



15th–19th November

EU Connect 2021

The Clinical Data
Science Conference

Real World Data, CDISC SDTM, HL7 FHIR, and the Future of Regulatory Submissions

Jozef Aerts
XML4Pharma

PHUSE EU Connect - November 2021

Real World Data, CDISC SDTM, HL7 FHIR, and the Future of Regulatory Submissions



Jozef Aerts

CEO XML4Pharma

A software and CDISC consultancy company

CDISC Volunteer for already 20 years

CDISC Authorized Instruction for ODM, Define-XML

FHIR-CDISC Integration Pioneer

Participant in the HL7 Vulcan Project

Mail: Jozef.Aerts@XML4Pharma.com

The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC, HL7 or PharmaSUG.

The use of Real World Data in Clinical Research

- The use of Real World Data (RWD) in clinical research is rapidly growing
- This is in part due to the fast growing use of HL7-FHIR for exchange of (even distributed) Electronic Health Records (EHRs)
- HL7 has also added new FHIR "Resources" specifically for Research (ResearchStudy, ResearchSubject)
- The current HL7-Vulcan project aims to bridge existing gaps between clinical care (FHIR standard) and clinical research (CDISC standards)
- This raises the question whether FHIR could also make it as a future format for submissions to the regulatory authorities (FDA, PMDA, ...)

The classic CDISC-based submission-generation process

- The protocol (Word, PDF) is translated into a set of CRFs and is fed into an EDC system.
 - Smarter companies use an MDR with SDTM pre-annotated questions and CRFs
- After finalizing data collection, EDC and other data is mapped to SDTM with output of XPT files
 - Smarter companies use specialized mapping tools to (partially) automate this, using a pre-generated define.xml as the mapping specification
- Analysis is performed with output of ADaM datasets, TFLs, ...

Current Problems with SDTM (personal opinion)

- **Table based**, with **implicit relations** between the tables (relations are only described in Implementation Guides)
=> Conformance rules must be developed to check the relations
- Mandated use of "punchcard-like" SAS Transport 5 format (XPT)
- Each new version of SDTM-IG is a "patch" on a prior version, correcting problems with the prior version - and introducing new ones ...
 - IMO, splitting up "Findings" into "domains" was a major design error
- Refusal to use modern terminologies, such as SNOMED, LOINC, UCUM, ... ("not-invented-here" syndrome)



Use of RWD in the "classic" Submission Process

- First possibility is to retrieve data from the EHRs and feed them into CRFs and the EDC system.
 - Requires an enormous (usually non-reusable) programming effort ...
- The other possibility is to generate **SDTM datasets directly from the EHRs**
 - modern APIs such as the FHIR API make this possible and automatable

Current Initiatives

- FHIR-CDISC Mapping
 - mapping between FHIR resources (content) and SDTM domains/variables
 - outsourced to [Gevity](#) (Lloyd McKenzie)
 - does not include **semantic mapping** (LOINC/SNOMED => CDISC)
- HL7 Vulcan
 - Part of the project is demonstrating automated generation of SDTM datasets from FHIR resources
 - current status: only CM

Our own work ...

- Extended the existing (by CDISC published) LOINC-CDISC **mappings** from 1,400 to **>9,600 distinct LOINC codes**
 - includes LB, MB-Corona, VS, partially EG
- Developed (open source) software to extract data from EHRs using the FHIR-API, categorize the retrieved information to domains, and automatically generate SDTM datasets
 - => DM, CM, MH, AE, LB, VS, MB, RELREC, SUPPxx
 - with output in CDISC's own Dataset-XML format with **embedded FHIR source records** (for **traceability**)
- Implemented the published SNOMED <=> MedDRA mappings as a private RESTful Web Service

Software Implementation

- Java-based, using a lot of APIs and RESTful web services, including from NLM (UCUM unit conversions for --STRESN), FHIR-API, HAPI-FHIR and our own LOINC-CDISC RESTful web services:
 - <https://ucum.nlm.nih.gov/ucum-service.html>
 - <https://www.hl7.org/fhir/http.html>
 - <https://hapifhir.io/>
 - <http://xml4pharmaserver.com/WebServices/index.html>
- Software is free and OPEN SOURCE:
 - <https://sourceforge.net/projects/fhirloinc2sdtm/>



FHIRLOINC2SDTM

Brought to you by: [xml4pharma](#)

Some screenshots ...

```
2021-08-30 09:27:34,919 INFO # of RELREC records created = 318
2021-08-30 09:27:34,997 INFO Wrapping up ...
2021-08-30 09:27:35,004 INFO # of DM records = 95
2021-08-30 09:27:35,004 INFO # of LB records = 0
2021-08-30 09:27:35,004 INFO # of MB records = 93
2021-08-30 09:27:35,004 INFO # of VS records = 93
2021-08-30 09:27:35,004 INFO # of CM records = 324
2021-08-30 09:27:35,004 INFO # of MH records = 545
2021-08-30 09:27:35,004 INFO # of AE records = 0
2021-08-30 09:27:35,004 INFO # of RELREC records = 318
2021-08-30 09:27:35,004 INFO # of SUPPMB records = 186
2021-08-30 09:27:35,004 INFO # of SUPPLB records = 0
2021-08-30 09:27:35,004 INFO # of SUPPVS records = 186
2021-08-30 09:27:35,004 INFO # of SUPPCM records = 282
***** All Done *****
```

Which FHIR-EHR
0 - SyntheaStu
2021-08-30 09:27:34,919 INFO # of RELREC records created = 318
race = null -
2021-08-30 09:27:34,997 INFO Wrapping up ...
2021-08-30 09:27:35,004 INFO # of DM records = 95
2021-08-30 09:27:35,004 INFO # of LB records = 0
2021-08-30 09:27:35,004 INFO # of MB records = 93
2021-08-30 09:27:35,004 INFO # of VS records = 93
2021-08-30 09:27:35,004 INFO # of CM records = 324
2021-08-30 09:27:35,004 INFO # of MH records = 545
2021-08-30 09:27:35,004 INFO # of AE records = 0
2021-08-30 09:27:35,004 INFO # of RELREC records = 318
2021-08-30 09:27:35,004 INFO # of SUPPMB records = 186
2021-08-30 09:27:35,004 INFO # of SUPPLB records = 0
2021-08-30 09:27:35,004 INFO # of SUPPVS records = 186
2021-08-30 09:27:35,004 INFO # of SUPPCM records = 282
Do you also want to use a selection of records?
Y
0 = all
1 = obs
0
Do you want to use a selection of records?
N

ion)
relations?

nverter
lts/MB.xml

Features of the Implementation

- Coverage: DM, CM, MH, LB, MB-Corona, VS, AE, RELREC, SUPPxx
- Choice between starting from A FHIR ResearchStudy/ResearchSubject or a SNOMED-CT code for a specific disease (Condition)
- Automated classification of observations into Findings domains
- Automated unit conversions for -STRESC (using UCUM conversions)
- Automated SNOMED-MedDRA transformation for AEs
 - using non-public RESTful web service
- FHIR source record is embedded in SDTM record (Dataset-XML format)

Visualization in the Open Source "Smart Submission Dataset Viewer"

DM	RELREC	AE	CM	MB	MH	VS	SUPPCM	SUPPMB	SUPPVS				
STUDYID	DOMAIN	USUBJID	MBSEQ	M...	MBTESTCD	MBTEST	MBTSTDTL						
HAPIFHIR	MB	1171862	1	c...	SAR2RNA	SARS-CoV-2 RNA	DETECTION						
HAPIFHIR	MB	1171862	1	4	SAR2RNA	SARS-CoV-2 RNA	DETECTION						
STUDYID	HAPIFHIR	MB	MBDTC	MBDY	MBEVLINT	MBRESTYP	MBRSLSCL	MBSYM	MBLOC	MBMETHOD	VIS		
HAPIFHIR	HAPIFHIR	MB	2020-03-19...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST			
HAPIFHIR	HAPIFHIR	MB	2020-01-14...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST			
HAPIFHIR	HAPIFHIR	MB	2020-01-11...	-1	PT	PRESENCE OR THRESHOLD	ORDINAL		IC)				
HAPIFHIR	HAPIFHIR	MB	2020-03-11...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		RS coronavirus 2 RNA:PrThr:Pt:Respiratory:Ord:Probe.amp.tar				
HAPIFHIR	HAPIFHIR	MB	2020-01-22...	-1	PT	PRESENCE OR THRESHOLD	ORDINAL		Name: SARS-CoV-2 (COVID-19) RNA [Presence] in Respiratory specimen by NAA				
HAPIFHIR	HAPIFHIR	MB	2020-01-20...	-1	PT	PRESENCE OR THRESHOLD	ORDINAL		on				
HAPIFHIR	HAPIFHIR	MB	2020-01-09...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		CRO (Microbiology)				
HAPIFHIR	HAPIFHIR	MB	2020-01-19...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		pound: SARS coronavirus 2 RNA (SARS-CoV-2 (COVID-19) RNA)				
HAPIFHIR	HAPIFHIR	MB	2020-03-04...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		resence or Threshold)				
HAPIFHIR	HAPIFHIR	MB	2020-02-03...	-1	PT	PRESENCE OR THRESHOLD	ORDINAL		Point in time (spot))				
HAPIFHIR	HAPIFHIR	MB	2020-02-04...	-1	PT	PRESENCE OR THRESHOLD	ORDINAL		ory (Respiratory specimen)				
HAPIFHIR	HAPIFHIR	MB	2020-01-24...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		al)				
HAPIFHIR	HAPIFHIR	MB	2020-03-02...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		np.tar (Nucleic acid amplification with probe detection)				
HAPIFHIR	HAPIFHIR	MB	2020-03-12...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST			
HAPIFHIR	HAPIFHIR	MB	2020-03-13...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST			
HAPIFHIR	HAPIFHIR	MB	2020-01-29...	1	PT	PRESENCE OR THRESHOLD	ORDINAL		RESPIRATORY SYSTEM	NUCLEIC ACID AMPLIFICATION TEST			

DM	RELREC	AE	CM	MB	MH	VS	SUPPCM	SUPPMB	SUPPVS				
STUDYID	DOMAIN	USUBJID	AESEQ	AEREFID	AETERM	AELLT	AELLTCD						
HAPIFHIR	AE	1171954	1	2407099	Headache	Headache	10019211						
HAPIFHIR	AE	1171865	1	2407210	Hair loss	Hair loss	10019045						
HAPIFHIR	AE	1171865	2	2407201	Application site rash	Application site rash	10003054						
		2020-03-12...	1	PT	PRESENCE OR THRESHOLD	ORDINAL							
		2020-03-13...	1	PT	PRESENCE OR THRESHOLD	ORDINAL							
		2020-01-29...	1	PT	PRESENCE OR THRESHOLD	ORDINAL							

FHIR and Clinical Research

- HL7 has developed a number of FHIR resources for use in Clinical Research: ResearchStudy, ResearchSubject, PlanDefinition, ActivityDefinition, Questionnaire ...
- We can now already represent >90% of a clinical research study definition as a set of (explicit) connected FHIR resources
- This triggers the question whether FHIR would also be suitable as a submission standard / format

FHIR for Regulatory Submissions: Advantages

- Relations between data points is EXPLICIT (not implicit as in the case of SDTM)
- Moving away from "table thinking" (A.Oliva: "the world is not flat")
- Use of modern exchange formats: XML, JSON, RDF-Turtle ...
- For observations/findings: no arbitrary classification in "domains" anymore
 - finding is specified by the LOINC code
- Better traceability
- Submission data can be remote, distributed, ... moving away from "files"
- Better support of "Rolling Data Review" (by RESTful-WS)

FHIR for Regulatory Submissions: Disadvantages

- No review tools available yet
- Reviewers are used to "pre-cooked" derived variables
 - they (or their tools) would need to derive these themselves
- Reviewers are used to work with SDTM
 - Education/training needed, change management
- Relation with ADaM is lost

Conclusions

- With the new FHIR "Research" resources, clinical studies can completely be planned and executed using FHIR as an exchange format
- Already for many domains, SDTM datasets can be generated fully automatically starting from EHRs that do have a FHIR interface
- FHIR is also suitable as an electronic submission format for clinical studies to the regulatory authorities, with many advantages
 - e.g. enabling distributed submission data and real "rolling data review"
- "Table" and "File thinking" however are serious hurdles at regulatory authorities to modernize how submissions are done

Background Information

- Presentations + demo movies

"Generating FDA-ready Submission Datasets directly from EHRs" - Part 1:

<https://www.youtube.com/watch?v=LX1XS18CeYQ&t=398s>

"Generating FDA-ready Submission Datasets directly from EHRs" - Part 2:

<https://www.youtube.com/watch?v=eLbtWmK12jg&t=152s>

"Fully automated Generation COVID-19 SDTM Datasets from EHRs"

<https://www.youtube.com/watch?v=6bg59Ot55nU>

- Open Source Software FHIRLOINC2SDTM:

<https://sourceforge.net/projects/fhirloinc2sdm/>

- Slides of the current presentation:

www.xml4pharma.com/publications/Phuse_EU_Connect_2021/PRE_RW04.pdf