour Response - 0.1" Repeating="No"/>
        <FormDef OID="F_OC14OFFSTUDY_01" Name="OC14 Off Study - 0.1" Repeating="No"/>
      </StudyEventDef>
      <StudyEventDef OID="SE_OFFSTUDY_2598" Name="Off Study" Repeating="No" Type="Scheduled">
        <FormDef OID="F_OC14HAEMATOL_01" Name="OC14 Haematology - 0.1" Repeating="No"/>
        <FormDef OID="F_OC14TUMOURRE_9904_01" Name="OC14 Tumour Response - 0.1" Repeating="No"/>
        <FormDef OID="F_OC14OFFSTUDY_01" Name="OC14 Off Study - 0.1" Repeating="No"/>
      </StudyEventDef>
    </Protocol>
  </Study>
</ODM>
```

Cycle 1 Study Event uses Haematology and Tumour Response Forms only
ItemGroups (Screen)

OC14 Tumour Response 0.1

CRF Header Info
Click the flag icon next to an input to enter/view discrepancy notes. Please note that you can only save the notes if CRF data entry has already started.

Target lesion ItemGroup

Non target lesion ItemGroup

Response ItemGroup
ItemGroups (XML)

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3" />

<Study OID="S_OC14">
  <GlobalVariables>
  </GlobalVariables>
  <BasicDefinitions>
    <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0"/>
  </BasicDefinitions>
  <FormDef OID="F_OC14HAEMATOL_01" Name="OC14 Haematology - 0.1" Repeating="No">
  </FormDef>
  <FormDef OID="F_OC14TUMOURRE_9904_01" Name="OC14 Tumour Response - 0.1" Repeating="No">
  </FormDef>
  <FormDef OID="F_OC14OFFSTUDY_01" Name="OC14 Off Study - 0.1" Repeating="No">
  </FormDef>
  <ItemGroupDef OID="IG_OC14T_TARGETLESIONS_1716" Name="TargetLesions" Repeating="Yes">
    <SASDatasetName>TARGETLE</SASDatasetName> <Comment>Target Lesions</Comment>
  </ItemGroupDef>
  <ItemGroupDef OID="IG_OC14T_NONTARGETLESIONS_4474" Name="NonTargetLesions" Repeating="Yes">
    <SASDatasetName>NONTARGE</SASDatasetName> <Comment>Non-Target Lesions</Comment>
  </ItemGroupDef>
  <ItemGroupDef OID="IG_OC14T_TUMOURRESPONSE_6359" Name="TumourResponse" Repeating="No">
    <SASDatasetName>TUMOURRE</SASDatasetName>
  </ItemGroupDef>
  <ItemGroupDef OID="IG_OC14H_UNGROUPIED" Name="IG_OC14H_UNGROUPIED" Repeating="No">
    <SASDatasetName>UNGROUIPE</SASDatasetName>
  </ItemGroupDef>
  <ItemGroupDef OID="IG_OC14O_UNGROUPIED" Name="IG_OC14O_UNGROUPIED" Repeating="No">
    <SASDatasetName>UNGRO001</SASDatasetName>
  </ItemGroupDef>
</Study>
```

5 ItemGroups
Form ItemGroups (XML)

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
    
    <Study OID="S_OC14">
        <GlobalVariables>
            <BasicDefinitions>
                <MetaDataVersion OID="v1.0.0" Name="MetaDataVersion_v1.0.0">
                    <Protocol>
                        <FormDef OID="F_OC14HAEMATOL_01" Name="OC14 Haematology - 0.1" Repeating="No">
                            <ItemGroupRef ItemGroupOID="IG_OC14T_TARGETLESIONS_1716" Mandatory="No" />
                            <ItemGroupRef ItemGroupOID="IG_OC14T_NONTARGETLESIONS_4474" Mandatory="No" />
                            <ItemGroupRef ItemGroupOID="IG_OC14T_TUMOURRESPONSE_6359" Mandatory="No" />
                        </FormDef>
                        <FormDef OID="F_OC14OFFSTUDY_01" Name="OC14 Off Study - 0.1" Repeating="No">
                            <ItemGroupDef OID="IG_OC14T_TARGETLESIONS_1716" Name="TargetLesions" Repeating="Yes">
                                <SASDatasetName>"TARGETLE" Comment="Target Lesions"</SASDatasetName>
                            </ItemGroupDef>
                            <ItemGroupDef OID="IG_OC14T_NONTARGETLESIONS_4474" Name="NonTargetLesions" Repeating="Yes">
                                <SASDatasetName>"NONTARGE" Comment="Non-Target Lesions"</SASDatasetName>
                            </ItemGroupDef>
                            <ItemGroupDef OID="IG_OC14T_TUMOURRESPONSE_6359" Name="TumourResponse" Repeating="No">
                                <SASDatasetName>"TUMOURRE"</SASDatasetName>
                            </ItemGroupDef>
                            <ItemGroupDef OID="IG_OC14H_UNGROUPED" Name="IG_OC14H_UNGROUPED" Repeating="No">
                                <SASDatasetName>"UNGROUPE"</SASDatasetName>
                            </ItemGroupDef>
                            <ItemGroupDef OID="IG_OC14O_UNGROUPED" Name="IG_OC14O_UNGROUPED" Repeating="No">
                                <SASDatasetName>"UNGRO001"</SASDatasetName>
                            </ItemGroupDef>
                        </FormDef>
                    </Protocol>
                </MetaDataVersion>
            </BasicDefinitions>
        </GlobalVariables>
    </Study>
</ODM>
```

Tumour Response Form contains 3 ItemGroups
ItemGroup items (Screen)

**Target Lesions ItemGroup** contains 2 items “site/organ” and “LongestDiameter”
ItemGroup fields (XML)

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3">
  ...
  <Study OID='S_OC14'>
    + <GlobalVariables>
      + <BasicDefinitions>
        - <MetaDataVersion OID='v1.0.0' Name='MetaDataVersion_v1.0.0'>
          ...
          - <ItemTypeDef OID='IG_OC14T_TARGETLESIONS_1716' Name='TargetLesions' Repeating='Yes'
            SASDatasetName='TARGETLESIONS' Comment='Target Lesions'>
            <ItemRef ItemOID='T_OC14T_TARGETLESIONSITEORGAN' OrderNumber='1' Mandatory='No' />
            <ItemRef ItemOID='T_OC14T_TARGETLESIONDIMETER' OrderNumber='2' Mandatory='No' />
          + <OpenClinica:ItemGroupDetails ItemGroupOID='IG_OC14T_TARGETLESIONS_1716'>
          </ItemTypeDef>
          ...
          + <ItemTypeDef OID='IG_OC14T_NONTARGETLESIONS_4474' Name='NonTargetLesions' Repeating='Yes'
            SASDatasetName='NONTARGET' Comment='Non-Target Lesions'>
            + <ItemGroupDef OID='IG_OC14T_TUMOURRESPONSE_6359' Name='TumourResponse' Repeating='No'
              SASDatasetName='TUMOURRESPONSE'>
              + <ItemGroupDef OID='IG_OC14T_UNGROUPED' Name='IG_OC14T_UNGROUPED' Repeating='No'
                SASDatasetName='UNGROUPED'>
                + <ItemGroupDef OID='IG_OC14T_UNGROUPED' Name='IG_OC14T_UNGROUPED' Repeating='No'
                  SASDatasetName='UNGROUPED'>
                  + <ItemRef ItemOID='T_OC14T_TARGETLESIONSITEORGAN' OrderNumber='1' Mandatory='No' />
                  + <ItemRef ItemOID='T_OC14T_TARGETLESIONDIMETER' OrderNumber='2' Mandatory='No' />
                  + <ItemRef ItemOID='T_OC14T_TARGETLESIONSITEORGAN' OrderNumber='1' Mandatory='No' />
                  + <ItemRef ItemOID='T_OC14T_TARGETLESIONDIMETER' OrderNumber='2' Mandatory='No' />
                </ItemTypeDef>
              </ItemRef>
            </ItemTypeDef>
          </ItemTypeDef>
          + <ItemTypeDef OID='IG_OC14T_NONTARGETLESIONSITEORGAN' Name='NonTargetLesionSiteOrgan' DataTy
```
Values for “Site/Organ” and “Longest Diameter” items
Item Data XML

```xml
<?xml version="1.0" encoding="US-ASCII" ?>
<ODM xmlns="http://www.cdisc.org/ns/odm/v1.3"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    <Study OID="S_OC14" />
    <AdminData StudyOID="S_OC14" />
    <ClinicalData StudyOID="S_OC14" MetaDataVersionID="v1.0.0" />
    <SubjectData SubjectKey="ISS_ANT" OpenClinica:StudySubjectID="ANT" OpenClinica:Status="available" />
    <StudyEventData StudyEventOID="SE_BASELINE_6245" OpenClinica:StartDate="2014-03-24" OpenClinica:EndDate="2014-03-24" OpenClinica:Status="data entry started" />
    <StudyEventData StudyEventOID="SE_CYCLE1_9588" OpenClinica:StartDate="2014-03-24" OpenClinica:Status="data entry started" />
    <FormData FormOID="F_OC14HAEMATOL_01" OpenClinica:Version="0.1" OpenClinica:Status="initial data entry" />
    <FormData FormOID="F_OC14TUMOURRE_9904_01" OpenClinica:Version="0.1" OpenClinica:Status="initial data entry" />
    <ItemGroupData ItemGroupOID="IG_OC14T_TARGET LESIONS_1716" ItemGroupRepeatKey="1"
                    TransactionType="Insert">
        <ItemData ItemOID="I_OC14T_TARGET LESIONS SITE ORGAN" Value="Neck">
            <OpenClinica:AuditLogs EntityID="I_OC14T_TARGET LESIONS SITE ORGAN" />
        </ItemData>
        <ItemData ItemOID="I_OC14T_TARGET LESION DIAMETER" Value="12">
            <MeasurementUnitRef MeasurementUnitOID="MU_MM_5672423" />
            <OpenClinica:AuditLogs EntityID="I_OC14T_TARGET LESION DIAMETER" />
        </ItemData>
    </ItemGroupData>
    <ItemGroupData ItemGroupOID="IG_OC14T_TARGET LESIONS_1716" ItemGroupRepeatKey="2"
                    TransactionType="Insert">
```

Values for “Site/Organ” and “Longest Diameter” items
CDISC ODM XML – data tables?
CDISC ODM XML – data tables
3 Transformation commands

- **VALUE OF**
  - `<xsl:value-of select=……>`
    - **Statement:** `<xsl:value-of select="/odm:ODM/odm:Study/odm:MetaDataVersion/odm:StudyEventDef/@Name">`
    - **Data:** Baseline, Cycle 1, Cycle 2, Off Study
    - **Result:** Baseline (first item found is selected)
### 3 Transformation commands

- **FOR EACH**
  - `<xsl:for-each select="……">`
    - **Statement:** `<xsl:for-each select="/odm:ODM/odm:Study/odm:MetaDataVersion/odm:StudyEventDef">`
    - **Data:** Baseline, Cycle 1, Cycle 2, Off Study
    - **Result:** Baseline, Cycle 1, Cycle 2, Off Study
3 Transformation commands

- **FILTER**
  - `<xsl:..... select=…[element=specific value]…>`
  - **Statement:** `<xsl:for-each select="/odm:ODM/odm:Study/odm:MetaDataVersion/odm:StudyEventDef[@name="Cycle 1"]/odm:FormRef">`
  - **Data:**
    - Baseline: Haematology, Cycle 1: Haematology,
    - Cycle 1: Tumour Response, Cycle 2: Haematology,
    - Cycle 2: Tumour Response, Off Study: Off Study
  - **Result:**
    - Cycle 1: Haematology, Cycle 1: Tumour Response
4 Transformation Steps

- For each ItemGroup definition
  - For each record for the current ItemGroup
    - For each column for the current ItemGroup
      - Find the data value for the current record and column
4 Transformation steps

1. For each ItemGroup

2. For each Record

3. For each column

4. Get the value

Data Values
CSV transformation

```xml
<?xml version="1.0" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  
  
  <xsl:for-each select="odm:ODM/odm:Study/odm:MetaDataVersion/odm:ItemGroupDef">
    
  
  <xsl:for-each select="/odm:ODM/odm:ClinicalData/odm:SubjectData/odm:StudyEventData/odm:FormData/odm:ItemGroupData[@ItemGroupOID=$ItemGroupOID]">
    
  <xsl:variable name="vSubjectKey">
    <xsl:value-of select="../../../../../@SubjectKey" />
  </xsl:variable>
    
  <xsl:for-each select="/odm:ODM/odm:Study/odm:MetaDataVersion/odm:ItemGroupDef[@OID=$ItemGroupOID]"/>
  
  <xsl:variable name="vItemOID">
    <xsl:value-of select="@ItemOID" />
  </xsl:variable>
  
  <xsl:value-of select="/odm:ODM/odm:ClinicalData[@StudyOID=$vStudyOID]/odm:SubjectData[@SubjectKey=$vSubjectKey]/odm:StudyEventData[@StudyEventOID=$vStudyEventOID and (@StudyEventRepeatKey=$vStudyEventRepeatKey or not (@StudyEventRepeatKey))]/odm:FormData[@FormOID=$vFormOID]/odm:ItemGroupData[@ItemGroupOID=$ItemGroupOID and (@ItemGroupRepeatKey=$vItemGroupRepeatKey or not(@ItemGroupRepeatKey))]/odm:ItemData[@ItemOID=$vItemOID]/@Value"/>
```
R Transformation

```xml
<?xml version="1.0" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

    <xsl:for-each select="odm:ODM/odm:Study/odm:MetaDataVersion/odm:ItemGroupDef">
        ...
    </xsl:for-each>

    <xsl:for-each select="/odm:ODM/odm:Study/odm:MetaDataVersion/odm:ItemGroupDef[@OID=$ItemGroupOID]/odm:ItemRef">
        ...
    </xsl:for-each>

    <xsl:for-each select="/odm:ODM/odm:ClinicalData/odm:SubjectData/odm:StudyEventData/odm:FormData/odm:ItemGroupData[@ItemGroupOID=$ItemGroupOID]">
        ...
    </xsl:for-each>

    <xsl:value-of select="odm:ItemData[@ItemOID=$vItemOID]/Value"/>

</xsl:stylesheet>
```
SAS (map) Transformation

```xml
<?xml version='1.0'?>
<xsl:stylesheet version='1.0' xmlns:xsl='http://www.w3.org/1999/XSL/Transform'>
  <xsl:for-each select='odm:ODM/odm:Study/odm:MetaDataSet/odm:ItemGroupDef'>
    ...
  </xsl:for-each>
  <xsl:for-each select='/odm:ODM/odm:Study/odm:MetaDataSet/odm:ItemGroupDef[@OID=$ItemGroupOID]/odm:ItemRef'>
    ...
  </xsl:for-each>
  <xsl:value-of select='/odm:ODM/odm:Study/odm:MetaDataSet/odm:ItemDef[@OID=$vitemOID]/@DataType'/>
</xsl:stylesheet>
```
SAS (data) Transformation

```xml
<?xml version="1.0" ?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
              xmlns:odm="odm:ODM/odm:ClinicalData/odm:SubjectData/odm:StudyEventData/odm:FormData/odm:ItemGroupData">

  ...

  <xsl:for-each select="odm:ItemData">
    ...
  </xsl:for-each>
</xsl:stylesheet>
```
### CSV Results

**TableName:** TumourResponseTargetLesions
- **SubjectID**, **EventName**, **StudyEventRepeatKey**, **CRFName**, **ItemGroupRepeatKey**, **TargetLesionSite**
- **Ant**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, **Neck**, 12
- **Ant**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 2, **Shoulder UL**, 10
- **Ant**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 1, **Neck**, 11.5
- **Ant**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 2, **Shoulder UL**, 9.2
- **Bat**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, **Left axillary node**, 5.5
- **Bat**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 2, **Right external iliac node**, 2.3
- **Bat**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 3, **Portocaval node**, 1.8
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 1, **Left axillary node**, 5.2
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 2, **Right external iliac node**, 1.5
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 3, **Portocaval node**, 0.5

**TableName:** TumourResponseNonTargetLesions
- **SubjectID**, **EventName**, **StudyEventRepeatKey**, **CRFName**, **ItemGroupRepeatKey**, **NonTargetLesionSite**
- **Ant**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, **Lung**, 12
- **Ant**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 1, **Lung**, 10
- **Bat**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, **Left Inguinal Node**, 1
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 2, **Right paratracheal**, 3
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 1, **Left Inguinal Node**, 1
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 2, **Right paratracheal**, 1

**TableName:** TumourResponse
- **SubjectID**, **EventName**, **StudyEventRepeatKey**, **CRFName**, **ItemGroupRepeatKey**, **AssessmentDate**, **TargetLesionResponse**, **NonTargetLesionResponse**, **OverallResponse**
- **Ant**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, 2014-03-24, **PR**, **notCRnorPD**, **PR**
- **Ant**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, 2014-05-09, **CR**, **CR**, **CR**
- **Bat**, **Cycle 1**, **OC14 Tumour Response** - 0.1, 1, 2014-04-28, **PR**, **notCRnorPD**, **PR**
- **Bat**, **Cycle 2**, **OC14 Tumour Response** - 0.1, 1, 2014-05-05, **PR**, **CR**, **PR**

**TableName:** Haematology
- **SubjectID**, **EventName**, **StudyEventRepeatKey**, **CRFName**, **ItemGroupRepeatKey**, **DateTaken**, **Hb**, **Platelets**
- **Ant**, **Baseline**, **OC14 Haematology** - 0.1, 2014-03-24, 70, 200
- **Ant**, **Cycle 1**, **OC14 Haematology** - 0.1, 2014-03-24, 75, 250
- **Ant**, **Cycle 2**, **OC14 Haematology** - 0.1, 2014-05-09, 78, 198
- **Bat**, **Baseline**, **OC14 Haematology** - 0.1, 2014-04-22, 61, 188
- **Bat**, **Cycle 1**, **OC14 Haematology** - 0.1, 2014-04-28, 98, 170
- **Bat**, **Cycle 2**, **OC14 Haematology** - 0.1, 2014-05-05, 68, 193

**TableName:** offStudy
- **SubjectID**, **EventName**, **StudyEventRepeatKey**, **CRFName**, **ItemGroupRepeatKey**, **OffStudy Date**
- **Ant**, **off study**, **OC14 off study** - 0.1, 2014-05-09
R Results

TumourResponseTargetLesions <- data.frame(
  SubjectID=c('ANT', 'ANT', 'ANT', 'ANT', 'BAT', 'BAT', 'BAT', 'BAT', 'BAT', 'BAT'),
  EventName=c('Cycle 1', 'Cycle 1', 'Cycle 2', 'Cycle 2', 'Cycle 1', 'Cycle 1', 'Cycle 1', 'Cycle 2', 'Cycle 2', 'Cycle 2'),
  StudyEventRepeatKey=c(NA, NA, NA, NA, NA, NA, NA, NA, NA, NA),
  CRFName=c('OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1'),
  ItemGroupRepeatKey=c(1,2,1,2,1,2,3,1,2,3),
  TargetLesionsSiteOrgan=c('Neck', 'Shoulder UL', 'Neck', 'Shoulder UL', 'left axillary node', 'Right external iliac node', 'Portocaval node', 'left axillary node', 'Right external iliac node', 'Portocaval node'),
  TargetLesionDiameter=c(12,10,11.5,9.2,5.5,2.3,1.8,5.2,1.5,0.5));

TumourResponseNonTargetLesions <- data.frame(
  SubjectID=c('ANT', 'ANT', 'BAT', 'BAT', 'BAT', 'BAT'),
  EventName=c('Cycle 1', 'Cycle 2', 'Cycle 1', 'Cycle 1', 'Cycle 2', 'Cycle 2'),
  StudyEventRepeatKey=c(NA, NA, NA, NA, NA, NA),
  CRFName=c('OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1'),
  ItemGroupRepeatKey=c(1,1,1,2,1,2),
  NonTargetLesionsSiteOrgan=c('Lung', 'Lung', 'Left Inguinal Node', 'Right paratracheal', 'Left Inguinal Node', 'Right paratracheal'),
  NonTargetLesionCount=c(12,10,NA,3,NA,1),
  NonTargetLesionMultiple=c(NA,NA,NA,1,NA,1,NA));

TumourResponse <- data.frame(SubjectID=c('ANT', 'ANT', 'BAT', 'BAT'),
  EventName=c('Cycle 1', 'Cycle 2', 'Cycle 1', 'Cycle 2'),
  StudyEventRepeatKey=c(NA, NA, NA, NA),
  CRFName=c('OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1', 'OC14 Tumour Response - 0.1'),
  ItemGroupRepeatKey=c(1,1,1,1),
  AssessmentDate=c(as.Date("2014-03-24"), as.Date("2014-05-09"), as.Date("2014-04-28"), as.Date("2014-05-05")),
  TargetLesionResponse=c('CR', 'CR', 'PR', 'PR'),
  NonTargetLesionResponse=c('notCRnorPD', 'CR', 'notCRnorPD', 'CR'),
  OverallResponse=c('PR', 'CR', 'PR', 'PR'));
SAS (map) results

```xml
<?xml version="1.0" encoding="utf-8" ?>
<sxlemap version="1.2">
  <table name="TumourResponseTargetLesions">
    <table-path syntax="XPATH">/OC14/IG_OC14T_TARGETLESIONS_1716</table-path>
    <column name="SubjectID">
      <path>/OC14/IG_OC14T_TARGETLESIONS_1716/SubjectID</path>
      <type>character</type>
      <datatype>string</datatype>
      <length>50</length>
    </column>
    + <column name="StudyEvent"/>
    + <column name="StudyEventRepeatKey"/>
    + <column name="ItemGroupRepeatKey"/>
    - <column name="TargetLesionSiteOrgan">
      <path>/OC14/IG_OC14T_TARGETLESIONS_1716/I_OC14T_TARGETLESIONSITEORGAN</path>
      <type>character</type>
      <datatype>string</datatype>
      <length>25</length>
    </column>
    - <column name="TargetLesionDiameter">
      <path>/OC14/IG_OC14T_TARGETLESIONS_1716/I_OC14T_TARGETLESIONDIAMETER</path>
      <type>numeric</type>
      <datatype>double</datatype>
    </column>
  </table>
  <table name="TumourResponseNonTargetLesions">
    <table-path syntax="XPATH">/OC14/IG_OC14T_NONTARGETLESIONS_4474</table-path>
    + <column name="SubjectID"/>
  </table>
</sxlemap>
```
SAS (data) results

```xml
<?xml version="1.0" encoding="utf-8" ?>
- <OC14>
  + <IG_OC14H_UNGROUPED>
  + <IG_OC14H_UNGROUPED>
  - <IG_OC14T_TARGETLESIONS_1716>
    <SubjectID>ANT</SubjectID>
    <StudyEvent><Cycle 1</StudyEvent>
    <StudyEventRepeatKey />
    <ItemGroupRepeatKey>1</ItemGroupRepeatKey>
    <I_OC14T_TARGETLESIONSITEORGAN>Neck</I_OC14T_TARGETLESIONSITEORGAN>
    <I_OC14T_TARGETLESIONDIAMETER>12</I_OC14T_TARGETLESIONDIAMETER>
  </IG_OC14T_TARGETLESIONS_1716>
  - <IG_OC14T_TARGETLESIONS_1716>
    <SubjectID>ANT</SubjectID>
    <StudyEvent><Cycle 1</StudyEvent>
    <StudyEventRepeatKey />
    <ItemGroupRepeatKey>2</ItemGroupRepeatKey>
    <I_OC14T_TARGETLESIONSITEORGAN>Shoulder UL</I_OC14T_TARGETLESIONSITEORGAN>
    <I_OC14T_TARGETLESIONDIAMETER>10</I_OC14T_TARGETLESIONDIAMETER>
  </IG_OC14T_TARGETLESIONS_1716>
  - <IG_OC14T_NONTARGETLESIONS_4474>
    <SubjectID>ANT</SubjectID>
    <StudyEvent><Cycle 1</StudyEvent>
    <StudyEventRepeatKey />
    <ItemGroupRepeatKey>1</ItemGroupRepeatKey>
    <I_OC14T_NONTARGETLESIONSITEORGAN>Lung</I_OC14T_NONTARGETLESIONSITEORGAN>
    <I_OC14T_NONTARGETLESIONCOUNT>12</I_OC14T_NONTARGETLESIONCOUNT>
  </IG_OC14T_NONTARGETLESIONS_4474>
  - <IG_OC14T_TUMOURRESPONSE_6359>
```
SAS results

The SAS System

<table>
<thead>
<tr>
<th>Directory</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Libref</td>
<td>OC14</td>
</tr>
<tr>
<td>Engine</td>
<td>XML</td>
</tr>
<tr>
<td>Access</td>
<td>READONLY</td>
</tr>
<tr>
<td>Physical Name</td>
<td>OC14</td>
</tr>
<tr>
<td>XMLType</td>
<td>GENERIC</td>
</tr>
<tr>
<td>XMLMap</td>
<td>MAP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Member Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>HAEMATOLOGY</td>
<td>DATA</td>
</tr>
<tr>
<td>2</td>
<td>OFFSTUDY</td>
<td>DATA</td>
</tr>
<tr>
<td>3</td>
<td>TUMOURRESPONSE</td>
<td>DATA</td>
</tr>
<tr>
<td>4</td>
<td>TUMOURRESPONSENONTARGETLESIONS</td>
<td>DATA</td>
</tr>
<tr>
<td>5</td>
<td>TUMOURRESPONSESETARGETLESIONS</td>
<td>DATA</td>
</tr>
</tbody>
</table>

Editor - Untitled1 * PROC DATASETS running

```plaintext
FileName OC14 'H:\OC14 Conference\SAS_OC14_data.xml';
FileName map 'H:\OC14 Conference\SAS_OC14_data_map.map';
libname OC14 xml xmlmap=map access=readonly ;

proc datasets library=OC14;
copy out=work;
run;
```
Code Lists (SAS)

```sas
proc datasets library=OC14;
  copy out=work;
run;

proc format;
  value $CL_4394_ "CR"="CR" "PR"="PR" "PD"="PD" "SD"="SD" "UNK"="UNK" ;
  value $CL_4395_ "CR"="CR" "notCRnorPD"="not CR and not PD" "PD"="PD" "UNK"="UNK" ;
  value $CL_4396_ "CR"="CR" "PR"="PR" "PD"="PD" "SD"="SD" "UNK"="UNK" ;
run;

data TumourResponse;
  set TumourResponse;
  format TargetLesionResponse $CL_4394_.;
run;

data TumourResponse;
  set TumourResponse;
  format NonTargetLesionResponse $CL_4395_.;
run;

data TumourResponse;
  set TumourResponse;
  format OverallResponse $CL_4396_.;
run;
```
Goals - Review

- Demonstrate XML transformations are a viable way to export OpenClinica data to any system

- Encourage further development in the following
  - Support for other formats/systems (e.g. SPSS, STATA, MS SQL)
  - More seamless integration