
Research and Applications

Automated production of research data marts from a canonical fast healthcare interoperability resource data repository: applications to COVID-19 research

Leslie A. Lenert,^{1,2} Andrey V. Ilatovskiy,^{1,2} James Agnew,³ Patricia Rudisill,^{1,2} Jeff Jacobs,² Duncan Weatherston,³ and Kenneth R. Deans Jr²

¹Biomedical Informatics Center, Medical University of South Carolina, Charleston, South Carolina, USA, ²Health Sciences South Carolina, Columbia, South Carolina, USA, and ³Smile CDR, Toronto, Ontario, Canada

Corresponding Author: Leslie A. Lenert, MD, MS, FACP, FACMI, Data Science and Informatics, Medical University of South Carolina, 22 West Edge Suite 13, Charleston, SC 29425, USA; Lenert@musc.edu

Received 3 March 2021 Accepted 14 May 2021 10052021

ABSTRACT

Objective: The rapidly evolving COVID-19 pandemic has created a need for timely data from the healthcare systems for research. To meet this need, several large new data consortia have been developed that require frequent updating and sharing of electronic health record (EHR) data in different common data models (CDMs) to create multi-institutional databases for research. Traditionally, each CDM has had a custom pipeline for extract, transform, and load operations for production and incremental updates of data feeds to the networks from raw EHR data. However, the demands of COVID-19 research for timely data are far higher, and the requirements for updating faster than previous collaborative research using national data networks have increased. New approaches need to be developed to address these demands.

Methods: In this article, we describe the use of the Fast Healthcare Interoperability Resource (FHIR) data model as a canonical data model and the automated transformation of clinical data to the Patient-Centered Outcomes Research Network (PCORnet) and Observational Medical Outcomes Partnership (OMOP) CDMs for data sharing and research collaboration on COVID-19.

Results: FHIR data resources could be transformed to operational PCORnet and OMOP CDMs with minimal production delays through a combination of real-time and postprocessing steps, leveraging the FHIR data subscription feature.

Conclusions: The approach leverages evolving standards for the availability of EHR data developed to facilitate data exchange under the 21st Century Cures Act and could greatly enhance the availability of standardized datasets for research.

Key words: Federated on FHIR architecture, FHIR Subscription, OMOP, PCORnet, data model, ETL, COVID

INTRODUCTION

The COVID-19 pandemic has illustrated the need for reliable rapidly accessible data from electronic health record (EHR) systems for research on risk factors, predictive models, and evaluation of emerging diseases. Moreover, the lack of reliable large datasets has led to

spurious research findings early in the COVID-19.¹ Two of the largest consortia leverage existing infrastructure for shared data collaboration. The National COVID Collaborative Cohort (N3C)² is an alliance among Clinical and Translational Research Grant Awardees sponsored by the National Center for Advancing Translational Sci-

ence (NCATS). This network leverages past experiences and infrastructure from the Accrual to Clinical Trials Network.³ N3C's preferred data model for accepting results is the Observational Medical Outcomes Partnership (OMOP) model maintained by the Observational Health Data Sciences and Informatics (OHDSI) collaborative.⁴ However, N3C accepts data in a variety of formats. A second consortium is based around the Patient-Centered Outcomes Research Network (PCORnet)⁵ and leverages prior investments on comparative effectiveness research across this large research network⁶ to create its database. There also are private large research networks, for example, TriNetX,⁷ that maintain a large data network of patients with COVID-19 for research from its clinical trial eligibility network.⁸ The FDA maintains several large networks for safety evaluation of drugs and devices that are also being applied to problems identified during the pandemic.⁹ In addition, some of the same partners in N3C are also using the integrating informatics for integrating biology and bedside (i2b2) platform¹⁰ to study COVID-19. Many networks have overlapping membership, and, as a result, members have to maintain duplicative data production processes.

As the pandemic has evolved rapidly, so have the requirements for rapid data updating in these large networks. Minimizing the lag between production of the data through the care of patients using an EHR system and its availability for research increases the relevance of the network to the evolving set of problems seen with COVID-19. In the prior modes of operation, Medical University of South Carolina (MUSC)'s ACT and PCORnet data networks could be used for data operations with lags of 3 or more months for production and cycles for new releases of datasets every quarter. In the COVID era, the specifications for the N3C network call for 2-week production cycles for data releases and 1-month lags between the closure of a record and its availability within the network. More current data might be even more valuable as new variants of the virus and new therapies emerge. This is a challenging task that requires the automation of processes for analytic database production.

Production of data for each network is, in itself, a multistep pipeline process that involves mapping and transformation of data to the preferred data model of a research network. Work for data production for different networks is often done in parallel, which is logistically challenging and consumes limited resources. Sometimes work is done in series, mapping from source data to 1 data model, and then another, which could potentially result in a loss of data through compression or inaccuracies in mapping. In this article, we describe the use of the Fast Healthcare Interoperability Resources (FHIR) standard data model as a *canonical model* for initial storage of the data for subsequent transformation to more analytics-oriented models (OMOP and PCORnet) as well as an architecture for multiple simultaneous largely automated translations from FHIR to these 2 CDMs. This is a particularly important task as the 21st Century Cures Act¹¹ will require availability from EHRs in FHIR standards for the United States Core Data for Interoperability standards,¹² which, while evolving, already has many of the required elements for research CDMs.

MATERIALS AND METHODS

The approach taken to standardize a data production pipeline for multiple analytic CDMs from FHIR builds on 1 of the central tools for the FHIR paradigm: a clinical data repository (CDR) designed to store, persist, retrieve, and deliver FHIR resources. A widely used implementation for this operation is the open-source HAPI FHIR engine.¹³ We built our system based on the Smile CDR platform¹⁴

that is powered by the HAPI FHIR engine. This platform can accept data in a variety of formats (JSON or XML-encoded FHIR objects, HL7 v2.x messages, flat comma-delimited files) and transform these data elements to FHIR resources that are stored in a proprietary relational format, for efficient search and retrieval. Alternatively, some vendors persist FHIR resources using a "data lake" approach, with extensive indexing but a minimal transformation of the data.¹⁵

A standards-specified feature of FHIR CDRs is automated tooling to allow *subscriptions* to specific FHIR resources.^{14,16} Subscriptions in the FHIR standard are triggers attached to FHIR data resources. Creating or updating a resource triggers a function that allows copying and transmission of the resource data object to another system. A common use for subscriptions in the FHIR standard is for notification of events. For example, if a patient is registered in an emergency department, a new FHIR resource with the registration information is created, and this then triggers sending a copy of the FHIR resource to another system via FHIR API with JSON payload or other interoperability protocol. This results in the second computer system becoming "aware" of the notification of registration.

We adapted this feature for use in data transformation in our "Federated on FHIR (F-on-F)" architecture.¹⁷ F-on-F is an architecture that replaces Health Sciences South Carolina (HSSC)'s legacy cross-institutional integrated data repository with a series of linked FHIR data repositories with a single centralized master person index maintained by subscriptions. F-on-F also uses subscriptions to admission discharge and transfer data for admission/discharge/transfer (ADT) notification and for automated updating of local repositories with centralized data on mortality, geocodes, and social determinants of health. A full description of F-on-F is beyond the scope of this article.

At the individual site level, whenever a new FHIR data resource is created in the clinical FHIR repository in our system, a copy of the resource is created in a second linked FHIR repository using the subscription mechanism. However, rather than persisting the object in the proprietary database format of the vendor, we codeveloped with Smile CDR rule-based transformations implemented in Java to persist data in 2 different targeted analytical models—OMOP and PCORnet. This results in a system of linked CDMs updated within milliseconds. The approach is illustrated in [Figure 1](#). Both OMOP and PCORnet models are extended to deal with identified data and data elements not covered by the CDM specifications. These are kept in separate tables to preserve the functioning of CDM quality inspection and analytics software; the approach also preserves subject anonymity. The specific examples of FHIR patient resource mapping to OMOP CDM tables and extensions are illustrated in [Supplementary Table S1](#).

The real-time transformation to analytic CDMs poses additional obstacles due to the transactional nature of the EHR data that evolves and expands for days or even weeks after a given patient encounter, while analytic data models assume a static self-consistent set of data. For example, the OMOP model assumes that each patient has a date of birth (known at least with a year's precision); in the EHR data the demographic details might be missing and such patients should be removed from the OMOP instance. Another example, in the OMOP model, both visits and associated clinical facts have patient IDs, allowing transitivity violations that occur in the EHR data due to patient merges. If data are incomplete or inconsistent at some point in time and evolve to completeness and correctness as time passes, the data in the live CDM instances are kept in sync with the source and all updates will be propagated via pipeline.

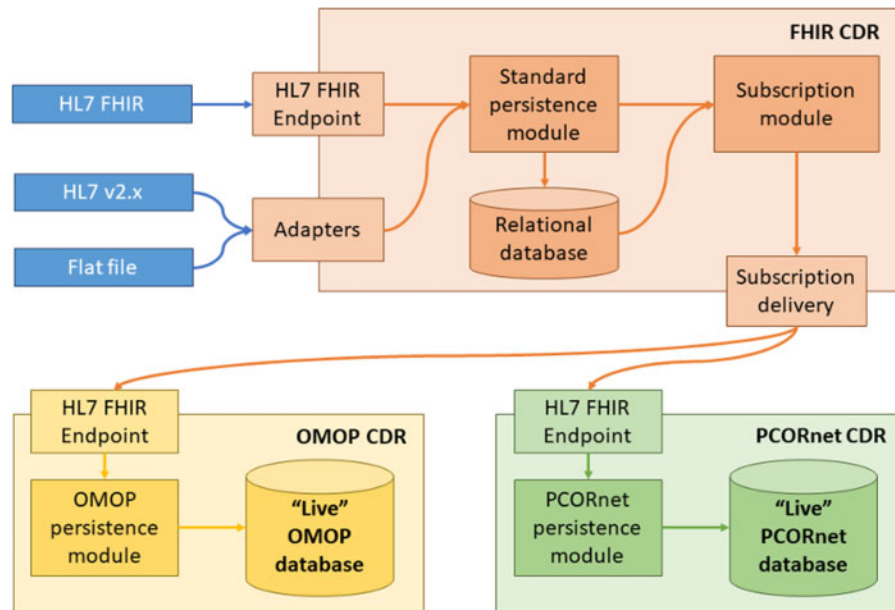


Figure 1. Adapting a FHIR CDR for real-time ETL to OMOP and PCORnet CDMs.

The issue is minimal for study feasibility queries; however, for longitudinal data analysis these inconsistencies need to be resolved. The solution we adopted was a production pipeline with separate static OMOP and PCORnet instances for postprocessing to reconcile the data.

Another issue in design was maintaining the robustness of the process to evolutionary changes in the CDMs. Ongoing changes in CDMs are the rule rather than the exception. The OHDSI group releases a new set of OMOP vocabularies weekly, with changes ranging from adding a few new concepts to complete redesign of the domain organization. Many other networks require frequent vocabulary updates (eg, once a month for N3C). In case of a major change, the postprocessing using an Extract Transform and Load (ETL) approach implemented via SQL was flexible enough to accommodate rapid changes in vocabularies, in contrast to the Java code transformations used in subscription-based mappings. Essentially, the postprocessing allows a quasi-incremental approach to the vocabulary updates: the clinical tables built with the old vocabularies (within the pipeline) are combined with a fresh set of vocabularies and all deviations from the new vocabularies are corrected, with the majority of the data being untouched. This flexibility requires preserving enough “rawness” of the clinical data so updates to an analytic CDM do not require a complete rebuilding of the database. The drawback is that “live” versions of the database produced through subscriptions cannot be used for complex analyses without postprocessing (although, again, quick study “feasibility” queries, such as counts of patients with specific e-phenotypes, are possible).

Data security is maintained using a variety of approaches. SmileCDR supports direct queries of the FHIR database using Smart-on-FHIR authentication.¹⁸ OMOP and PCORnet databases have been extended with patient data elements in separate data tables. Databases retain their original clinical temporal labels and, as such, are not truly deidentified datasets. Access to data is controlled by governance, including investigator data use agreements and by honest brokers who produce datasets based on institutional review board-approved protocols. Additionally, PCORnet and OMOP ac-

cess tools (SAS and Atlas) are restricted to the standard CDM tables with limited identifying information.

To test the feasibility of this architecture, we converted the Medical University of South Carolina research data warehouse (RDW) and operational production of PCORnet and OMOP CDMs for MUSC to use a novel process based on this concept model. Specific versions details of this implementation: FHIR version is 3.0.2; OMOP CDM version is currently 5.3.1; and the PCORnet CDM is version 6.0. The database for data management operations and for PCORnet queries is Oracle 19.6.0. Oracle tools are used for maintenance and data manipulation. The Smile CDR also uses Oracle to persist the data but it is not limited to this platform. At MUSC analytic operations for OMOP transform and persist reference datasets using SQL Server 2019. The FHIR CDR is assessed remotely via virtual private network linkages to the HSSC’s data center at Clemson University. SAS, used to run the PopMedNet queries, is version 9.4.

The transformation process and maintenance pipeline was based upon the export of a series of flat pipe-delimited files that were loaded into the FHIR CDR but supports many options for importing EHR data, including HL7 FHIR transactions and HL7 v2.x transactions. In this instance, large export files were then processed to instantiate the repository with preexisting data. An incremental updating approach based on extracting new data into a flat file from the RDW was also developed. This incremental update is extracted daily and loaded into the FHIR repository. The FHIR repository and “live” CDMs are updated incrementally. In FHIR, any changes from the previously persisted data are recorded as new versions of the resources. OMOP and PCORnet “live” versions of the databases only store the most recent value. Restoration of data elements in the FHIR database automatically results in updating of the other 2 CDMs via the subscription mechanism. The staging and production OMOP and PCORnet instances for release are rebuilt *de novo* for quarterly releases. The environments for FHIR, OMOP, and PCORnet require about 2, 0.5 and 0.4 terabytes, respectively.

Composite time for database production, including the application of postprocessing steps in the pipeline, was observed. For the

OMOP instance, data quality measures were computed using both the in-house reports and OHDSI Achilles v1.6.7 and DQD v1.0 (develop) tools. For the PCORnet instance, standardized database quality assessment routines were computed and applied. Results of prior assessments of data quality for PCORnet certification were compared to this new approach for the generation of the database.

The iterative data quality assessment cycles resulted in numerous improvements that were mostly implemented as postprocessing steps with the expectation that they will be pushed upstream into the main pipeline if they are not constrained by the transactional nature of the data. For instance, the mappings to the expected terminologies were gradually improved in terms of completeness and correctness and then could be applied at any stage. In contrast, the data cleanup steps that remove data of insufficient quality were limited to the analytic production instances only since future data updates to the “live” instances might improve the data quality and thus save these bits of information.

Work on and participation in the HSSC data warehouse program is conducted under IRB PRO0009273.

RESULTS

Figure 2 shows the time delays with different approaches to capture of “raw” EHR data. Flat-file exports result in delays of days, while data accumulate for export at each stage. Export from our Epic EHR to Epic’s Clarity database occurs nightly. Data are extracted from this database daily and stored in a linked clinical model based on the EHR data model with minimal transformation and then exported as flat files for conversion to FHIR. There are other available approaches, such as HL7 v2.x or FHIR data streams, for other contributing data partners. Early data products support (blue shading in the figure) trial feasibility studies (counts); final products meet network quality requirements and support longitudinal analyses. Processing for conversion to FHIR, PCORnet, and OMOP occurs at HSSC’s Clemson database facility designed to support multiple institutions, each with their own segmented, but linked, FHIR infrastructure.

Table 1 shows the main administrative and clinical data domains implemented in the F-on-F architecture, applied to the MUSC data for the period from July 1, 2014 to December 31, 2020. The results reveal 2 primary findings. First, the data entry location indicates how the data are represented in each layer. In most cases, there is a 1:1 correspondence between the CDMs, but certain domains are not straightforward. For instance, the vital signs and the lab results are separate domains but exist as common observation resources in FHIR, stored in both the Measurement and Observation tables in OMOP, and in the individual tables only in PCORnet.

Second, the majority of the metrics are highly consistent between all stages of the pipeline. There are about 11M visits for slightly over 1M patients, associated with 18M diagnoses and 22M procedures. The differences between the CDMs are due to several factors. Most of them are common to all domains: there are certain source data entries that are not processed into the pipeline; on the other hand, the OMOP and PCORnet CDMs require certain data entries to be removed. The most dramatic differences are OMOP-specific due to the domain assignments (eg, a diagnosis code might be classified as an observation concept) and one-to-many mappings.

In addition to the main resources/tables shown in Table 1, the F-on-F architecture implementation involved other resources/tables, including linked resources based on the relational model (eg, FHIR DiagnosticReports), extensions to capture the data elements not covered by the standard specifications, and mapping-related supporting elements (eg, FHIR ConceptMaps). The specific examples of FHIR to OMOP mapping are illustrated in Supplementary Table S1.

With regard to specifics of data quality, the OHDSI DQD had 3312 data quality checks, including conformance and plausibility tests, and the OMOP instance passed 3092 (93%). Some of the failed checks were due to DQD technical errors (submitted on Github). Some data issues cannot be resolved without significant effort for a limited impact (eg, medication mapping is still somewhat incomplete, with an additional 20K+ codes needed for comprehensive coverage for the last 5 percent by volume of medication prescriptions and administrations). The PCORnet data quality was verified after each quarterly refresh by executing the Empirical Data

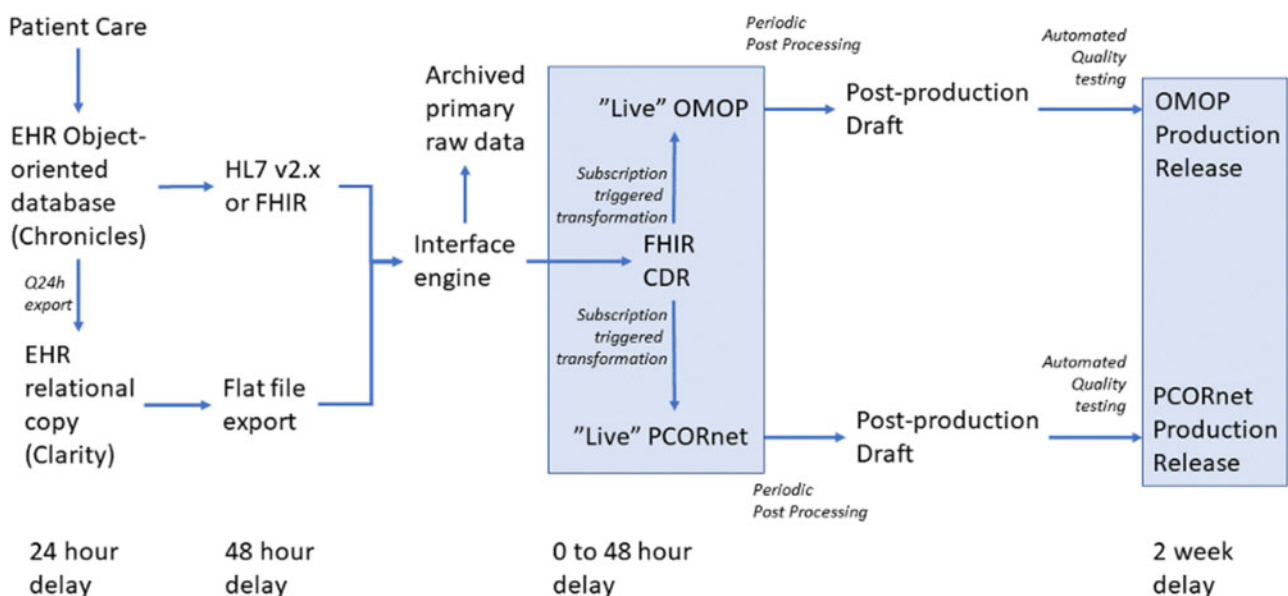


Figure 2. Computational pipeline for simultaneous multi-CDM production. Shaded boxes show technically available computational products. Approximate delays (relative to the previous step) are shown in the bottom.

Table 1. Comparison of source RDW, FHIR, OMOP, and PCORnet CDMs for MUSC

Domain	Source	FHIR		OMOP		PCORnet	
		Count	Resource	Count	Table	Count	Table
Patient	1 078 964	Patient	1 063 886	Person	1 059 009	Demographic	1 063 891
Visit	10 746 491	Encounter	10 636 834	Visit_Occurrence	10 628 243	Encounter	10 636 928
Diagnosis	18 402 862	Condition	17 593 910	Condition_Occurrence	14 254 546	Diagnosis	17 594 043
Procedure	23 029 835	Procedure	22 246 999	Procedure_Occurrence	16 838 402	Procedures	22 247 280
MedOrder	37 948 002	MedicationRequest	32 693 681	Drug_Exposure	31 829 609	Prescribing	32 703 095
MedAdmin	84 450 696	MedicationAdministra tion	43 577 863	Drug_Exposure	39 562 252	Med_Admin	39 574 769
Vital	16 917 012	Observation	16 127 660	Measurement + Observation	16 128 048	Vital	16 127 920
Lab	137 984 163	Observation	124 870 259	Measurement + Observation	154 831 038	Lab_result_CM	119 992 129

Check (EDC). SAS package was provided by the PCORnet Distributed Research Network Operations Center. There were 1450 data checks validating that all data elements were populated as expected, column lengths were correct, mappings conformed to the specification, relationships were logical, referential integrity was respected and more. Our PCORnet instance passed 1424 (98%) of those checks. An extended metrics report (generated by the custom SQL) is provided in the [Supplementary Table S2](#).

Initial loading of the backlog of 5 years of EHR data into the FHIR instance required several weeks. However, once loaded, the approach resulted in significant improvements in the time required to produce quarterly updates of the PCORnet instance and improved the concurrency of data in refreshes. As shown in [Figure 3](#), preparation time was far less. Significant work was still required in postprocessing but now could be focused on data quality.

DISCUSSION

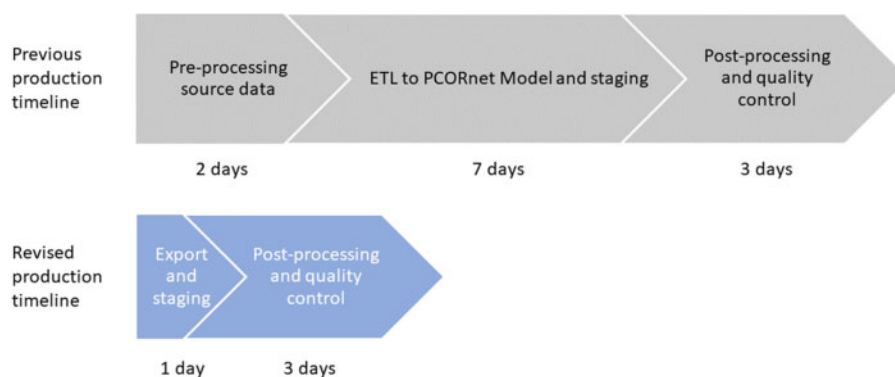
The sustainability of research networks for COVID-19 and emerging disorders is an important issue. Costs arise in part due to custom data modeling, ETL tasks, and repetitive data integration tasks required for operations. Further, research infrastructure has to be replicated at each site, requiring significant additional investments. A unique feature of the design is that rather than rely upon the FHIR representation as to the means for query and retrieval of data, the architecture uses FHIR subscriptions to trigger continuous transformations to other common data models that are maintained in synchrony. This approach also allows the selection of data for specific subsets of patients. Linked databases are available for a query with minimal time lag behind the data source for queries for counts and

other simple operations. When an analytical-grade quality of the instance is required, the postprocessing can be applied on-demand to produce analytical datasets.

The primary innovations in this work are the use of FHIR as an initial canonical data model and FHIR subscription protocols for the transformation and synchronization of multiple data models. In future work, we will explore the use of subscription models to distribute data across networks and to maintain shared data elements, such as mortality status and social determinants of health data. We will also explore the use of this approach to federate clinical data across sites by maintaining a single master patient identifier and consistent supporting demographic information.

Prior approaches to the problem of maintenance of multiple linked data models in a data repository have focused on other canonical models and *automation* of data both transformation for queries and production of datasets. For example, work from the i2b2 group at Harvard has used the i2b2 data format as the canonical representation to support dynamic ETL from that format to the OMOP and to FHIR.¹⁹ Ong et al use the OMOP model as their canonical representation of data and support dynamic queries mapped from the PCORnet CDM.²⁰ Choi et al²¹ have developed automated mapping functions from OMOP to FHIR for computation.

There is also prior work with the use of FHIR as a meta map for ETL operations between clinical data models. Pfaff et al²² describe the use of CampFHIR, a tool for guiding ETL for conversions of different models. FHIR concept representation aids and speeds a largely manual ETL process. This approach guides current N3C efforts.² The F-on-F approach described herein was developed contemporaneously with CampFHIR¹⁷ with both efforts benefiting from collaborations. F-on-F differs in that it is focused on ongoing

**Figure 3.** Production timelines for PCORnet database release.

data production through automated maintenance of parallel CDM instances. Data transformations are implemented in Java code and executed in near real time. The FHIR CDR servers also provide reference storage and master data management of EHR data and a live canonical representation of the data. The approach may allow us to better understand what is lost in translation.

F-on-F combines near real-time data transformations with post-processing for production release databases. We envision the scope of real-time processing as being pragmatic, scaled to the intended use cases. For example, if a site wanted to run OMOP models for prediction of sepsis from clinical data, then real-time processing might need to be expanded to meet those needs.

The use of FHIR as a data source from EHRs is important from a policy perspective as Information Blocking Statutes stemming from the 21st Century Cures Act specify a range of clinical data to be available from EHRs for downloading and data exchange (the USCDI).¹¹ They further require that EHRs respond to queries for USCDI elements in the FHIR standard both at an individual patient query level and population level starting in December of 2022. Providers who cannot offer data access in this format may be subject to fines for “information blocking.” EHR vendors must offer these capabilities to maintain certification of their systems as compliant with Meaningful Use regulations. As a result, many barriers for data production for research and safety will be overcome if the starting point for data transformation operations for computational models is the FHIR standard.

The broad future availability of data in the FHIR standard raises the question of whether other analytically oriented models are necessary. OMOP and PCORnet are highly evolved models refined for their purpose.²³ As analytical models, they are optimized for efficient and unbiased analysis of large volumes of longitudinal normalized data. The FHIR model is an object-oriented data model, focused on the accurate expression of clinical events, not computation. F-on-F envisions a best of both worlds approach, with flexible representation and optimized computation.

Limitations

The above approach is standards-based but leverages proprietary extensions of the FHIR subscription specification. As discussed above, there are inherent limitations in the speed with which clinical data can be integrated into any analytic model such as OMOP or PCORnet. For example, orders or laboratory data cannot be linked to an encounter that does not (yet) exist. The use of the persistence module for the transformation of data is novel and computationally efficient but implements rules in compiled Java code, where changes may be more difficult. Ultimately, some manual ETL was still deemed optimal in the production pipeline; however, future work may reduce these requirements.

CONCLUSION

The use of FHIR standard as a canonical representation of clinical data with the subsequent dynamic transformation to other research CDMs for analytics is a practical approach to accelerate the availability of data for research and may be particularly useful for evolving diseases such as COVID-19. While it is theoretically possible to fully automate transformation to near real-time versions of OMOP or PCORnet databases, it is more practical given the evolving nature of data to take a staged approach for models for longitudinal data analysis applications.

FUNDING

This project was supported by grants to Health Sciences South Carolina from the Duke Endowment and by the National Center for Advancing Translational Sciences of the National Institutes of Health under Grant Number UL1 TR001450.

AUTHOR CONTRIBUTIONS

LAL designed the study, helped obtain funding, and wrote major parts of the manuscript. AVI, PR, and JJ contributed to the conduct of the study and the writing of the manuscript. KD helped obtain funding for the study and contributed to the writing of the manuscript. JA and DW contributed to the design, the conduct of the study, and to the writing of the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *Journal of the American Medical Informatics Association* online.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

CONFLICT OF INTEREST STATEMENT

Dr. Lenert, Dr. Ilatovskiy, Ms. Rudisill, Mr. Jacobs, and Mr. Deans have no competing interests. Mr. Agnew and Mr. Weatherston have financial interests in Smile CDR and are employed by the company.

REFERENCES

1. Piller C. Disgraced COVID-19 studies are still routinely cited. *Science* 2021; 371: 331–2.
2. Haendel MA, Chute CG, Bennett TD, *et al.*; the N3C Consortium. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc* 2021; 28 (3): 427–43. doi:10.1093/jamia/ocaa196
3. Visweswaran S, Becich MJ, D'Itri VS, *et al.* Accrual to Clinical Trials (ACT): a clinical and translational science award consortium network. *JAMIA Open* 2018; 1: 147–52.
4. Hripcsak G, Duke JD, Shah NH, *et al.* Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform* 2015; 216: 574–8.
5. Block JP, Marsolo KA, Nagavedu K, *et al.* Characteristics of 24,516 patients diagnosed with COVID-19 illness in a national clinical research network: results from PCORnet. *bioRxiv* 2020; doi:10.1101/2020.08.01.20163733.
6. PCORnet: Progress, challenges, and opportunities ahead. 2015. <https://www.pcori.org/blog/pcornet-progress-challenges-and-opportunities-ahead> Accessed November 24, 2019
7. TriNetX readies its real-world data platform and global network of healthcare organizations to support COVID-19 Clinical Research. 2020. <https://trinetx.com/covid-19/> Accessed January 31, 2021
8. Taquet M, Luciano S, Geddes JR, *et al.* Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry* 2021; 8 (2): 130–40.
9. FDA Sentinel System's Coronavirus (COVID-19) activities. 2020. <https://www.sentinelinitiative.org/assessments/coronavirus-covid-19> Accessed January 31, 2021
10. Brat GA, Weber GM, Gehlenborg N, *et al.* International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *NPJ Digit Med* 2020; 3: 109.

11. Mandl KD, Gottlieb D, Mandel JC, *et al.* Push button population health: The SMART/HL7 FHIR bulk data access application programming interface. *NPJ Digit Med* 2020; 3 (1): 151.
12. U.S. Core Data for Interoperability (USCDI) | Interoperability Standards Advisory (ISA). <https://www.healthit.gov/isa/us-core-data-interoperability-uscdi> Accessed November 24, 2019
13. Braunstein ML. SMART on FHIR. In: *Health Informatics on FHIR: How HL7's New API is Transforming Healthcare*. Cham, Switzerland: Springer; 2018: 205–25. doi:10.1007/978-3-319-93414-3_10.
14. Braunstein ML. FHIR. In: *Health Informatics on FHIR: How HL7's New API is Transforming Healthcare*. Cham, Switzerland: Springer; 2018: 179–203.
15. Azure API for FHIR. <https://azure.microsoft.com/en-us/services/azure-api-for-fhir/> Accessed February 14, 2021
16. Subscription - FHIR v4.0.1. <https://www.hl7.org/fhir/subscription.html> Accessed November 24, 2019.
17. Health Sciences South Carolina lights up a FHIR-based clinical data repository. 2018. <https://www.healthcareitnews.com/news/health-sciences-south-carolina-lights-fhir-based-clinical-data-repository> Accessed January 30, 2021
18. Mandel JC, Kreda DA, Mandl KD, *et al.* SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc* 2016; 23: 899–908.
19. Klann JG, Phillips LC, Herrick C, *et al.* Web services for data warehouses: OMOP and PCORnet on i2b2. *J Am Med Inform Assoc* 2018; 25: 1331–8.
20. Ong TC, Kahn MG, Kwan BM, *et al.* Dynamic-ETL: a hybrid approach for health data extraction, transformation and loading. *BMC Med Inform Decis Mak* 2017; 17: 134.
21. Choi M, Starr R, Duke J. OMOP on FHIR as an enabler for analytics-as-a-service. Washington, DC: AMIA; 2017.
22. Pfaff ER, Champion J, Bradford RL, *et al.* Fast Healthcare Interoperability Resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. *JMIR Med Inform* 2019; 7: e15199.
23. Weeks J, Pardee R. Learning to share health care data: a brief timeline of influential common data models and distributed health data networks in US Health Care Research. *EGEMS (Wash DC)* 2019; 7 (1): 4.