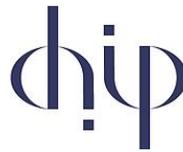


Push Button Population Health Data: Extending the HL7® FHIR® Standard to Support Bulk Data Export

April 2018

Daniel Gottlieb¹, MPA, Alyssa Ellis¹, Wayne Kubick², MBA, Grahame Grieve³, Josh Mandel⁴, MD, Kenneth D. Mandl^{1,4}, MD, MPH



Funded by ONC Cooperative Agreement 90AX0019/01-01

¹ *Computational Health Informatics Program, Boston Children's Hospital, Boston, MA*

² *Health Level Seven International*

³ *Health Intersections*

⁴ *Department of Biomedical Informatics, Harvard Medical School, Boston, MA*

Contents

About Health Level Seven International	3
About SMART	3
Overview	3
Bulk Data in Healthcare	4
Current Challenges	6
Opportunity with FHIR	7
FHIR and SMART	8
Extending FHIR to Support Bulk Data Access	9
Privacy and Security Considerations	10
Project Scope and Timeline	11
Conclusion	12
Appendix 1: Supplemental Technical Materials	13

About Health Level Seven International

Founded in 1987, Health Level Seven International (HL7) is a not-for-profit, ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. "Level Seven" refers to the seventh level of the International Organization for Standardization (ISO) seven-layer communications model for Open Systems Interconnection (OSI) - the application level. The application level interfaces directly to and performs common application services for the application processes. Although other protocols have largely superseded it, the OSI model remains valuable as a place to begin the study of network architecture. Fast Health Interoperability Resources (FHIR) is an HL7 specification for Healthcare Interoperability.¹

About SMART

SMART Health IT is an open, standards based technology platform, that builds on the FHIR API to enable innovators to create apps that seamlessly and securely run across the healthcare system. Using an electronic health record (EHR) system or data warehouse that supports the SMART standard, patients, doctors, and healthcare practitioners can draw on this library of apps to improve clinical care, research, and public health.

The SMART platform is composed of open standards, open source tools for developers building apps and a publicly accessible app gallery. The SMART on FHIR specification extends the FHIR standard to better support application use cases by providing user experience integration, authorization support, and single sign on capabilities based on the OAuth standard. To date, dozens of clinical applications have been built on this platform and SMART applications are being used to provide clinical care at leading healthcare institutions. The SMART Health IT Project is run out of the not for-profit institutions, Boston Children's Hospital Computational Health Informatics Program and the Harvard Medical School Department for Biomedical Informatics.²

Overview

¹ <http://www.hl7.org/>

² <https://smarthealthit.org/>

The Open APIs called for in the 21st Century Cures Act present an opportunity to create the learning healthcare system that has been long envisioned. A learning healthcare system must be able to do more than conduct individual queries on one patient; it requires the ability to aggregate and analyze data at a population level, to better manage population health, deliver value-based care, and conduct discovery science. Access to data is a fundamental requirement for improving the health of individuals and populations. Lowering the financial and technical barriers to allow health data to become more accessible and usable will create a more robust marketplace for innovative health care technologies. APIs for working with population level data, often referred to as bulk data, are needed to support this system. The existing FHIR and SMART APIs work well for accessing small amounts of data, but large exports perform poorly, requiring an impractical number of API requests to be issued serially. By adding asynchronous primitives to FHIR and defining an export operation, the Bulk Data API enables secure integration of third-party, externally-hosted applications into diverse EHR and data warehouse environments.

On behalf of the Office of the National Coordinator for Health Information Technology (ONC) , The Boston Children's Hospital Computational Health Informatics Program and SMART hosted a meeting in December 2017 to discuss standardizing bulk data exports from EHR systems and data warehouse environments. This meeting brought together key stakeholders from across health care, including the National Coordinator for Health Information Technology, top representatives from HL7, payer organizations, health systems, EHR vendors, as well as other leading health technology innovators.

This document explores use cases for EHR bulk data export, and provides details about the proposed technical roadmap and regulatory environment for population level data export from health information systems using FHIR.

Bulk Data in Healthcare

Recent Centers for Medicare and Medicaid Services (CMS) payment programs and models require new capacities to administer population and value-based payments. CMS now emphasizes expanded use of health IT and a more robust data infrastructure to ensure transparency of accurate quality and cost information to monitor value-based care programs. CMS uses data to monitor trends in critical health measures among priority populations; monitor health status, health care, and health policies at the national, state, local, and tribal levels; and conduct in-depth studies of population health at the community level, and for specific groups of individuals.³ Other payer organizations also use bulk data to measure the clinical performance of their plan and networks, especially related to value-based care. CMS and private payers are interested in analyzing the data to better understand transitions of care, gaps in care, and risks associated with certain patient groups. The ability to access and analyze large volumes of data

³ <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf>

can help payers to initiate timely interventions that reduce visits to the hospital and emergency department.

In 2016, employer-based insurance covered 55.7 percent of the population.⁴ Employers have a strong interest in the cost of health care, and use data for various components of population health management to understand and improve their health plan performance. Many employers are now encouraging Accountable Care Organizations (ACO) and incorporating value-based benefits options for employees. ACOs are groups of hospitals and healthcare providers who come together voluntarily to provide coordinated high quality care, ensuring that all patients, especially the chronically ill, get the right care at the right time, while avoiding unnecessary duplication of services and preventing medical errors.⁵ For this type of coordinated care to be successful, member organizations need to implement HIT solutions that can insert quality measures into the clinical workflow, and deliver actionable data to all members of the care team in real time. This requires the consolidation of clinical and claims data from multiple data sources.

More automated processes for gathering data required by CMS and private payers, and for computing quality measures will relieve much of the burden placed on the health system in order to meet these requirements. Data must be integrated from multiple sources, often requiring multiple extract, transform and load (ETL) operations, in order to accurately assess both a patient's health and the quality and value of care received. Payers provide claims data to help supplement the data available in a provider's EHR system, because it is rarely the case that a patient only receives care at one institution, with all of their health information available within one electronic record. A patient with a chronic condition, such as diabetes, will likely receive care from a number of providers with different specialties, probably not all at the same healthcare organization, so sharing data between EHR systems is critical in order for each provider to have access to the 360 degree view of the patient's care and conditions.

Outside of caring for specific patients, health care organizations also export bulk data for transfer into disease registries, clinical research or pharmacovigilance databases, and into data warehouses for operational or clinical analytics. Interfacing registries with EHRs has been an ongoing challenge, because EHRs are used to collect data on individual patients, and registries are often built to understand population-level data. The process to extract data from EHRs into registries is complex and time-consuming and often introduces data errors and redundancies that have to be addressed. A common way that disease registries are populated with data from the EHR involves interfacing with a Health Information Exchange (HIE) network. This approach involves exporting bulk data from the EHR into the HIE, which also collects data from other sources, such as claims data from payers, pharmacy data, laboratory data, and data from other EHRs. It also requires that the data be matched to the patient and physician, and pushed at regular intervals to the registry. These networks supply the enormous amount of aggregated

⁴ <https://www.census.gov/library/publications/2017/demo/p60-260.html>

⁵ <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/>

data necessary to apply machine learning for population health analytics, however, the processes for importing and exporting data are still highly complex.

Current Challenges

Today, bulk data export is often accomplished through proprietary pipelines, which are labor intensive to create and maintain. Data extraction and import is different for every system, and custom data fields and proprietary data models require manual mapping to and from delimited text formats like comma separated values (CSV). IT departments have to allocate teams of engineers for each project, so it is very expensive for the health system to implement these pipelines, and each transfer requires manual ETL processes. Health organizations will enthusiastically support any standardization that will help to lower costs and make these processes more efficient, and major EHR vendors have acknowledged that supporting a standardized bulk data API that could improve the efficiency and consistency of the data extract process.

The original use cases for health information exchange were simply to move data between electronic health records, but today's Health Information Technology (HIT) landscape requires much more flexibility. The 21st Century Cures Act requirement that EHR vendors make patient data accessible "without special effort, through the use of application programming interfaces"⁶ means that health care providers should be able to easily exchange consistent patient information across different vendor systems and provide data access for patients. Most providers already have interchange formats in place, but these are custom and proprietary for each data source. This type of architecture is expensive and requires constant maintenance, making it unsustainable going forward. We propose instead that providers should have access to multiple data sources through a single HL7 FHIR interface.

Insurance providers routinely share data back and forth with provider organizations, including claims information and information from care management systems. However, one payer has claimed that more than 50% of provider organizations have never even downloaded the information that is sent. These organizations are not set up with the technical architecture to import and analyze the data in a meaningful way. Because population-level health data is so valuable, some organizations have internally invested the time and resources to build HIEs to facilitate the transfer of their data, and HIT systems to record, aggregate, and analyze it. These systems can work very well for the cluster of organizations that they serve, but they aren't all interoperable; the data cannot be shared and presented across disparate healthcare settings regardless of the application or vendor.

The Massachusetts eHealth Collaborative (MAeHC) receives about 30 million health records per year or about 100,000 per day. The data is sent electronically or entered manually, in

⁶ <https://www.congress.gov/bill/114th-congress/house-bill/34/>

different formats from different sources. About 60% comes from various EHRs and 40% via intermediaries. They have to maintain many custom interfaces to support these data flows. Micky Tripathi, President & CEO of MAeHC, says an API-based approach that doesn't require engineering support to build an interface for each source system would make bulk data more efficient. This could be done using RESTful web resources and the FHIR standards.⁷

Opportunity with FHIR

Federated data networks are allowing researchers and clinicians to access and analyze large standardized datasets from many sources, but these networks do not have a sustainable architecture to handle population-level health data going forward. It involves expensive, manual and customized work to get the data out of the EHR and into these networks. Using FHIR for such processes would make existing networks more sustainable, and allow much more data to become available for clinical research and population health analytics.

There are a number of innovative organizations that have built their own solutions based on the FHIR standard, and shown that it works. Cambia Health Solutions built an encounter query API using FHIR, to deliver Admit-Discharge-Transfer notifications (ADTs) to an external business partner for care management. Bulk data on members was pulled in on ADTs from clearinghouses and other facilities. It was pulled in via CSVs and then mapped to FHIR resources for internal RESTful API access.⁶ This has proven to be an effective strategy, but it is not an automatic process. Going forward, a user should be able to export any type or amount of data from the EHR in FHIR format, so that it is ready to access, analyze, and share through standardized pipelines.

Before Epic started rolling out FHIR support, Rush Health created their own FHIR server to extend their Epic EHR to connect to external RUSH apps and 3rd party health apps. Rush is using SMART on FHIR and CDS Hooks⁸ to help patients needing medical and social assistance. FHIR resources containing patient information from the EHR and other information systems are pushed from Rush to a social welfare system, which finds donors to assist patients. Patients receive real-time notification about assistance.⁶

FHIR could provide a sustainable, standard data transfer process, enabling the type of data liquidity that is common in other industries. A FHIR-based data extraction approach creates standard components and modules that can be put together into pipelines to support different scenarios, such as: creating a de-identified view for researchers, sharing data between provider organizations, and sharing data between payers and providers. This means data entered once can be used by downstream applications and systems without any special effort to retrieve the

⁷ <http://smarthealthit.org/wp-content/uploads/Population-Level-Data-Export-Meeting-Summary-Report.pdf>

⁸ <https://cds-hooks.org/>

data. These capabilities would have a huge impact for clinical trials, personalized medicine, disease registries, and population health analytics.

Recognizing the value this approach could bring into the healthcare ecosystem, The ONC have asked the FHIR community to add new capabilities to the FHIR standard to increase support for API-based access to bulk data.

FHIR and SMART

The HL7 FHIR standard defines a set of data models referred to as FHIR Resources that, either by themselves or when combined, satisfy the majority of common health use cases. Each resource has a tag that acts as a unique identifier, just like the URL of a web page, meaning health data can be shared and accessed using a standard browser and url, in a format that is already widely understood and used in other industries, regardless of the underlying infrastructure or operating system. FHIR overcomes the challenges that have been encountered with other HL7 standards, such as the C-CDA, because the data is modularized, and does not have to be extracted from a static document. More than 100 FHIR resources (data models) have been developed, which can reference other resources by their URL. For example, a FHIR resource for a vital sign could reference a resource about the patient, the provider who took the vital sign, and the encounter where it was taken. Current FHIR resources cover about 80% of the use case, but are extendable.

The SMART Health IT project builds on top of FHIR to enable substitutable health applications that can run on any vendor system that supports SMART. As of March 2018, there are 54 commercial and open source apps listed in the SMART App Gallery⁹, many of these apps are also listed in app galleries hosted by EHR vendors who support SMART on FHIR standards.

The Argonaut Project¹⁰ has been a driving force behind FHIR and SMART, working to develop a set of FHIR profiles (constraints on the standard) specific to use in the United States, and standards implementation guides to ensure that technologies that support the FHIR standard will actually be interoperable in production. The Argonaut Project was formed by a group of HIT vendors and healthcare organizations supporting HL7 development of FHIR standards to create usable, secure avenues to access data from various sources through APIs. The Argonaut participants, the HL7 FHIR team, and the SMART Health IT team have worked collaboratively to develop a set of standards and profiles that enable plug-and-play interoperability for health applications.

⁹ <http://apps.smarthealthit.org/>

¹⁰ http://argonautwiki.hl7.org/index.php?title=Main_Page

SMART on FHIR applications are changing the way that health care data is accessed and used. By reducing the need for custom work and expensive implementation projects, innovation is happening faster than ever. Expanding these standards to support bulk data export will allow population-level data to flow fluidly between HIT systems from different vendors, so that it can be used for its full potential, without special effort.

Extending FHIR to Support Bulk Data Access

The SMART Health IT project in the Computational Health Informatics Program at Boston Children's Hospital is working with HL7, the ONC, EHR vendors, payers, and the FHIR community to extend FHIR and the SMART authorization model to support bulk data access. The goal is to enable automated communication between backend services and clinical systems. These systems should be able to communicate with each other without a user having to log in once a connection has been configured. One key design goal is to limit the complexity of the API to prevent implementation of the specification from being burdensome. The project is also focused on using mature, stable technologies, reusing existing FHIR data models, API formats, and data types wherever possible, and leveraging the existing standards for authentication and authorization.

File Request

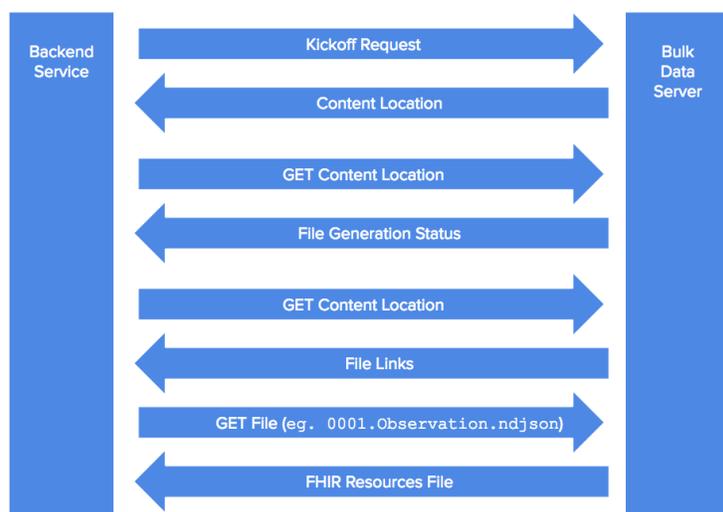


Figure 1 shows the envisioned request flow between a backend service and a bulk data server.

The initial design proposes that backend service make a http request to the FHIR bulk data server to kick off the data export. The requested data can be filtered based on attributes like the creation date, FHIR data type and patient population group. If the request is successful, the FHIR server will return a URL that the backend service can then poll periodically to check the status of the export. When the export has been generated, the FHIR server returns URLs for the

data files to the backend service which can then retrieve these files. A detailed technical description of this flow and the file format is available in Appendix 1.

The entire transaction can be secured using the SMART Backend Services OAuth profile (see Appendix 1). In this model, the backend service submits a signed request using a pre-configured certificate pair to the FHIR server and receives a token that it can use in subsequent requests.

This project involves extending the FHIR specification to make the bulk data approach work. There will need to be a FHIR operation to get all data on all patients. Today in FHIR it is possible to request all data on an individual patient, but not all data on every patient. There will also need to be a FHIR operation to obtain data on groups of patients. FHIR already has the concept of defining groups of patients and the intended functionality will support restricting the data export to a single one of these groups, representing a concept such as the patients associated with a particular payer. There will need to be support for files in a format that supports reading and writing the data without loading the entire file into memory. Lastly, FHIR will need to be able to perform large requests without keeping the connection between the client application and the FHIR server open while the query is processed.

These extensions to the FHIR standard correspond to the use cases outlined earlier with the export of all data on all patients facilitating scenarios such as moving data from a transactional system to a research database, and the capability to export data on groups of patients supporting uses such as data exchange on subsets of patients between payers and providers, as well as transfer of data to disease registries or even internal applications focused on a specific population.

Privacy and Security Considerations

Using the FHIR bulk data standards to share population-level health data assumes that an appropriate legal framework is in place between parties. A recommended approach for secure, compliant access to consented EHR data might involve a business associate agreement (BAA), service-level agreement (SLA), and/or data use agreement (DUA).

Most data in the electronic health record is collected for clinical care and not as part of a research study, so consent to share the data is not typically obtained at the time it is collected. In order to export and share data from the EHR, most Institutional Review Boards (IRBs) would require that the data is de-identified before it is exported to registries or other organizations, so that no individual patient can be easily identified from the data that is shared. Covered Entities, which include providers, health care clearinghouses, and health plans must comply with HIPAA to uphold the privacy and security of PHI, and under the HITECH Act, business associates must also comply with HIPAA regulations.

To grant access to population-level data, covered entities should ensure that they have a current HIPAA BAA in place with each of their partners. In addition to the BAA, there should be

agreements in place that address the method of data transfer and storage, terms of use, and oversight of data access. It is recommended that a DUA or similar type of data sharing agreement is executed for each organization requesting access to data, which clearly states why the receiving party is requesting access to the data, and how the data will be used. The requestor should attest to their ability to keep the data contained on encrypted servers and devices, and agree to establish appropriate administrative, technical, and physical safeguards to protect the confidentiality of the data and to prevent unauthorized use or access to it. The data provider should implement a mechanism to allow different levels of access for different users. Some partners may only need access to aggregate data, whereas other users may have the approvals in place to access individual patient data. A data management process to allow for periodic auditing of access is also recommended. This process should include the ability to impose sanctions when necessary.

Project Scope and Timeline

There are several important areas related to bulk data that are out of scope for the technical bulk data implementation guide, but represent potential future projects. The technical standard will not attempt to define specific legal and regulatory requirements for EHR bulk data sharing, however, in the future, organizations may wish to collaborate on creating standard legal agreements to use with partners. The bulk data standard will initially focus on moving bulk data in batches, rather than in real time. However, even if the batch data is loaded nightly, synchronous FHIR API calls can be used to pull in any data that has changed since the last batch upload in order to provide up to date data on individual patients. Data transformation, such as de-identification and encryption, will not be built into the bulk data export specification, but can be addressed via modules that connect together by importing and exporting bulk data. HIEs and other systems that import data from multiple sources use specific data elements for patient matching, such as a subscriber number. These data elements will come through when data is pulled via the FHIR bulk data API, however this specification will not include any special measures for handling data elements for patient matching.

Timeline

December 2017	<ul style="list-style-type: none"> ● ONC meeting on bulk data
January 2018	<ul style="list-style-type: none"> ● SMART Bulk Data Reference Implementation Server ● Initial Connectathon for interested parties to start experimenting with proposed technical specification
2018	<ul style="list-style-type: none"> ● FHIR Connectathon events to refine and develop the technical specification ● Argonaut Project to bring together pilot implementations

	<ul style="list-style-type: none"> • Publish Draft Implementation Guide
2019	<ul style="list-style-type: none"> • Publish version 1.0 of FHIR Implementation Guide For Bulk Data

Table 1 shows the planned timeline for key events in the development of the FHIR bulk data Implementation Guide.

A series of FHIR Connectathon events have been planned to promote community involvement and further refine and develop the FHIR bulk data technical specification. The January 2018 Connectathon was held in New Orleans, and was well-attended by interested parties from across the health care community. Changes discussed at the connectathon have already been incorporated into the draft Bulk Data Implementation Guide and SMART Reference Server. EHR vendors and other members of the FHIR community are working on server and backend service pilot implementations.

Conclusion

With a shift toward value based care, the healthcare system needs better data analytics and population health management tools. Population level data must move fluidly between different systems, so that payers, employers and providers can work together to identify and manage high-risk patients, assess and improve care quality and construct efficient, high quality care plans. Major health systems are already using FHIR APIs to export population-level data and aggregate data from multiple sources, in order to support innovative new tools and solutions. A FHIR bulk data API will allow for liquidity of population-level data, which means lower costs, increased competition in the health IT market, and a spur in innovation. Whether the data is being used for reporting quality measures, machine learning, disease registries, or data mining to identify high-risk patients, the ability to combine data from multiple sources in a standardized way is a big step toward the learning health care system that has been envisioned.

The SMART and FHIR standards have already been adopted by major EHR vendors and health systems throughout the United States, and SMART apps are being used today to help clinicians better utilize the data and information available to make decisions at the point of care. Using web and mobile applications written against the SMART on FHIR standards, which have adopted the profiles outlined in the Argonaut Implementation Guide¹¹, patients and providers are accessing and using health data to improve the health and lives of individuals. SMART will continue to work with the ONC, HL7 and the FHIR community to extend the FHIR standard to support bulk data export, leveraging existing FHIR and SMART work to improve the liquidity of population level data.

¹¹ <http://www.fhir.org/guides/argonaut/r2/>

Appendix 1: Supplemental Technical Materials

- Draft Bulk Data Implementation Guide:
<https://github.com/smart-on-fhir/fhir-bulk-data-docs>
- Authentication approach with SMART backend services specification:
<http://docs.smarthealthit.org/authorization/backend-services/>
- SMART Server Reference Implementation:
<https://bulk-data.smarthealthit.org>
- SMART Client Reference Implementation:
<https://github.com/smart-on-fhir/sample-apps-stu3/tree/master/fhir-downloader>
- Other implementations:
[https://github.com/smart-on-fhir/fhir-bulk-data-docs/blob/master/README.md#server-
implementations](https://github.com/smart-on-fhir/fhir-bulk-data-docs/blob/master/README.md#server-implementations)
- Discussion Group:
<https://chat.fhir.org/#narrow/stream/bulk.20data>