CDISC®

CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM

The CDISC vision is to inform patient care & safety

through higher quality medical research.

Strength *through Collaboration*

A new transport standard for electronic submissions in XML: Dataset-XML

Presented by FH-Prof.Dr. Jozef Aerts
University of Applied Sciences FH Joanneum

Graz, Austria

FH | JOANNEUM
University of Applied Sciences

Strength *through Collaboration*

What (the hell) is Dataset-XML

- A new transport format for Tabular data
  - SDTM
  - SEND
  - ADaM

- Draft name was SDS = Study Data Set
  - so not "Submission Data Set" as the FDA re-baptised it
  - but will often be used as such

- Based on ODM 1.3.1/ODM 1.3.2
  and in full alignment with define.xml 2.0

Dataset-XML Example

```xml
<?xml version="1.0" encoding="UTF-8"?>
<ODM xmlns:data="http://www.cdisc.org/ns/Dataset-XML/v1.0"
     xmlns="http://www.cdisc.org/ns/odm/v1.3"
     data:DatasetXMLVersion="1.0.0"
     ODMVersion="1.3.2"
     FileType="Snapshot"
     FileOID="SDTM-XML"
     CreationDateTime="2012-12-08T23:06:03.046+01:00">
    <ClinicalData StudyOID="LZZT" MetaDataVersionOID="CDISC.SDTM.3.1.0">
      <ItemGroupData data:ItemGroupDataSeq="1" ItemGroupOID="EX">
        <ItemData ItemOID="EX.STUDYID" Value="CDISCPILOT01"/>
        <ItemData ItemOID="EX.DOMAIN" Value="EX"/>
        <ItemData ItemOID="EX.USUBJID" Value="01-701-1015"/>
        <ItemData ItemOID="EX.EXSEQ" Value="1"/>
        <ItemData ItemOID="EX.EXTRT" Value="PLACEBO"/>
        <ItemData ItemOID="EX.EXDOSE" Value="0"/>
        <ItemData ItemOID="EX.EXDOSU" Value="mg"/>
        <ItemData ItemOID="EX.EXDOSFRM" Value="PATCH"/>
        <ItemData ItemOID="EX.EXDOSFRQ" Value="QD"/>
        <ItemData ItemOID="EX.EXROUTE" Value="TRANSDERMAL"/>
        <ItemData ItemOID="EX.VISITNUM" Value="3"/>
        <ItemData ItemOID="EX.VISIT" Value="BASELINE"/>
        <ItemData ItemOID="EX.VISITDY" Value="1"/>
        <ItemData ItemOID="EX.EXSTDTC" Value="2014-01-02"/>
        <ItemData ItemOID="EX.EXENDTC" Value="2014-01-16"/>
        <ItemData ItemOID="EX.EXSTDY" Value="1"/>
        <ItemData ItemOID="EX.EXENDY" Value="15"/>
      </ItemGroupData>
```

Dataset-XML Advantages

Next 30 slides …

But will limit to the most important ones

- Really vendor neutral
  - Did you ever try to generate Java/C#/C++ software for reading or writing SAS-XPT?

- Modern transport format => huge opportunities

- No more XPT limitations
  - we get rid of 8-, 40-, and 200-character limitations

- Easy validation against define.xml
  - same technology as for ODM validation

CDISC

Dataset-XML Disadvantages (?)

- File size ? Or is that a Myth?

	XPT	CSV	Dataset-XML	Dataset-XML combined	Dataset-XML combined & zipped
DM	0.11	0.06	0.4	0.4	0.01
SUPPDM	1.0	0.1	0.6		
AE	1.5	0.3	2.0	1.8	0.05
SUPPAE	1.0	0.1	0.7		
VS	23.5	4.8	34.1	35.7	0.07
LB	33.5	8.3	65.0	79.8	1.6
SUPPLB	55.4	7.1	37.3		
QS	33.1	2.0	122.5	122.5	2.3

Updated Pilot LZZT 2013 - file sizes in MB

CDISC

Format and file size - does it matter?

- it does not unless you are still using a telephone modem …

- Zipping XML files is extremely efficient

- What matters is:

  MEMORY USAGE IN TOOLS!

The Myth of filesize - memory usage

- Relative <u>memory usage</u> for tools for different transport formats
- Assuming that software is correctly designed and implemented …

- CSV   1.0
- XPT   1.0
- XML   1.0
- Zipped XML   1.0

Opportunities

- Next 40 slides …

- But I will limit myself to the most important ones …

CDISC

Audit Trails & Signatures

```xml
<ItemData ItemOID="AE.AESEV" Value="MILD">
    <AuditRecord EditPoint="Monitoring" UsedImputationMethod="Yes">
        <UserRef UserOID="ZBIuser000"/>
        <LocationRef LocationOID="XML4PharmaLocation"/>
        <DateTimeStamp>2013-12-21T11:59:59.9+01:00</DateTimeStamp>
        <ReasonForChange>Originally classified as moderate, then
        corrected to mild as subject had worked the whole day in the
        forest</ReasonForChange>
    </AuditRecord>
    <Signature>
        <UserRef UserOID="XML4Pharmauser000"/>
        <LocationRef LocationOID="XML4PharmaLocation"/>
        <SignatureRef SignatureOID="XML4PharmaSignature000"/>
        <DateTimeStamp>2013-12-31T11:59:59.9+01:00</DateTimeStamp>
    </Signature>
</ItemData>
<ItemData ItemOID="AE.AESER" Value="N"/>
```

Integration with electronic health records (EHRs)

```xml
<ItemGroupData ItemGroupOID="VS">
    <ItemData ItemOID="VS.STUDYID" Value="CDISCPILOT01"/>
    <ItemData ItemOID="VS.DOMAIN" Value="VS"/>
    <ItemData ItemOID="VS.USUBJID" Value="01-701-1015"/>
    <ItemData ItemOID="VS.VSSEQ" Value="2"/>
    <ItemData ItemOID="VS.VSTESTCD" Value="DIABP"/>
    <ItemData ItemOID="VS.VSTEST" Value="Diastolic Blood Pressure"/>
    <ItemData ItemOID="VS.VSPOS" Value="STANDING"/>
    <ItemData ItemOID="VS.VSORRES" Value="83"/>
    <ItemData ItemOID="VS.VSORRESU" Value="mmHg"/>
    <ItemData ItemOID="VS.VSSTRESC" Value="83"/>
    <ItemData ItemOID="VS.VSSTRESN" Value="83"/>
    <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 STANDING FOR 1 MINUTE"/>
    <ItemData ItemOID="VS.VSTPTNUM" Value="816"/>
    <ItemData ItemOID="VS.VSELTM" Value="PT1M"/>
    <ItemData ItemOID="VS.VSTPTREF" Value="PATIENT STANDING"/>
    <cda:observation classCode="OBS" moodCode="EVN">
        <cda:id root="E6F45742-BA20-7224-15A0-6D8836E96789"/>
        <cda:code code="8462-4" displayName="Intravascular Diastolic" codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC">
            <cda:originalText>
                <cda:reference value="#vitsigtype3"/>
            </cda:originalText>
        </cda:code>
        <cda:text>
            <cda:reference value="#vitsig3"/>
        </cda:text>
        <cda:statusCode code="completed"/>
        <cda:value xsi:type="PQ" value="83" unit="mm[Hg]"/>
    </cda:observation>
</ItemGroupData>
```

FHIR would even do better

The same in detail …

```xml
<ItemData ItemOID="VS.VSTPTREF" Value="PATIENT STANDING"/>
<cda:observation classCode="OBS" moodCode="EVN">
    <cda:id root="E6F45742-BA20-7224-15A0-6D8836E96789"/>
    <cda:code code="8462-4" displayName="Intravascular Diastolic"
    codeSystem="2.16.840.1.113883.6.1" codeSystemName="LOINC">
        <cda:originalText>
            <cda:reference value="#vitsigtype3"/>
        </cda:originalText>
    </cda:code>
    <cda:text>
        <cda:reference value="#vitsig3"/>
    </cda:text>
    <cda:statusCode code="completed"/>
    <cda:value xsi:type="PQ" value="83" unit="mm[Hg]"/>
</cda:observation>
</ItemGroupData>
```

P.S.: CDA data point is on the record level (DIABP)

Tools for working with SDS-XML

- XPT2DatasetXML
  - transforms XPT datasets into SDS-XML datasets
  - does not take advantage of advantages SDS-XML
  - so don't use if you do not really need to …
- Smart Dataset-XML Viewer
  - like the SAS Viewer
  - but much much smarter
- SAS Clinical Standards Toolkit
  - coming …
  - Special thanks to Lex Jansen
- EZ Convert (Sally Cassells)
  - converts Datset-XML files into SAS Datasteps
  - Allows easy import into SAS
- R4CDISC
  - Reads Dataset-XML into the R statistical software

The Smart Dataset-XML Viewer

- Open Source (similar to OpenCDISC license)
- For working with define.xml + Dataset-XML files
  - also reads zipped Dataset-XML
- Understands SDTM, SEND and ADaM

- Freely available at:
  http://sourceforge.net/projects/smart-sds-xml-viewer/

- Pluggable into any modern software

CDISC

Features of the Smart Dataset-XML Viewer

- Simple validation
  - But OpenCDISC will do more …
- Display of metadata on column headers
- Coloring of non-standard variable values in parent domain
- Highlighting out-of-normal-range values
- Toggling between SUPPxx /RELREC and parent records
- Linking between datasets – e.g.
  - CTRL-D: show DM record for current record
  - e.g. show first EX record information in DM dataset
  - Calculate --DY on the fly

Hey, we can soon get rid of --DY, VISIT (name), and all redundant variables !

CDISC

Features of the Smart Dataset-XML Viewer

Features of the Smart Dataset-XML Viewer

- Filtering   (example: Alzheimer study)
  - Load DM
  - Select all subjects >80 years old
  - Load CM data (only for subjects > 80 years old)
  - Select all subjects (>80 years old) that had Hydrocortisone as concomitant medication
  - Load VS data for subjects >80 years old that had Hydrocortisone as concomitant medication
  - Select …
  - Load …

Smart Dataset-XML Viewer
Features soon coming

- Look up detailed information for –LOINC codes over web service

- Later:
  - look up web services for SNOMED-CT, ATC, WHODRUG, …
  - create and populate relational database, data warehouse …

- … can grow into a modern review environment …

Smart Dataset-XML Viewer Availability

- Executables: now (SourceForge website)

- Source code: as soon as Dataset-XML is released as „final"

- We (the University of Applied Sciences FH Joanneum) will provide services to sponsors who would like to extend the Smart Dataset-XML viewer
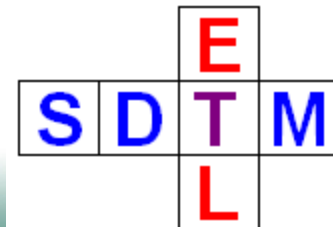 - E.g. simple analysis, distributions, web services …

FH | JOANNEUM
University of Applied Sciences

CDISC

The Smart Dataset-XML Viewer

- Some movies at:
    - http://www.youtube.com/watch?v=wRBktVtB47s

- And a demo after this meeting if you want…
  (I only leave on Friday …)

CDISC

EZ Convert

- Converts Dataset-XML into SAS Datasteps
- Author: Sally Cassells
- Movie available at:
  http://www.youtube.com/watch?v=XIky6ePfo1I

SDTM-ETL 3.0 now also generates Dataset-XML

- My own mapping tool for generation of SDTM datasets from ODM

- v.3.0 supports define.xml 2.0 and SDTM 1.4 (SDTM-IG 3.2)

- Choice between output in SAS-XPT and Dataset-XML

- Smart Dataset-XML viewer integrated

- Other vendors will also soon support Dataset-XML

Dataset-XML – the opportunity to innovate …

Thank you for your attention