2022-01 Vulcan - Real World Data (RWD) Submission to FDA

- Short Description
- Long Description
- Type
- Submitting Work Group/Project/Accelerator/Affiliate/Implementer Group
- Track Lead(s)
- Track Lead Email(s)
- Related Tracks
- FHIR Version
- Specification(s) this track uses
- Artifacts of focus
- Expected participants
- Zulip stream
- Track Details
- Kick-off Call
- Artifacts from Previous Tracks
- Useful artifacts for January 2022 Track

**Short Description**

The purpose of this track is to develop methods to use the HL7 FHIR standard to retrieve relevant data from Real World Data sources – specifically for this track, Electronic Health Record (EHR) systems. The destination format for this track is the SDTM (Study Data Tabulation Model) standard, created by the Clinical Data Interchange Standards Consortium (CDISC) standards development organization, which optimized for clinical research and regulatory uses and is the data standard used for regulatory submissions of study data to the US Food and Drug Administration.

**Long Description**

Real World Data can be considered data created in the “real world” of everyday experience, such as a routine patient visit to a healthcare provider, as opposed to data created under clearly defined protocols typical of controlled clinical trials. The primary purpose for such data, collected for a purpose other than use in a clinical trial, is in support of clinical care of patients and knowledge for their healthcare providers. However, large amounts of such information could potentially be used for the secondary purpose of supporting clinical research to analyze the data and generate supporting evidence for, as an example, a new indicated use for an already approved pharmaceutical drug or safety-related analyses.

Many sources of RWD exist, but for the current phase of work, the scope of the track is firmly limited to the use of Electronic Health Record (EHR) systems as sources of RWD. Additionally, broad use case is currently limited to the use of EHRs for retrospective analysis of data (to generate evidence for new indications, comparisons, and/or safety) and preparation of such data for submission to governmental regulatory bodies covering pharmaceutical approvals such as the United States Food and Drug Administration (FDA). The use of EHRs as a mode of direct data collections for traditional prospective clinical trials (sometimes called “electronic source data” or “eSource” activities) is not currently in scope. However, we consider it highly likely that types of solutions developed for eSource and for RWD will have significant overlap.

The challenge faced by the track is: How, with the aide of HL7 FHIR, can we most efficiently and comprehensively migrate data and bridge the many syntactic and semantic gaps from the healthcare data sphere to the research and regulatory sphere?

**Previous Work**

January 2022 starts the second year of this connection track. In the first year of this work, to better understand the end-to-end challenges faced when moving information from a healthcare setting to a research and regulatory setting, we focused on the specific data domain of concomitant medications (other medications taken by people at the same time as a medication of interest). Specifically, we focused on what it takes to find, retrieve, and transform, relevant medication information for patients from an EHR data source to an acceptable SDTM format while not losing critical supplemental information in the process.

Using available test data set of fictional patients, we initially retrieved medications information from the following resources: MedicationStatement; MedicationRequest. Subsequently, we expanded to allow retrieval from additional resources, MedicationDispense and MedicationAdministration, since we found that different EHRs implement and use the suite of these 4 FHIR resources (referred to as Medication[]() in highly variable ways. We explored the challenges of differing levels of “certainty” in EHRs regarding the actual ingestion of a drug, how best to package results of data queries in FHIR for transport and transformation, and the use of SDTM SUPPxx supplemental domains to represent additional supplemental when such information semantically different from that typically stored in the core domains of SDTM.

Finally we identified key attributes of the Medication[]() Resources that we felt to be critical for implementation in EHRs if such data is to support clinical research and submissions. Some of these findings resulted in new proposal submissions to the US Core Data for Interoperability (USCDI) requirements created by the US Office of the National Coordinator for Health IT (ONC). Specific attributes such as Status and Patient for MedicationRequest and MedicationAdministration are now in queue for potential inclusion in USCDI version 3.

<table>
<thead>
<tr>
<th>Date Submitted</th>
<th>Data Class</th>
<th>Data Element</th>
<th>Status</th>
<th>Classification Level</th>
<th>LINK</th>
</tr>
</thead>
<tbody>
<tr>
<td>09/30/2021</td>
<td>Medications</td>
<td>Medication Prescription Patient</td>
<td>Published</td>
<td>Level 2</td>
<td>Link</td>
</tr>
<tr>
<td>09/30/2021</td>
<td>Medications</td>
<td>Medication Administration Patient</td>
<td>Published</td>
<td>Level 1</td>
<td>Link</td>
</tr>
<tr>
<td>09/30/2021</td>
<td>Medications</td>
<td>Medication Administration Status</td>
<td>Published</td>
<td>Level 2</td>
<td>Link</td>
</tr>
<tr>
<td>09/30/2021</td>
<td>Medications</td>
<td>Medication Prescription Status</td>
<td>Published</td>
<td>Level 2</td>
<td>Link</td>
</tr>
<tr>
<td>09/30/2021</td>
<td>Medications</td>
<td>Medication Prescription Do-Not-Perform</td>
<td>Published</td>
<td>Level 2</td>
<td>Link</td>
</tr>
<tr>
<td>09/30/2021</td>
<td>Medications</td>
<td>Reported Medication (unique)</td>
<td>Published</td>
<td>Level 2</td>
<td>Link</td>
</tr>
</tbody>
</table>

**Plan for January 2022 Connection Track**

In the new year, we aim to move beyond the focus on a single data domain and, instead, take a practical and realistic, query-based approach. Specifically, we intend to begin with a data query that can ask a (very small) question, based on some existing study or protocol, which calls on very specific information from a range of different data domains. This will allow us to increase our breadth of understanding of the challenges and solutions. Another Vulcan track, Schedule of Activities, is working with a public domain study protocol H2Q-MC ZZT[c], which will serve as the protocol and will also create a cross-cutting synergy between the two tracks. Additionally we will be exploring revising the technical approach by working with a software developer who has worked on this set of issues (see his background and coding approach here: https://sourceforge.net/projects/ParlondStdm/).

**Type**

Test the design of a Resource/set of Resources
<table>
<thead>
<tr>
<th><strong>Submitting Work Group / Project / Accelerator / Affiliate / Implementer Group</strong></th>
<th>Vulcan Accelerator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Track Lead(s)</strong></td>
<td>Scott Gordon, Lauren McCabe</td>
</tr>
<tr>
<td><strong>Track Lead Email(s)</strong></td>
<td><a href="mailto:Gideon.Gordon@fda.hhs.gov">Gideon.Gordon@fda.hhs.gov</a>; <a href="mailto:Lauren.McCabe@pfizer.com">Lauren.McCabe@pfizer.com</a></td>
</tr>
<tr>
<td><strong>Related Tracks</strong></td>
<td></td>
</tr>
<tr>
<td><strong>FHIR Version</strong></td>
<td>FHIR R4 is current primary focus, also US Core. However, since many other implementations exist, we are not limiting this exclusively. Further, as R5 moves forward in balloting, we will keep an eye on relevant changes.</td>
</tr>
<tr>
<td><strong>Specification(s) this track uses</strong></td>
<td>Expanding to multiple resources. Medication[x]; Probably Observations, Patient, etc.</td>
</tr>
<tr>
<td><strong>Artifacts of focus</strong></td>
<td>Vulcan members, Regulators, EHR Vendors, EDC Vendors, Academic Medical Centers, Pharma Companies. Additionally:</td>
</tr>
<tr>
<td></td>
<td>• People who have FHIR data (FHIR servers, files, reference EHRs) that they would like to test against the xml4pharma implementation (see Useful Artifacts section below) being used for this track. This can be used both to check the ability of this implementation to pull desired information from a given data set and also exercise the mappings of the recently published CDISC-FHIR mappings.</td>
</tr>
<tr>
<td></td>
<td>• People wanting to test their own implementations against our FHIR sample data set. This data set is currently very small and custom created (and also synthetic), yet optimized for maximal usage from a clinical research perspective. In other words, this FHIR sample patient set (which will be present on a FHIR server or as a file-based option) does not represent typical EHRs of today - with widely variable implementation and population of FHIR resources - but, instead, a deliberative and consistent use of canonical FHIR R4 resources which we hope to exercise and revise to demonstrate a preferred FHIR profiles approach that will best support the needs of clinical research and ultimate use in a pharmaceutical regulatory submission.</td>
</tr>
<tr>
<td><strong>Zulip stream</strong></td>
<td>#Vulcan/RWD</td>
</tr>
<tr>
<td><strong>Track Details</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Kick-off Call</strong></td>
<td>Link to Recording in Zoom: <a href="https://hl7-org.zoom.us/rec/share/lNI2b0bSkeKdWio0gN7vAdP4Os19WeOIB_0ZiQIEpqaj6A6Bzr9JL36C8xq3qQ-WipZa8-XmYG9vU?startTime=1641311841000">https://hl7-org.zoom.us/rec/share/lNI2b0bSkeKdWio0gN7vAdP4Os19WeOIB_0ZiQIEpqaj6A6Bzr9JL36C8xq3qQ-WipZa8-XmYG9vU?startTime=1641311841000</a></td>
</tr>
<tr>
<td></td>
<td>Slides from call: <a href="https://hl7-org.zoom.us/rec/share/lNI2b0bSkeKdWio0gN7vAdP4Os19WeOIB_0ZiQIEpqaj6A6Bzr9JL36C8xq3qQ-WipZa8-XmYG9vU?startTime=1641311841000">Vulcan RW...05-22.pdf</a></td>
</tr>
<tr>
<td><strong>Artifacts from Previous Tracks</strong></td>
<td>Draft of decision algorithms based of Medication[x] Status field</td>
</tr>
<tr>
<td></td>
<td>med-status.xlsx</td>
</tr>
<tr>
<td></td>
<td>Draft identifying all elements of all Medication[x] resources which we recommend be implemented and utilized consistently (where relevant) in all EHRs in order to make concomitant medication determination more reliable. (There are absolutely strong pure healthcare justifications as well for these elements being present).</td>
</tr>
<tr>
<td><strong>Useful artifacts for January 2022 Track</strong></td>
<td>September 2021 updated CDISC-FHIR Mapping</td>
</tr>
<tr>
<td></td>
<td>LZZT Study protocol</td>
</tr>
<tr>
<td></td>
<td>This is the LZZT protocol, <a href="https://wiki.ihe.net/images/4/47/Lzzt_protocol_redacted.pdf">https://wiki.ihe.net/images/4/47/Lzzt_protocol_redacted.pdf</a> Lilly donated it years ago. There is a database of the SDTM on GitHub - Transcelerate - not sure who (T). CDISC has it on the website years ago but it is no longer there.</td>
</tr>
<tr>
<td></td>
<td>Sample SoA Spreadsheet: ScheduleOfActivities_01.xlsx</td>
</tr>
</tbody>
</table>
Sample LZZT Variable Mapping: LZZT_Variable_Mappings.xlsx
Sample ODM File: LZZTTrial.xml
Sample CRF String: _AnnotatedCRF.pdf
Sample Study Protocol: Lzzt_protocol_redacted.pdf

LZZT Datasets in .xpt
- aa.xpt
- cm.xpt
- define.xml
- dm.xpt
- ds.xpt
- ex.xpt
- lb.xpt
- mh.xpt
- suppdb.xpt
- suppdb2.xpt
- suppdm.xpt
- suppa.xpt
- ss.xpt
- sc.xpt
- relrec.xpt
- qgrp.xpt
- qgrmm.xpt
- qsci.xpt
- qsci.xpt
- qsci.xpt

LZZT Datasets in .xls
- LZZT_Excel.zip

LZZT Blank CRF
- blankcrf.pdf

FHIRLOINC2SDTM
Open source software brought to you by: xml4pharma

Manual for use with the LZZT FHIR resources:

RWD_demo_FHIR2...DTM_manual.pdf

(last update 2022-01-11)

Executables:
- Retrieve and transform FHIR dataset to SDTM: FHIR2SDTM open source software
  https://sourceforge.net/projects/fhirloinc2sdtm/
- Viewer for SDTM output files (optional, can also view files manually): Open source “Smart Submission Dataset Viewer”, https://sourceforge.net/projects/smart-submission-dataset-viewer/
- The semantic mappings and the API for it can be found at:
  http://xml4pharmaserver.com/WebServices/LOINC2CDISC_webServices.htm

Demonstration videos
PharmaSUG:
- presentation part 1: https://www.youtube.com/watch?v=LX1Xs18cYlA&t=398s
- presentation part 2 (containing the actual demo): https://www.youtube.com/watch?v=eLbtWmK12js&t=152s

CDISC Interchange 2020: generating COVID-19 SDTM datasets from EHRs:
https://www.youtube.com/watch?v=tNPvE4mKRUJ&t=17s (also includes a demo).
- FHIR-LOINC RESTful web Service: https://loinc.org/fhir/

**Test bundle for the LZZT study**

Created from Source data: See above under “LZZT Study Protocol”

(last update: 2022-01-08):

**JSON format:**

```
LZZT_FHIR_Bundle...Resources.json
```

**XML format:**

```
LZZT_FHIR_Bundle...Resources.xml
```

Server-based link to FHIR bundle:  https://demo1.aha.accenture.com/thrivulcan/r4/
(no web front end)

Containing:

- ResearchStudy for LZZT study - courtesy SoA team
- ResearchSubjects for LZZT study (10+1(screenfailure) subjects) - courtesy SoA team
- PlanDefinition resources for LZZT study - courtesy SoA team
- Observation resources (reverse-engineered from LZZT LB and VS-SDTM dataset)
- AdverseEvent resources (reverse-engineered from LZZT AE dataset)
- Condition resources (reverse-engineered from LZZT MH dataset)
- MedicationStatement resources (reverse-engineered from LZZT CM dataset)
- ObservationDefinition resources for 7 vital signs tests and 5 laboratory tests (to be extended)
The following queries (set of LOINC codes) will be used for Vital Signs and Laboratory observations:

Results:
Latest set of results of the generation of SDTM datasets and records from the FHIR Bundle for the LZZT study, including a define.xml:

Format is CDISC Dataset-XML format (https://www.cdisc.org/standards/data-exchange/dataset-xml) as FHIR source records are embedded within the SDTM records (impossible with SAS Transport). "Smart" Visualization can be performed using the open source "Smart Submission Dataset Viewer" - available from: https://sourceforge.net/projects/smart-submission-dataset-viewer/files/.

SDTM Data Set in CSV format:

SPECIAL NOTE: Please be aware that the CSV files, exported from the .xml file tables, has all dates/times represented in the required ISO 8601 date format (ie, 2022-01-25). We noted that when loading the CSV files into Microsoft Excel, Excel automatically converts some dates to local format and not others. Attempts to manually convert columns to YYYY-MM-DD resulted in dates that were just a year (ie, "2013") to be converted into very incorrect dates. We didn't have sufficient time to figure the magic words to make Excel do what we wanted. The data in the actual CSV files is in the correct format.
LZZT_FHIR_Elements was created by

- Reading FHIR Resource, FHIR Element and Domain from HL7_FHIR_to_CDISC_Mapping.xlsx
- "Pivoting" to make Domain columns with 'Y' flag
- Doing a full join with all elements used in the LZZT bundle
- Adding extension URL

SDTM Review Comments

Post-Connectathon developed mapping overview table Jozef Aerts, also revealing gaps