

The Smart Dataset-XML Viewer: Support for FHIR EHR Records

Document last update: 2018-06-21

Introduction

Already now, electronic health records (EHRs) are an important source of data in clinical research. In such a case, the original, primary source is not the Case Report Form (CRF) anymore, but the EHR, even if the data point is (automatically) copied into the CRF. But even the use of CRFs will in future decrease, as more and more trials will be "remote" trials, in which the data is collected by devices (e.g. wearables) and the data directly send to the data capture system without any form of "form".

In both cases, the [HL7 FHIR standard](#) is very important: it is both used in classic EHR systems as well as a data transport format in devices such as wearables.

In SDTM, traceability to the source is obtained through the "annotated CRF", which is mandated by the FDA to be a PDF file with "pages", even when paper was not used at all, and the eCRFs are essentially "screens". A ridiculous situation indeed: it would be much easier (and be machine-readable) when the traceability is given by a link (e.g. using XPath or XQuery) to the study design in [ODM, which can also be annotated with SDTM information](#).

So what if there is no CRF anymore, a case that will in my opinion in future become the normal case, and the CRF the exception. How can traceability to the source data point than be guaranteed?

The answer is simple: by embedding into the SDTM record.

FHIR resources within SDTM records

The FDA currently still mandates the use of SAS Transport 5 (also named "XPT"), a 30-year old, completely outdated format¹. This format is a fixed-record-length format, [very inefficient](#), not allowing embedding anything other than a value itself, and especially not an EHR record.

The newer, modern format, is Dataset-XML. It was developed by CDISC as a replacement for XPT, and [takes all the disadvantages away of XPT](#). It is based on XML (eXtensible Markup Language) and thus is ... extensible.

So how would a FHIR data point (named "resource" in FHIR) look like when embedded into an SDTM record in Dataset-XML format? Here is an example:

¹ FDA is the only organization in the world still using this (NOT vendor-neutral) format. All other industries have already for a long time wiped SAS Transport 5 (and SAS Transport 8) out.

```

<ItemGroupData data=ItemGroupDataSeq="1" ItemGroupOID="VS">
  <ItemData ItemOID="VS.STUDYID" Value="CDISCPIL0701"/>
  <ItemData ItemOID="VS.DOMAIN" Value="VS"/>
  <ItemData ItemOID="VS.USUBJID" Value="01-701-1015"/>
  <ItemData ItemOID="VS.VSSEQ" Value="1"/>
  <ItemData ItemOID="VS.VSTESTCD" Value="DIABP"/>
  <ItemData ItemOID="VS.VSTEST" Value="Diastolic Blood Pressure"/>
  <ItemData ItemOID="VS.VSPOST" Value="SUPINE"/>
  <ItemData ItemOID="VS.VSORRES" Value="64"/>
  <ItemData ItemOID="VS.VSORRESU" Value="mmHg"/>
  <ItemData ItemOID="VS.VSSTRESC" Value="64"/>
  <ItemData ItemOID="VS.VSSTRESN" Value="64"/>
  <ItemData ItemOID="VS.VSSTRESU" Value="mmHg"/>
  <ItemData ItemOID="VS.VISITNUM" Value="1"/>
  <ItemData ItemOID="VS.VISIT" Value="SCREENING 1"/>
  <ItemData ItemOID="VS.VISITDY" Value="-7"/>
  <ItemData ItemOID="VS.VSDTC" Value="2013-12-26"/>
  <ItemData ItemOID="VS.VSDY" Value="-7"/>
  <ItemData ItemOID="VS.VSTPT" Value="AFTER LYING DOWN FOR 5 MINUTES"/>
  <ItemData ItemOID="VS.VSTPTNUM" Value="815"/>
  <ItemData ItemOID="VS.VSELTN" Value="PT5M"/>
  <ItemData ItemOID="VS.VSTPTREF" Value="PATIENT SUPINE"/>
  <Observation xmlns="http://hl7.org/fhir">
    <id value="blood-pressure"/>
    <meta>
      <profile value="http://hl7.org/fhir/StructureDefinition/vitalsigns"/>
    </meta>
    <text>
      <status value="generated"/>
      <div xmlns="http://www.w3.org/1999/xhtml">
        <p><b>Generated Narrative with Details</b></p>
        <p><b>id</b> : blood-pressure</p>
        <p><b>code</b> : Blood pressure diastolic supine<span> (Details : {LOINC code '8455-8' = 'Diastolic blood pressure diastolic supine'})</span></p>
        <p><b>value</b> : 64 mmHg<span> (Details : UCUM code mm[Hg] = 'mmHg')</span></p>
        <p><b>interpretation</b> : Below</p>
      </div>
      <system value="urn:ietf:rhc:3986"/>
      <value value="urn:uuid:187e0c12-8dd2-67e2-99b2-bf273c878281"/>
    </identifier>
    <!-- demonstrating the use of the baseOn element with a fictive identifier -->
    <basedOn>

```

Followed by:

```

<ItemData ItemOID="VS.VSTPTREF" Value="PATIENT SUPINE"/>
<Observation xmlns="http://hl7.org/fhir">
  <id value="blood-pressure"/>
  <meta>
    <profile value="http://hl7.org/fhir/StructureDefinition/vitalsigns"/>
  </meta>
  <text>
    <status value="generated"/>
    <div xmlns="http://www.w3.org/1999/xhtml">
      <p><b>Generated Narrative with Details</b></p>
      <p><b>id</b> : blood-pressure</p>
      <p><b>code</b> : Blood pressure diastolic supine<span> (Details : {LOINC code '8455-8' = 'Diastolic blood pressure--supine'})</span></p>
      <p><b>value</b> : 64 mmHg<span> (Details : UCUM code mm[Hg] = 'mmHg')</span></p>
      <p><b>interpretation</b> : Below</p>
    </div>
    <system value="urn:ietf:rhc:3986"/>
    <value value="urn:uuid:187e0c12-8dd2-67e2-99b2-bf273c878281"/>
  </identifier>
  <!-- demonstrating the use of the baseOn element with a fictive identifier -->
  <basedOn>
    <identifier>
      <system value="https://acme.org/identifiers"/>
      <value value="1234"/>
    </identifier>
  </basedOn>
  <status value="final"/>
  <category>
    <coding>
      <system value="http://hl7.org/fhir/observation-category"/>
      <code value="vital-signs"/>
      <display value="Vital Signs"/>
    </coding>
  </category>
  <code>
    <coding>
      <system value="http://loinc.org"/>
      <code value="8455-8"/>
      <display value="Diastolic blood pressure--supine"/>
    </coding>
    <text value="Diastolic blood pressure--supine"/>
  </code>

```

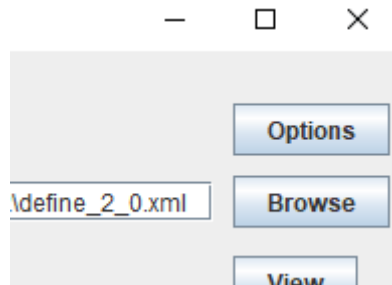
When the data point is a "Finding", the source FHIR resource will usually be an "Observation" resource. In very many cases, the test performed is characterized by a LOINC code. So LOINC is really NOT only about lab tests.

A FHIR resource always contains 2 parts: a human-readable part (within the <text> element) and a machine-readable part. Both essentially have the same contents, with the machine-readable part being highly coded. For humans however, the human-readable part is of importance.

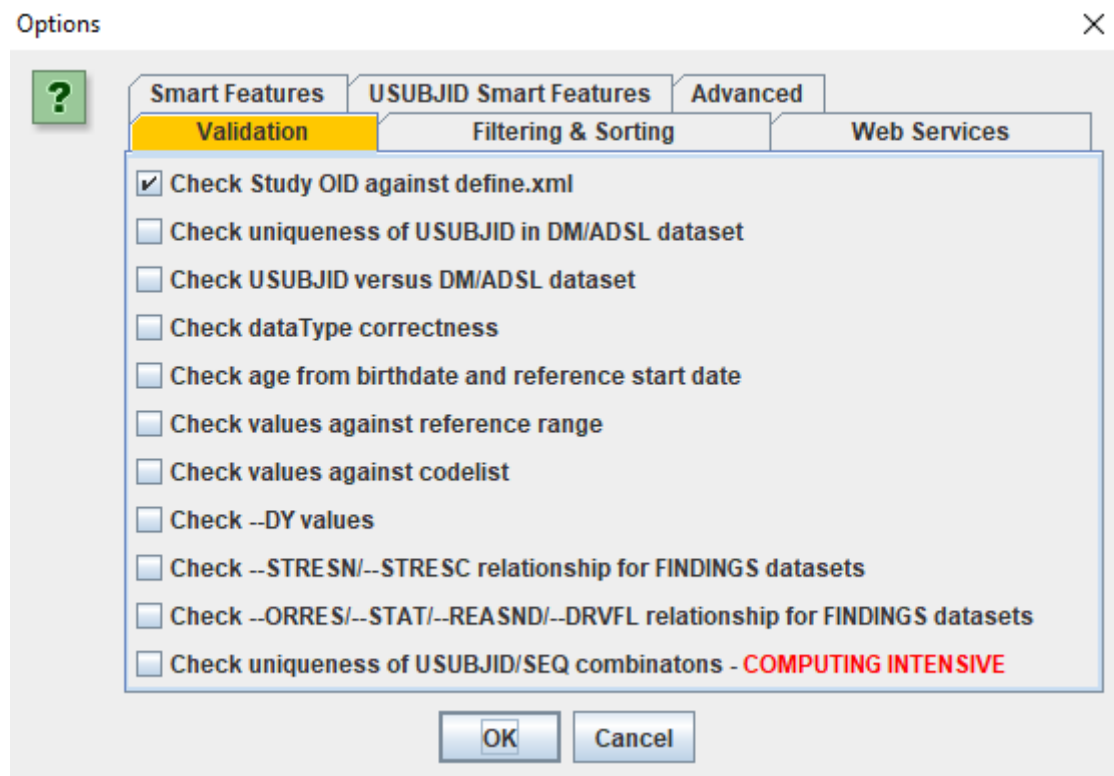
So, when the source is an EHR or a device with no form involved, and the source data point can be represented by a FHIR resource, it can simply be embedded in the XML element ("ItemGroupData" element) representing the SDTM record.

Displaying the source EHR in the "Smart Dataset-XML Viewer"

When FHIR resources are embedded as source data points in the SDTM records, they can easily be visualized by the "Smart Dataset-XML Viewer". To enable this feature, select your define.xml and SDTM datasets as usual, and then click the "Options" button:

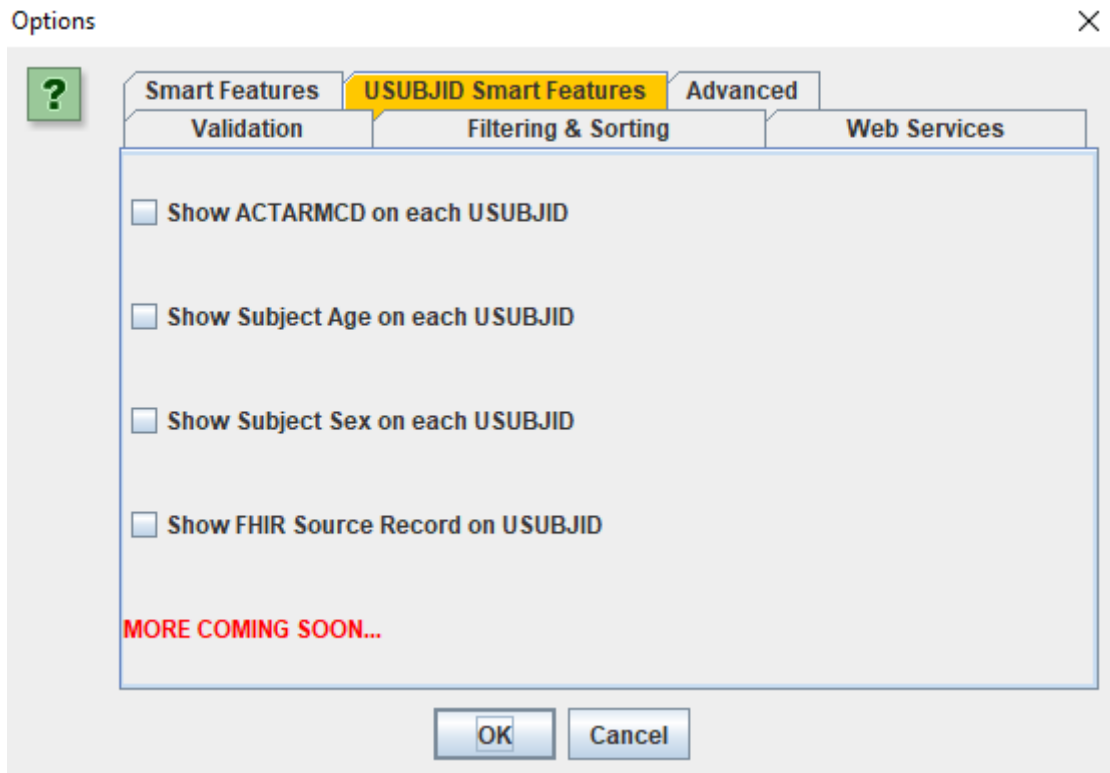


A dialog with different tabs appears:

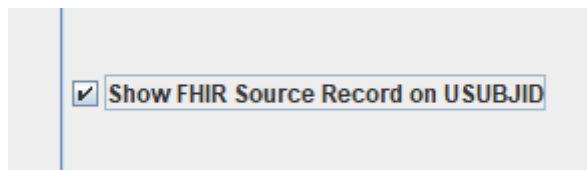


Click the "USUBJID Smart Features" Tab². The following set of options is shown.
a new screen is displayed with the sets of rules:

² We might rename this tab into e.g. "Smart Features" in near future.



In order to enable displaying the FHIR source record, check the checkbox "Show FHIR Source Record on USUBJID":



Then click on "Ok" to confirm.

When now loading SDTM datasets that have FHIR resources attached to the records³, these will be visualized by holding the mouse over the "USUBJID" cell. For example:

³ Such an example dataset "Files_from_LZZT_Pilot_2013_Dataset-XML_EHR.zip" can be downloaded from [the SourceForge project website](#), containing a VS dataset where each record contains the source FHIR data point as a FHIR resource.

DOMAIN	USUBJID	VSSEQ	VSTESTCD	VSTEST	VSPOS	VSORRES	VSORRESU	VSSTRESC	VSSTRESN	VSSTRE
VS	01-705-1282	88	SYSBP	Systolic Blo...	STANDING	125	mmHg	125	125	mmHg
VS	01-705-1282	89	SYSBP	Systolic Blo...	SUPINE	120	mmHg	120	120	mmHg
VS	01-705-1282	90	SYSBP	Systolic Blo...	STANDING	120	mmHg	120	120	mmHg
VS	01-705-1282	91	SYSBP	Systolic Blo...	STANDING	120	mmHg	120	120	mmHg
VS	01-705-1282	92	SYSBP	Systolic Blo...	SUPINE	120	mmHg	120	120	mmHg
VS	01-705-	01-705-1282 (USUBJID) FHIR record: Generated Narrative with Details id : blood pressure meta : identifier : urn:uuid:c016c0e3-88cd-4ced-8891-4a7dfcafe86e status : final category : Vital Signs (Details : {http://hl7.org/fhir/observation-category code 'vital-signs' = 'Vital Signs', given as 'Vital Signs'}) code : systolic blood pressure - standing (Details : {LOINC code '8460-8' = 'Systolic blood pressure--standing'}) subject : Patient/example effective : 2012-12-24 performer : Practitioner/example value : 120 - mmHg (Details: UCUM code mm[Hg] = 'mmHg')								
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									
VS	01-705-									

Showing the human-readable part of the source FHIR resource. The machine-readable part is not displayed, as it is mean for ... machines. However, one could imagine that also tools of the reviewers use this machine-readable part e.g. for statistical or other analysis.

Conclusion

When using Dataset-XML format instead of SAS Transport 5 ("XPT") format, source data can be embedded within the SDTM records. This is impossible with XPT.

The "Smart Dataset-XML Viewer" allows to visualize such source records in the case of FHIR resources. Principally, this is not limited to FHIR, but can be done for any other kind of source data in any format. Implementing such a new feature in the "Smart Dataset-XML Viewer" source code is just a matter of a few hours.