Paper TT16

Use of HL7 FHIR as eSource to Pre-populate CDASH Case Report Forms using a CDISC ODM API

Sam Hume, CDISC, State College, PA, USA
Jeff Abolafia, Rho, Chapel Hill, NC, USA
Geoff Low, Medidata Solutions, London, UK

ABSTRACT

Use of Fast Healthcare Interoperability Resources (FHIR) in the Generation of Real World Evidence (RWE) demonstrated that electronic CRF data could be populated by mapping FHIR resources to CDASH/SDTM variables. To grow the use of FHIR for eSource beyond pilot projects, existing standards and workflows must be adapted to enable repeatable and scalable processes. New APIs will improve the flow of data from FHIR-exposed EHR content to ODM-based CRFs. This project demonstrates how data standards, workflows, and APIs can be used together to reduce implementation barriers while advancing the automation opportunities available for implementing FHIR-based eSource for clinical research. This project develops and demonstrates the use of new extensions to the CDISC ODM and CDASH standards metadata to enable information systems to retrieve and apply FHIR resource content to serve a variety of eSource and RWE use cases.

INTRODUCTION

In this project we develop a prototype that enables a site coordinator to drive the integration of patient data from an EHR into a clinical research EDC system eliminating the need to manually re-enter this data. This project extends a previous version that sought to test the feasibility of mapping HL7 FHIR into the CDISC CDASH and SDTM data standards. This project adds the use of the FHIR API to pre-populate CRFs in an EDC system. This paper references the previous Research on FHIR project paper (Zopf, Abolafia, & Reddy, 2017) as it provides useful background information and supplements the topics covered in this paper.

Pre-populating CRFs using patient data available in EHR systems, also called eSource, has long been a goal of clinical researchers seeking to improve data collection efficiency and quality. Researchers, including many that use the CDISC ODM standard (Hume, Aerts, Sarnikar, & Huser, 2016), have tested a variety of eSource approaches. Currently, most clinical research studies redundantly enter data available in Meaningful Use conformant EHR systems into research focused EDC systems. The EHR data represents original patient records that in many cases are the same information needed to complete the EDC-based CRFs for the research study. Today, integration between EHR and EDC systems predominantly relies on redundant data entry.

Pre-populating CRFs with EHR data pilots go back over a decade, but the difficulty setting up the data exchange process and mapping the EHR data has ensured that these previous efforts have not evolved beyond pilots. HL7 FHIR and the CDISC ODMv2 standard promise to provide the infrastructure needed to scale this process towards common usage. This project (1) develops a prototype that examines how these new standards can work together to advance the state of pre-populating CRF forms with EHR data, (2) identifies gaps in automation using these new standards, and (3) recommends extensions to existing CDISC standards to better support eSource processes.

BACKGROUND

CDISC CLINICAL DATA ACQUISITION STANDARDS HARMONIZATION (CDASH)

The CDISC CDASH standard defines clinical trial data collection standards and describes the implementation of CRFs that apply the standard (CDISC, 2018b). CDASH establishes a common way to collect data across studies and sponsors that aligns the data for use in SDTM. Major new versions of the CDASH standard, the CDASH Implementation Guide v2.0 and the CDASH Model v1.0, were released in late 2017.

CDISC OPERATIONAL DATA MODEL (ODM)

The CDISC ODM standard is a vendor-neutral, platform-independent data exchange standard for clinical and translational research data, metadata, administrative data, reference data, and audit trail information (CDISC, 2018c). ODM supports the regulatory-compliant acquisition, archival and exchange of study metadata and data. Used as much for metadata as data, ODM has become the language of choice for representing case report form metadata and has been implemented in many data capture software tools. CDISC publishes CDASH metadata as

ODM to provide a machine-readable representation of this foundational standard. ODMv2 is a major new version of the standard currently under development. ODMv2 will provide the specification for a standardized REST API, new mechanisms for representing the semantics found in controlled terminologies, and the means to represent metadata to support retrieving FHIR-exposed EHR data.

ODM APPLICATION PROGRAMMING INTERFACE (API)
The ODMv2 API provides a standard specification for exchange ODM content, including metadata, data, reference data, and administrative data. The purpose of the ODMv2 API is to promote dynamic, software-driven data exchange between clinical research applications. The ODMv2 API seeks to promote dynamic messaging between a wide range of compliant clinical research data applications. The first version of the ODM API will be released with ODMv2 and will support XML, JSON, and RDF media-types.

CDISC SHARE
CDISC SHARE is a cloud-based CDISC standards metadata repository used to curate, manage, and publish standards metadata in machine-readable formats (Hume et al., 2018). CDISC SHARE is a curated resource that simplifies the retrieval and implementation of CDISC standards metadata in clinical research information systems such as clinical data management systems, mobile apps, and learning health systems (CDISC, 2018a). In addition to providing a resource to software developers, data managers, and biostatisticians, CDISC SHARE also supports the standards development process by enabling the implementation of new software tools and quality improvement mechanisms.

HL7 FAST HEALTHCARE INTEROPERABILITY RESOURCES (FHIR)
HL7 published FHIR Release 3 in March 2017 (HL7, 2018) as a standard for exchanging healthcare information electronically using a REST API and data organized into Resources. Resource content provides a focused, independent block of information that stands alone, or when combined with other Resources forms a useful patient record. Resources are new with FHIR, but their development has been informed by the previous generations of HL7 standards. FHIR seeks to broaden the scope of data sharing across organizations, disciplines, devices, and platforms, as well as enabling developers to integrate data faster than previously possible. FHIR aims to maintain relative simplicity by addressing the most common 80% of healthcare data exchange needs, while leaving the other 20% to extensions. FHIR has been developed based on today's web technologies to maintain a low learning curve, leverage existing software libraries, and maximize implementation feasibility.

ESOURCE
The FDA defines electronic source data, or eSource, as data initially recorded in electronic format, and EHR systems may function as an eSource data originator (FDA, 2013). The EHR functions as an electronic source when software transfers EHR data into an EDC system's CRF (FDA, 2013). The term eSource often refers to the pre-population of EDC-based CRFs using EHR data (Mitchel, 2015). The FDA claims eSource benefits may include: eliminating unnecessary data duplication, reducing transcription errors, capturing source data at the time of the patient visit, promoting remote monitoring of data, encouraging real-time data reviews, and improving the accuracy and completeness of the data (FDA, 2013). The TransCelerate BioPharma eSource Initiative survey showed near consensus on the importance of clinical data standards in realizing EHR eSource integration for clinical research (Kellar et al., 2016).

REAL WORLD EVIDENCE (RWE)
RWE can come from a variety of sources including EHRs, payer administration claims, and patient registries (Sherman, 2016). One of the challenges in obtaining RWE from existing data sources stems from the diverse array of applications used to collect and store healthcare data. The HL7 FHIR standard can serve as an interface between diverse EHR applications delivering real world data (RWD) in support of RWE. The FDA currently uses RWE to monitor post-market safety and adverse events. The FDA has also indicated that data derived from RWE, under the right circumstances, can be used to support a marketing application. Furthermore, the 21st Century Cures Act, passed in 2016, emphasizes the use of RWE to support regulatory decisions.

SYNTHEA SYNTHETICMASS DATASET
The SyntheticMass simulated Health Information Exchange (HIE) open-source dataset populated with EHR data for over 1 million simulated residents of Massachusetts statistically mirrors the real population in terms of demographics, disease burden, vaccinations, medical visits, and social determinants (Zopf et al., 2017). The Synthea synthetic patient population simulator generated the synthetic patients within SyntheticMass dataset. Synthea generates synthetic (not real), but realistic patient data and associated health records in several formats, including HL7 FHIR 3.0.1 (MITRE, 2017).

HEALTHCARE SERVICES PLATFORM CONSORTIUM SANDBOX
In 2013 Intermountain Healthcare, Louisiana State University, and the U.S. Department of Veteran Affairs founded the Healthcare Services Platform Consortium (HSPC) to refocus how healthcare applications are developed. Today over 270 contributors have joined the consortium to help develop a healthcare services platform community that

supports the development of interoperable applications existing in a service-oriented architecture and knowledge-enabled model. HSPC's mission is to improve patient health by creating an "open ecosystem of interoperable applications, knowledge, content, and services" (HSPC, 2017). The HSPC platform enables developers to create sandboxes consisting of an isolated environment, a partial SyntheticMass dataset, and software tools that support the development of healthcare applications.

METHODS

CRF PRE-POPULATION PROTOTYPE: AN OVERVIEW OF THE RESEARCH ON FHIR ADAPTER

This project developed and demonstrated new features of the CDISC ODM standard to reference eSource content using the HL7 FHIR data exchange standard. The software prototype, called the Research on FHIR Adapter (RoF Adapter), retrieves the FHIR resource mapping for specified CDASH variables using the draft ODMv2 API, generates FHIR API requests based on the resource mapping metadata to retrieve patient data from the EHR, and posts this mapped patient data to the EDC system using the draft ODMv2 API. Draft ODMv2 metadata describes the mapping of CDASH variables to FHIR resource properties. The RoF Adapter prototype uses the FHIR API to retrieve patient data from a sandbox EHR provided by HSPC and populated with data from the open-source Synthea Synthetic Patient Platform. The RoF Adapter takes the patient data retrieved from the HSPC sandbox and populates the ODM and CDASH-based CRFs. The draft ODMv2 API simulates an EDC system for the purposes of this project. **Figure 1** shows this basic process.

RESEARCH ON FHIR ADAPTER FOR PRE-POPULATING CRFS

Figure 1. High-level data flow for RoF Adapter

The RoF Adapter application implements the same study criteria that the initial project paper (Zopf et al., 2017) used to identify adult subjects with Diabetes Mellitus Type 2 and at least two encounters in the HSPC Synthea data. Table 1 in the previous paper (Zopf et al., 2017) lists the specific selection criteria and the CDASH fields targeted for FHIR data extraction. However, the current project retrieves the Synthea patient data from an HSPC sandbox using the FHIR API, instead of running a simulation using a Ruby script to clone the Synthea database. Table 2 in Zopf et al. (2017) provided the mapping metadata for populating CDASH variables using FHIR resources and attributes.

This project also used new draft FHIR resources, *ResearchStudy* and *ResearchSubject*, proposed to support the secondary use of EHR data for clinical research. The RoF Adapter application generated these resources and posted them to the HSPC sandbox using the FHIR API. RoF Adapter data collection forms captured the data needed to generate both new resources. Each *ResearchSubject* FHIR resource required an EHR patient be matched to an EDC subject by the site coordinator. The patient and subject lists were retrieved using the FHIR and ODMv2 APIs, respectively.

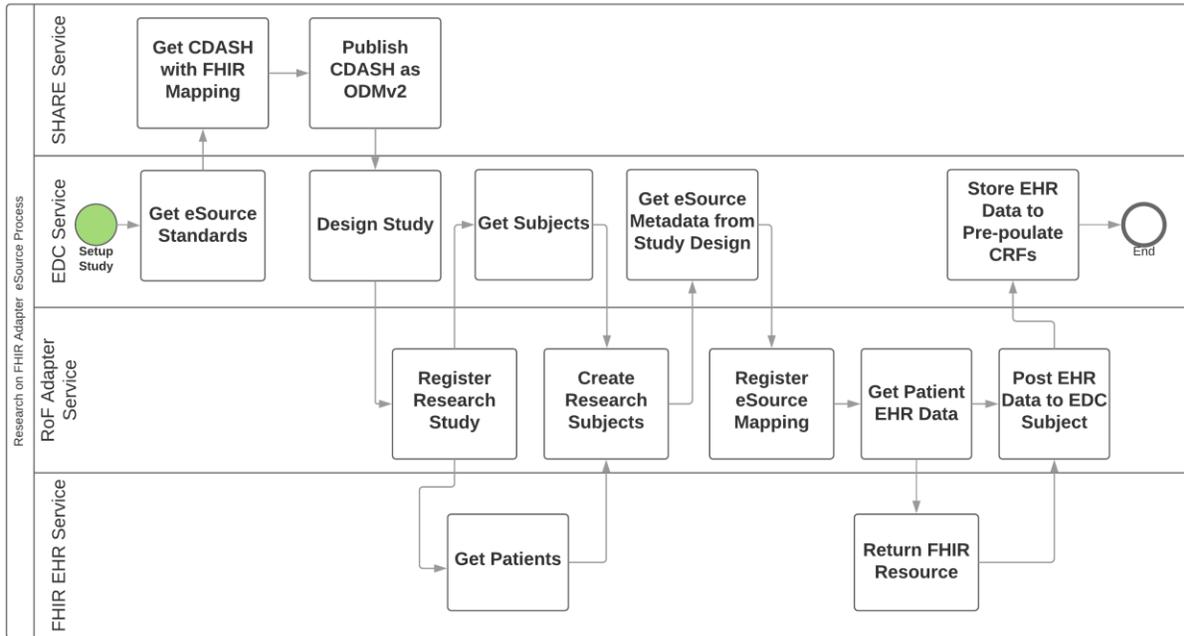RESEARCH ON FHIR ADAPTER ESOURCE PROCESS

Figure 2. RoF Adapter eSource process diagram

Figure 2 shows a more detailed view of the process implemented by the RoF Adapter eSource application. FHIR, ODMv2, and CDISC SHARE REST APIs provide the service interfaces that drive the data flow for this process. Although an EDC system might retrieve the study metadata indirectly from CDISC SHARE, this simplified view shows the EDC study design tool retrieving the standard study metadata directly using the CDISC SHARE API. This study metadata includes the mapping metadata needed for the RoF Adapter to retrieve the patient data needed to pre-populate a CDASH variable using the FHIR API. Prior to using this mapping metadata to retrieve EHR data, the RoF Adapter retrieves EDC subjects using the ODMv2 API and EHR patients using the FHIR API and the site coordinator does the matching using the RoF Adapter. The patient-subject matching shown in **Figure 2** can occur before or after the registration of eSource mapping metadata step since these steps are independent of one another.

The RoF Adapter prototype was developed using Python 3.6.2 and Flask 0.12.2. RoF Adapter used the Python SMART on FHIR client 3.0.0 (Wiffin, Mandel, Schwertner, Pfiffner, & Baranika, 2018) to retrieve and access EHR content. The Python *fhirclient* works with FHIR servers that support the SMART on FHIR protocol and was used to access the HSPC sandbox FHIR server. The HSPC sandbox with Synthea data implements FHIR v3.0.1 and is openly accessible at https://api-stu3.hspconsortium.org/phusefhir/open. The RoF Adapter used *fhirclient* to retrieve JSON-encoded FHIR resources from the HSPC sandbox and post JSON content to the EDC service using the ODMv2 API.

MAPPING FHIR RESOURCE PROPERTIES TO CDASH VARIABLES
As shown in **Figure 2**, after the EDC service retrieves the FHIR mapping metadata using the CDISC SHARE API the RoF Adapter retrieves the mapping metadata from the EDC service using the ODMv2 API. **Figure 3** shows the Height data collection variable in ODM including extensions for the FHIR resource and attribute mapping metadata. **Figure 3** also shows an ODM extension that represents the coding for the EHR content for this variable. The RoF Adapter uses the FHIR resource and attribute metadata with the LOINC coding to generate the API request to retrieve the EHR content to pre-populate this variable for a subject. The *odmv2* prefix shows the new, extended ODM content used by the RoF Adapter.

```
1.   <ItemDef OID="ODM.IT.VS.HEIGHT.VSORRES" Name="Height" DataType="float">
2.     <Description>
3.       <TranslatedText xml:lang="en">
4.         Result of the vital signs measurement as originally received or collected.
5.       </TranslatedText>
6.     </Description>
7.     <Question>
8.       <TranslatedText xml:lang="en">Height</TranslatedText>
9.     </Question>
10.    <Alias Context="CDASH" Name="HEIGHT.VSORRES"/>
11.    <Alias Context="CDASH/SDTM" Name="VSORRES+VSORRESU"/>
12.    <odmv2:Origin Type="Collected" Source="Investigator">
13.      <odmv2:FHIR Resource="Observation" Attribute="valueQuantity.value"/>
14.    </odmv2:Origin>
15.    <odmv2:Coding Code="8302-2"  CodeSystem="urn:oid:2.16.840.1.113883.6.1"  CodeSystemName="LOINC"
       CodeSystemVersion="2.61"/>
16.  </ItemDef>
```

Figure 3. ODMv2 extensions demonstrating FHIR mapping and eSource coding semantics

RETRIEVING EHR DATA USING THE HSPC FHIR API

The RoF Adapter uses the ODMv2 metadata shown in **Figure 3** to generate FHIR API calls using the Python *fhirclient*. **Figure 4** shows a simple *fhirclient* example using Python code to retrieve the EHR content from the HSPC sandbox. The RoF Adapter prototype code is more sophisticated than the example in **Figure 4**, but **Figure 4** concisely demonstrates how *fhirclient* works. Notice that the *api_base* in **Figure 4** references the HSPC sandbox that functions as the EHR service for this project. The HSPC sandbox functions as the EHR service for this project by providing a FHIR STU 3 (v3.0.1) interface and a subset of the full Synthea data. The RoF Adapter reads patient data from the HSPC sandbox and writes the new *ResearchStudy* and *ResearchSubject* resources to the sandbox.

```python
1.   from fhirclient import client
2.   import fhirclient.models.observation as ob
3.   settings = {
4.       'app_id': 'demonstration',
5.       'api_base': 'https://api-stu3.hspconsortium.org/phusefhir/open'
6.   }
7.   smart = client.FHIRClient(settings=settings)
8.   search_obs = ob.Observation.where(struct=dict(patient="SMART-665677", code="8302-2", date="2001-05-09"))
9.   print(search_obs.perform_resources(smart.server).pop().as_json())
```

Figure 4. Example Python fhirclient code needed to retrieve a specific patient height observation value

Research Subject Resource

Subject

[trace010002 ▼]

Status

[enrolled ▼]

Study Identifier

[16676]

patient

[Frank Taylor ▼]

Assigned Arm

[Treatment]

Actual Arm

[Treatment]

[Save ResearchSubject]

| Select Study | Add Sponsor | Add Investigator | Add Site | Add ResearchStudy | Add ResearchSubject |

Figure 5. RoF Adapter form for creating a basic ResearchSubject resource

THE RESEARCHSTUDY AND RESEARCHSUBJECT FHIR RESOURCES

In addition to the patient data maintained in the EHR, this project utilized the *ResearchStudy* and *ReasearchSubject* resources created to support clinical research. No test servers supporting these resources were identified, including the HSPC sandbox, when this project started. Therefore, the RoF Adapter prototype generated very basic instances of these resources to support the needed functionality. The *ResearchStudy* resource contains or references much of the fundamental information about the study needed by the site. The *ResearchSubject* references the *ResearchStudy* and performs the important task of representing the patient-subject link. **Figure 5** shows the form used by the RoF Adapter to create a basic *ReseachSubject* resource for each subject participating in the study. RoF Adapter used the set of *ResearchSubject* resources to direct the transfer of eSource data from the EHR to the EDC service. The *ResearchStudy* and *ResearchSubject* resources both come under the HL7 Biomedical Research and Regulatory (BR&R) sub-committee and are being maintained on their behalf by the BRIDG management team. These resources are under re-evaluation to finalize the specifications and be properly aligned with the current needs for clinical research.

RETRIEVING METADATA AND POSTING DATA TO AN EDC SERVICE USING THE ODMV2 API

Prior to retrieving the EHR patient data using a FHIR request, the EDC study metadata describing the CDASH variable FHIR mappings must be retrieved. The ODMv2 API supports creating, reading, updating, and deleting study metadata, data, administrative data, and reference data. The RoF Adapter reads study metadata by retrieving the content of the study *MetaDataVersion* from the EDC service. **Figure 6** shows the basic structure of the *MetaDataVersion* REST API.

MetaDataVersion   ∨

POST	`/studies/{studyOID}/mdv`	Creates a new MetaDataVersion
GET	`/studies/{studyOID}/mdv/{mdvOID}`	Returns the metadata for the specified MetaDataVersion
PUT	`/studies/{studyOID}/mdv/{mdvOID}`	Update a MetaDataVersion
DELETE	`/studies/{studyOID}/mdv/{mdvOID}`	Removes the specified MetaDataVersion element

Figure 6. The ODMv2 draft REST API for MetaDataVersion

The ODMv2 API returns the study *MetaDataVersion* content as JSON. The RoF Adapter then identifies the CRFs and variables that contain eSource mappings in the form of the FHIR metadata shown in **Figure 3** and lists them in the form displayed in **Figure 7**.

Research On FHIR eSource: Retrieve Study Metadata

Study ID: 16676

FHIR Endpoint: https://api-stu3.hspconsortium.org/phusefhir/open/

Register Study and Retrieve Study Metadata

URL to Retrieve EDC Metadata

http://127.0.0.1:8080/odm1-3-2-api/

Study OID

16676

Username

Tom Swift

Password

••••••••

Load Study Metadata

Metadata

```
ODM.IT.LB.LBDTC FHIR:Observation.effectiveDateTime
ODM.IT.LB.WBC.VSORRES FHIR:Observation.valueQuantity.value
ODM.IT.LB.WBC.LBORRESU FHIR:Observation.valueQuantity.unit
ODM.IT.LB.ALB.VSORRES FHIR:Observation.valueQuantity.value
ODM.IT.LB.ALB.LBORRESU FHIR:Observation.valueQuantity.unit
ODM.IT.LB.GLUC.VSORRES FHIR:Observation.valueQuantity.value
ODM.IT.LB.GLUC.LBORRESU FHIR:Observation.valueQuantity.unit
ODM.IT.LB.HBA1CHGB.VSORRES FHIR:Observation.valueQuantity.value
ODM.IT.LB.HBA1CHGB.LBORRESU FHIR:Observation.valueQuantity.unit
ODM.IT.LB.ALBCREAT.VSORRES FHIR:Observation.valueQuantity.value
ODM.IT.LB.ALBCREAT.LBORRESU FHIR:Observation.valueQuantity.unit
ODM.IT.CM.CMTRT FHIR:Medication.display
ODM.IT.CM.CMDOSE FHIR:MedicationStatement.doseage.doseQuantity.value
ODM.IT.CM.CMDOSFRQ FHIR:MedicationStatement.doseage.timing
ODM.IT.CM.CMROUTE FHIR:MedicationStatement.doseage.route
```

Select Study | Register Study | Transfer Data

Figure 7. RoF Adapter form for registering the study metadata

For each *ResearchSubject* resource, the Rof Adapter uses the FHIR API to request EHR data for each study variable containing eSource mapping. Each FHIR request returns a full resource. The RoF Adapter uses the FHIR attribute to retrieve the value of interest from the resource for insertion into an EDC CRF. In cases where the FHIR resource attribute is embedded in another container attribute, a "." notation is used. For example, **Figure 3** shows an attribute value of *valueQuantity.value* that indicates the *value* attribute is contained within *valueQuantity*. In cases where a FHIR request returned multiple instances of a requested resource, the RoF Adapter used the first instance.

After retrieving the targeted FHIR resource attribute, the RoF Adapter uses this value to create the associated ODM *ItemData* element. The RoF Adapter supports ODMv2 clinical data transactions at the CRF level to limit the number of requests sent to the EDC service. After retrieving the EHR data for each eSource variable defined for a specific CRF, the RoF Adapter creates and submits an ODMv2 Transaction API post to pre-populate the CRF in the EDC system. **Figure 8** shows the basic structure of the ODMv2 draft Transaction API where transactions are specified in the body of the post using ODM (CDISC, 2018c). Once the RoF Adapter has transferred each CRF with eSource

content for each subject, that completes the process addressed by this project. The previous project addressed mapping the data into SDTM as well as generating an annotated CRF that displays both the FHIR and SDTM mappings (Zopf et al., 2017).

Figure 8. The ODMv2 draft REST API for Transaction

RESULTS
The RoF Adapter successfully transferred patient data from a FHIR-exposed EHR service to an ODMv2 API exposed EDC service. The RoF Adapter was able to generate the FHIR API requests to retrieve patient data using the ODM-based FHIR mapping metadata for CDASH variables. The RoF Adapter retrieved the ODM metadata, including the FHIR mapping metadata, using the ODMv2 draft API. Using the ODM metadata, the RoF Adapter successfully posted the EHR patient data into the EDC service. The results of the data transfer for one patient are shown in the status form in **Figure 9**.

Figure 9. RoF Adapter data transfer status form

The new ODM elements representing FHIR references to eSource data for specified CDASH variables demonstrates a standards-based mechanism supporting studies seeking to pre-populate CRFs with EHR patient data. In addition to adding FHIR resource mapping metadata, ODM represents the terminologies found in EHR data, such as LOINC and SNOMED-CT. The *ResearchStudy* and *ResearchSubject* FHIR resources provided the EHR with information needed to transfer patient data from the EHR service into the EDC service.

Although the RoF Adapter prototype performs the essential tasks shown in **Figure 2**, fully automating all eSource data transfer tasks requires additional features. The RoF Adapter does not attempt to map or transform the eSource terminology found in the Synthea data. This data is frequently coded using terminologies not routinely used in regulated clinical research, such as SNOMED-CT and ICD-10. Patient-subject matching was a manual process occurring during the creation of a *ResearchSubject* resource. Automated matching support would simplify the site coordinator's task of mapping patients to enrolled subjects. The RoF Adapter allows the site coordinator to select the visit type and visit date to use when transferring data into the EDC service. The FHIR data request can include a date filter. More sophisticated selection requirements will require additional features. For example, in cases where multiple values exist for an observation in the EHR data, yet the EDC CRF expects only one variable, it may be necessary to display the EHR values to allow the site coordinator to select the desired data to transfer. Finally, the RoF Adapter process flows one way, providing a one-time transfer of data from the EHR service into the EDC service, and does not provide a mechanism for identifying and remedying specific data corrections in either the EHR or EDC services.

DISCUSSION

BENEFITS OF A SITE-DRIVEN APPROACH
Our prototype and workflow have the sites, as opposed to sponsors or agents of the sponsor, using the RoF Adapter to pull data from an EHR. This approach alleviates administrative and security hurdles surrounding access to a site's EHR. Accessing a site's EHR is a more difficult issue to resolve than the technology challenges presented. Additionally, a site driven approach lets sites keep doing what they do, and will not drastically alter a site's current workflow. Sites now typically use their EHRs and other paper forms as the source to enter data into a study's EDC database. Using the RoF Adapter, site personnel will be able to automate much of the manual data entry into an EDC system.

HOW TO DETERMINE WHICH FIELDS USE ESOURCE CONTENT
EHRs contain a wealth of health-related information on subjects and providers. So, for a given protocol, how do we decide which data fields in a subject's EHR can be used? First, there are administrative/security constraints. Will a site give you access to its EHR database? At this point in time, the answer is much more likely to be no than yes. If a site will give you access, what data will the site give you access to? Also, compliance with the US Health Insurance Portability and Accountability Act (HIPAA) regulations will limit the data fields that you can extract and store. Second is the issue of data quality. The quality of data in EHR systems varies considerably. As a result, one must evaluate each data field of interest to determine if it is of sufficient quality to generate scientifically valid conclusions for regulatory decision making. A future approach could involve development of clinical research profiles on top of the core FHIR Resources, which mask or remove the personally identifiable information resource attributes, thereby de-risking the use of APIs.

VISIT VS. VISITLESS (LONGITUDINAL) STUDIES
While EHRs contain a wealth of health data, data in EHRs are typically collected for the purposes of health care delivery, reimbursement and administration. Therefore, EHRs do not contain many of the protocol-related concepts needed for clinical research. Such concepts as the schedule of events, arm, and protocol-defined visits are rarely part of a subject's EHR. This could limit data usability dependent on study design as protocols typically call for evaluating safety and efficacy parameters at protocol-defined visits. However, in many cases, visit or assessment dates could be used instead of visit number or serve as a proxy for visit number. In a recent kidney transplant study, lab values were extracted from EHRs pre and post-surgery. While we were not able to get visit number out of the EHR, we were able to get assessment dates, and calculate the number of days from surgery to lab assessment. For this study design and indication, we were able to meet the goals of the study without visit number. Increasingly data collection in clinical studies is changing toward more of a continuum model, whereby the data (in person in site) is augmented with streams of non-site data, such as mobile health devices or electronic patient reported outcomes. The significance of a 'visit' in studies of the future may be diminished; the industry will need to plan for this eventuality.

ESOURCE AND CONSENT
In addition to the technical, administrative, and security requirements for accessing EHR data for research, there are legal requirements for using EHR data for research. The bottom line is that patients control the use of their data. The HIPAA Privacy Rule requires that data contained in a subject's EHR cannot be used for a research project unless the subject provides informed consent. This is a tremendous barrier to overcome when using EHR data retrospectively. When the goal is to use EHR data for a prospective study, typically institutional review boards require that the data is only used for the study that the subject consents to, and not for secondary uses. However, an update to HIPAA now allows subjects to give consent for future research, as well as to withdraw consent. The new European General Data Protection Regulation (GDPR) is similar to HIPPA in that it requires specific consent from subjects regarding the use of their personal data, but also makes certain allowances for research. This should remove some of the burden of having to re-contact subjects and obtain consent for each new research project. We expect that the outcome of the recently (Apr 2018) initiated pilot by the Sync4Science (S4S, 2018) project will help inform the wider consent narrative, specifically looking at how the patient can be more involved and more informed about what data is being used and by whom. FHIR already includes a *Consent* resource, and the SMART-on-FHIR project is making use of this.

HEALTHCARE / RESEARCH TERMINOLOGY DIFFERENCES
In our pilot project we successfully mapped all but one variable of interest from FHIR to CDASH/SDTM. However, this required a manual process that is not scalable and therefore limits the use of EHR data in a real world clinical trial setting. While most of our variables of interest are represented in FHIR and CDASH/SDTM, in many cases they lack harmonization. Most concepts have variable names that are different in FHIR and CDASH/SDTM and contain controlled terminology that is not harmonized. For example, the concept "administrative gender" is called "gender" in FHIR and has controlled terminology "male" and "female". While in CDASH/SDTM it is called "SEX" and has controlled terminology 'M' and 'F'. Furthermore, coding dictionaries often differ in healthcare and research databases.

In addition, there will need to be a strategy for the pre- and post-coordinated terminology concepts. Clinical research generally uses post-coordinated terminologies; this contrasts with the primarily pre-coordinated terminologies in use in routine healthcare. Laboratory and vital sign parameters are specified in the Synthea FHIR data by a LOINC code. A given LOINC code provides a description of the parameter, the unit of measurement, the specimen type, the method of the test and the category for the test or vital sign. Given the lack of harmonization between FHIR and CDASH/STDM controlled terminology, a single LOINC code had to be manually mapped to five distinct CDASH/STDM variables. This will have implications for Source Data Verification strategies, as the pre-coordinated source data would need to be decomposed to be reported in a CRF. Can the decomposed elements also be classed as eSource? This is an important question to answer.

Medications and medical diagnoses were only available in SNOMED terminology and ICD-10 coding in the Synthea database. This is consistent with most EHR systems that commonly capture diagnoses using ICD-10 coding for billing purposes. In the pharmaceutical industry, adverse event data related to drug exposure is provided to regulatory authorities using MedDRA terminology and for medications using the WHO Drug Dictionary. A new variable to capture the original SNOMED or ICD-10 terminology would need to be created, along with an explanation of the process to convert SNOMED/ICD-10 to MedDRA and/or WHO Drug coding. A standard approach to "translation" between dictionaries would greatly enhance the ability to use RWE in drug development.

It is also worth noting that we also used a single EHR database for our pilot, the Synthea database. The US lacks a universal health care system and fully interoperable data standards for medical record systems. There is a great deal of variability between EHR systems. Differences in data standards, data collection/entry, and data quality add additional complexities for multi-site and multi-geographical region clinical studies. Much of the ability to execute on this will hinge on open accessibility of terminology systems using APIs pointing to persistent URIs; it is an example where machine learning/artificial intelligence will shine.

OPEN, STANDARD APIS

Open, standard APIs for clinical research will drive improvements in data exchange by increasing the levels of automation possible and simplifying access to EHR and EDC content. The FHIR and ODMv2 APIs are essential to creating a RoF Adapter style application that can be used with different EHR and EDC systems. **Figure 2** clearly shows the importance the REST APIs have in exposing the services needed to support eSource. Open, standard APIs like FHIR and ODMv2 make it possible for eSource applications to scale beyond point-to-point integrations that work integrating one EHR to one EDC system. The FHIR and ODMv2 APIs should make it possible to create applications that communicate to all conformant EHR and EDC systems, respectively.

The mapping burden and additional study setup historically required to conduct eSource studies has limited broad acceptance of eSource implementations. The RoF Adapter prototype achieved the primary project goals by using new extensions to the CDISC standards to generate calls to the FHIR and ODMv2 APIs. This process increases the automation available to studies using eSource. The use of open, standard APIs will simplify eSource implementations. Three APIs were involved in this project: CDISC SHARE, ODMv2, and FHIR. A deliverable of this project was to add the eSource mapping metadata into CDISC SHARE and make it available via the CDISC SHARE API. These enhancements are under development by the CDISC SHARE team. Open, standard APIs, when combined with the CDISC and HL7 data standards and associated workflows, promise to significantly improve the state of data interchange between routine healthcare and clinical research.

PROTOTYPE LIMITATIONS

Currently, the RoF Adapter allows the site coordinator to select a visit from the EDC metadata to load EHR data. It also provides a basic ability to filter EHR data by event date. Simple date filtering and visit type selection may not be sophisticated enough to map the event-driven healthcare data into a prospective protocol-driven study design. Manual support from site coordinators or more sophisticated mapping automation maybe required. For this reason, many RWD studies do not use patient visits, and pragmatic studies may use very simple visit structures to simplify the process of moving EHR data into the EDC service. In a study of normal complexity, the RoF Adapter may need to load the EDC system to aid the site coordinator in determining where in the study the data should be loaded. In cases where multiple EHR observations exist that match the eSource mapping metadata, those values may need to be displayed to allow the site coordinator to choose the data that is transferred into the EDC system. The availability of data and metadata from the source and target systems should make the building of 'expert' systems possible which will be able to use machine learning/artificial intelligence techniques to streamline this bridging.

Not all mapped FHIR resources were available to retrieve the EHR data, and in some cases, alternatives were used in place of the missing resources. For example, many of the concomitant medication variables were mapped to the properties in the *MedicationStatement* resource. However, for many patients no *MedicationStatement* resources were available. The lack of uniform availability of the *MedicationStatement* in the Synthea data is likely due to the fact that it is not part of the prescribe/dispense/administer sequence, but instead is a report from a patient, significant other, or clinician that one or more of the actions in this sequence have occurred (HL7, 2018). In these cases, the *MedicationAdministration* resource was used instead. As future versions of this project test other EHR systems,

additional resource substitutions may be necessary and content nuances will need to be addressed. The availability of alternative FHIR Profile implementations, such as the US Core Patient or Structured Data Capture profiles, may also impact how the RoF Adapter maps FHIR content to an EDC service.

USE OF TEST EDC SYSTEMS IN ESOURCE STUDY SETUP TO AVOID MAPPING PROBLEMS AND DATA ISSUES
When clinical data is collected via entry into an EDC, extensive User Acceptance Testing (UAT) is performed before the system goes into production. More extensive testing may be necessary when pulling data from EHRs and pushing the data into an EDC. UAT, as well as a pilot, should be performed to avoid transmission, mapping and data issues when the system goes live. If you don't plan to perform a pilot, your initial pull/push at a given site is likely to serve as a pilot. We recommend setting up a test instance of both EHR and EDC databases. The software and process should be tested in-house on dummy data (or perhaps real data if the data is used retrospectively) before anything is deployed to the sites. It is important to ensure that all partners are onboard with the decisions – this will be a new paradigm for Sites, Sponsors, Contract Research Organizations, Vendors and Regulators. As shown by this and the preceding project – the technology exists; it will be a change to existing processes and that can often be very disruptive.

When taking a site-driven approach (as described above), an installation qualification (IQ), in addition to the typical UAT should be performed. The IQ should test whether the software executing the pull/push is installed and configured properly at each site. Next each site should perform a pilot to determine if the software works as documented and that the EDC data extracted from EHRs complies with specifications in the Data Management Plan. At a minimum, this would include checking that the system functions as documented, subjects and variables are extracted as expected and variable mappings are executed according to specifications.

CONCLUSION

This project extends the conclusions drawn from the previous Research on FHIR project by introducing the use of open, standard APIs and automating eSource data retrieval tasks. The RoF Adapter prototype demonstrated populating CRF fields of interest using ODMv2 / CDASH metadata and the FHIR and ODMv2 APIs. Taken together, the prototype software in these two projects could retrieve patient data from an EHR, post that data to an EDC system, and then convert the EDC data to SDTM. As we incrementally address the limitations discussed in the previous section, more complete process automation will be possible.

Another outcome of the project was to test the use of draft ODMv2 features. As draft features of ODMv2, these features are likely to change prior to final release, but support for FHIR eSource is an ODMv2 priority. Support for the FHIR metadata mapping will be included in the publically available ODMv2 draft, as will the ODMv2 API. The CDISC SHARE API mechanisms to retrieve the CDASH FHIR mapping are part of the CDISC SHARE v2.x development plans. Together, these new standards features support the use of FHIR for eSource simplifying future implementations. Combining the open, standard ODMv2 and FHIR APIs should have significant long-term impact on moving eSource studies from pilots to widespread adoption. Future versions of this project will seek to address gaps in the standards that limit FHIR-based eSource process automation.

Future work should include testing on multiple EHR systems. Major EHR vendors, such as EPIC and Cerner, have FHIR server sandbox environments that should be investigated as suitable test EHR platforms. The current development draft version of ODMv2 used in this project will likely be updated in the near future. Additional testing against future versions of the ODMv2 API will be necessary, including testing of EDC applications that support the new API.

REFERENCES

CDISC. (2018a). CDISC SHARE. Retrieved from https://www.cdisc.org/standards/share
CDISC. (2018b). Clinical Data Acquisition Standards Harmonization (CDASH). Retrieved from https://www.cdisc.org/standards/foundational/cdash
CDISC. (2018c). Operational Data Model (ODM). Retrieved from https://www.cdisc.org/standards/transport/odm
FDA. (2013). *Guidance for Industry: Electronic Source Data in Clinical Investigations*. FDA Retrieved from https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm328691.pdf.
HL7. (2018). FHIR Release 3. Retrieved from http://hl7.org/fhir/STU3/index.html
HSPC. (2017). Healthcare Services Platform Consortium Sandbox. Retrieved from https://sandbox.hspconsortium.org
Hume, S., Aerts, J., Sarnikar, S., & Huser, V. (2016). Current applications and future directions for the CDISC Operational Data Model standard: A methodological review. *Journal of Biomedical Informatics*. doi:doi:10.1016/j.jbi.2016.02.016
Hume, S., Chow, A., Evans, J., Malfait, F., Chason, J., Wold, D., . . . Becnel, L. (2018, March 15, 2018). *CDISC SHARE, a Global, Cloud-based Resource of Machine-Readable CDISC Standards for Clinical and Translational Research.* Paper presented at the AMIA Informatics Summit 2018, San Francisco, CA.

Kellar, E., Bornstein, S. M., Caban, A., Célingant, C., Crouthamel, M., Johnson, C., . . . Wilson, B. (2016). Optimizing the Use of Electronic Data Sources in Clinical Trials:The Landscape, Part 1. *Therapeutic Innovation & Regulatory Science, 50*(6), 682-696. doi:10.1177/2168479016670689

Mitchel, J. T., Helfgott, J., Haag, T., Cappi, S., McCanless, I., Kim, Y.J., Choi, J., Cho, T., Gittleman, D.A. (2015). eSource Records in Clinical Research. *Applied Clinical Trials*.

MITRE. (2017). SyntheticMass.   Retrieved from https://syntheticmass.mitre.org/index.html

S4S. (2018). Sync for Science.   Retrieved from http://syncfor.science/

Wiffin, E., Mandel, J., Schwertner, N., Pfiffner, P., & Baranika, T. (2018). Python SMART on FHIR client. github.com. Retrieved from https://github.com/smart-on-fhir/client-py

Zopf, R., Abolafia, J., & Reddy, B. (2017). *Use of Fast Healthcare Interoperability Resources (FHIR) in the Generation of Real World Evidence (RWE)*. Paper presented at the PhUSE Annual Conference 2017, Edinburgh, Scottland. https://www.cdisc.org/sites/default/files/resource/Use_of_Fast_Healthcare_Interoperability_Resources_in_the_Generation_of_Real_World_Evidence.pdf

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the team co-lead at:

Trisha Simpson
UCB
8010 Arco Corporate Drive
Suite 100
Raleigh, NC 27617
+ 919-767-3107
Trisha.Simpson@ucb.com

Brand and product names are trademarks of their respective companies.