Development of a Generalizable Multi-site and Multi-Modality Clinical Data Cloud Infrastructure for Pediatric Patient Care

Andrew Hornback
ahornback6@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Wenqi Shi
wshi83@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Felipe O. Giuste
fgiuste@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Yuanda Zhu
yzhu94@gatech.edu
Georgia Institute of Technology
Atlanta, GA

Ashley M. Carpenter
acarpenter@shrinenet.org
Shriners Children's Headquarters
Tampa, FL

Coleman Hilton
chilton@shrinenet.org
Shriners Children's Headquarters
Tampa, FL

Vinieth N. Bijanki
vbijanki@shrinenet.org
Shriners Children's St. Louis
St. Louis, MO

Hiram Stahl
hstahl@shrinenet.org
Shriners Children's St. Louis
St. Louis, MO

Gary S. Gottesman
gsgottesman@wustl.edu
Shriners Children's St. Louis
St. Louis, MO

Chad Purnell
cpurnell@uic.edu
Shriners Children's Chicago
Chicago, IL

Henry J. Iwinski
hiwinski@shrinenet.org
Shriners Children's Lexington
Lexington, KY

J. Michael Wattenbarger
mwattenbarger@shrinenet.org
Shriners Children's Greenville
Greenville, SC

May D. Wang
maywang@gatech.edu
Georgia Institute of Technology
Atlanta, GA

ABSTRACT

World-renowned pediatric patient care in scoliosis, craniofacial, orthopedic, and other life-altering conditions is provided at the international Shriners Children's hospital system. The impact of scoliosis can be extreme with significant curvature of the spine that often progresses during childhood periods of growth and development. Gauging the impact of treatment is vital throughout the diagnostic and treatment process and is achieved using radiographic imaging and patient reported feedback surveys. Surgeons from multiple clinical centers have amassed a wealth of patient data from more than 1,000 scoliosis patients. However, these data are difficult to access due to data heterogeneity and poor interoperability between complex hospital systems. These barriers significantly decrease the value of these data to improve patient care. To solve these challenges, we create a generalizable multi-site and multi-modality cloud infrastructure for managing the clinical data of multiple diseases. First, we establish a standardized and secure research data repository using the Fast Health Interoperability Resources (FHIR) standard to harmonize multi-modal clinical data from different hospital sites. Additionally, we develop a SMART-on-FHIR application with a user-friendly graphical user interface (GUI) to enable non-technical users to access the harmonized clinical data. We demonstrate the generalizability of our solution by expanding it to also facilitate craniofacial microsomia and pediatric bone disease imaging research. Ultimately, we present a generalized framework for multi-site, multi-modal data harmonization, which can efficiently organize and store data for clinical research to improve pediatric patient care.

CCS CONCEPTS

• **Information systems → Extraction, transformation and loading**; **Users and interactive retrieval**; **Data structures**.

KEYWORDS

ETL, Data Repository, Data Interoperability, User Interface, FHIR, SMART-on-FHIR, Pediatric Patient Care

ACM Reference Format:
Andrew Hornback, Wenqi Shi, Felipe O. Giuste, Yuanda Zhu, Ashley M. Carpenter, Coleman Hilton, Vinieth N. Bijanki, Hiram Stahl, Gary S. Gottesman, Chad Purnell, Henry J. Iwinski, J. Michael Wattenbarger, and May D. Wang. 2022. Development of a Generalizable Multi-site and Multi-Modality Clinical Data Cloud Infrastructure for Pediatric Patient Care. In *13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '22), August 7–10, 2022, Northbrook, IL, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3535508.3545565

BCB '22, August 7–10, 2022, Northbrook, IL, USA

ACM ISBN 978-1-4503-9386-7/22/08.
https://doi.org/10.1145/3535508.3545565

1 INTRODUCTION

Electronic health records (EHRs) contain digital patient medical records, including demographics, medical history, vital signs, lab tests, diagnoses, treatment procedures, medications, radiology images, pathology images, and free-text clinical notes [8]. EHRs were originally designed to automate business processes and convert paper-based documentation into equivalent electronic forms with minimal thought for clinical decision support [17]. EHR sharing between institutions has solved data access issues associated with paper-based records, prevents costly re-admissions, reduces medication and prescription errors, improves disease diagnoses, and decreases duplicate lab testing [20]. To advance pediatric patient care and clinical research, clinicians and researchers seek to improve healthcare delivery by enhancing medical decisions by combining targeted clinical knowledge with health information enabled by comprehensive and effective EHR systems.

While EHRs provide efficiencies, and their adoption is widespread, there is no prescribed standard protocol. This has resulted in a variety of data storage and organization formats, which prevents many healthcare providers from realizing the full potential of EHRs. In comparison to other industries, the healthcare industry has been slow to adopt technological advancements and has lacked widely accepted standards to facilitate multi-site and multi-modal data studies [5]. Furthermore, due to a lack of data interoperability, the majority of existing clinical data is difficult to exchange, analyze, and interpret because it is hidden in isolated databases, incompatible systems, and proprietary software [12]. Consequently, the lack of health information exchange and data interoperability impedes digital innovations envisioned for future medicine, as technologies that rely on big data, such as artificial intelligence, clinical decision support, and mobile applications, cannot be utilized to their full potential.

Shriners Children's (SC) is a nationwide healthcare organization that has collected large amounts of multi-site and multi-modal electronic data from pediatric patients. For example, through the Setting Scoliosis Straight (SSS) Surgeon Performance Program (SPP) quality improvement registry, SC-Greenville and SC-Lexington have accumulated comprehensive examination information for all patients undergoing spine fusion. Clinicians at both sites have archived the medical records of over 1,000 individual patients and tracked each patient's two to seven separate visits in Microsoft Excel files, including pre-operative, post-operative, 6-month, 1-year, 2-year, 5-year and 10-year marks. Currently, scoliosis patient information is manually organized and disparately distributed in individual SC sites with heterogeneous formats only for routine patient care. It is neither stored nor organized in a manner conducive to efficient provider retrieval. In addition, there is no interoperable application for clinicians to store, share, and retrieve data from a centralized repository. The lack of standardization, exchange, and interoperability result in a significant underutilization of clinical data for pediatric scoliosis patient care and clinical research.

To address the challenges of healthcare information exchange and interoperability, we proposed a healthcare information system for pediatric patient care leveraging a new health care data exchange standard, FHIR from Health Level-7 (HL7). The overall system configuration diagram is shown in Figure 1. Our objective was to create a generalizable clinical infrastructure that would enable the harmonization of heterogeneous multi-site, multi-modal data to a centralized data repository and facilitate data interoperability through a web interface for both clinical and research purposes. First, we extracted existing data from local storage in multiple sites, transformed heterogeneous multi-modal data into FHIR resources for standardized structures, and loaded transformed data to the centralized data repository. This proposed extract, transform, and load (ETL) process could serve as an efficient pipeline for the data harmonization and standardization to create a FHIR-enabled clinical research data repository. Second, we developed a SMART-on-FHIR application with an interactive GUI design as a frictionless mechanism to access and retrieve patient data for clinical and research studies. The web application development enables clinicians and researchers to create, read, upload, and delete (CRUD) information, as well as facilitate interoperable data query and cohort definition. Furthermore, this multi-site and multi-modality clinical data cloud infrastructure could be generalized to multiple diseases (e.g., scoliosis, craniofacial, and orthopedic) for advanced pediatric patient care. The main contribution of our work is three-fold:

- We established an ETL framework to effectively harmonize heterogeneous multi-site, multi-modal data into standardized data structures leveraging FHIR resources for pediatric patient data management and clinical insight generation.
- We developed a SMART-on-FHIR web application for more efficient and timely clinical data access with a user-friendly graphic user interface to improve data interoperability and advance pediatric patient care for clinicians and researchers.
- We illustrate the generalizability of our system design within multi-site and multi-model healthcare information systems to achieve data interoperability to improve patient outcomes across diverse treatment areas without the need to create new infrastructure for each area.

2 RELATED WORKS

The wealth of health data available and collected daily around the world provides enormous opportunities for applications in artificial intelligence, cloud computing and machine learning to create solutions for millions of patients worldwide – if and only if the data can be turned into useful information [12].

2.1 Healthcare Information Exchange

Due to the diversity, volume, and distribution of ingested data, clinical sites typically operate independently, amplifying the challenge of data harmonization and sharing. To address this issue, researchers integrated FHIR into ETL tasks to achieve data mapping from multiple sites to an integrated data repository [9, 11]. HL7 FHIR is a healthcare information exchange standard describing data formats and elements (i.e., resources) for EHRs. For example, a *Patient* resource in FHIR is a data structure that contains basic patient information such as name and contact information. This strategy challenges traditional methods of storing data in various document formats used by the majority of healthcare providers and is widely embraced by a number of leading players in the health care informatics field [15].

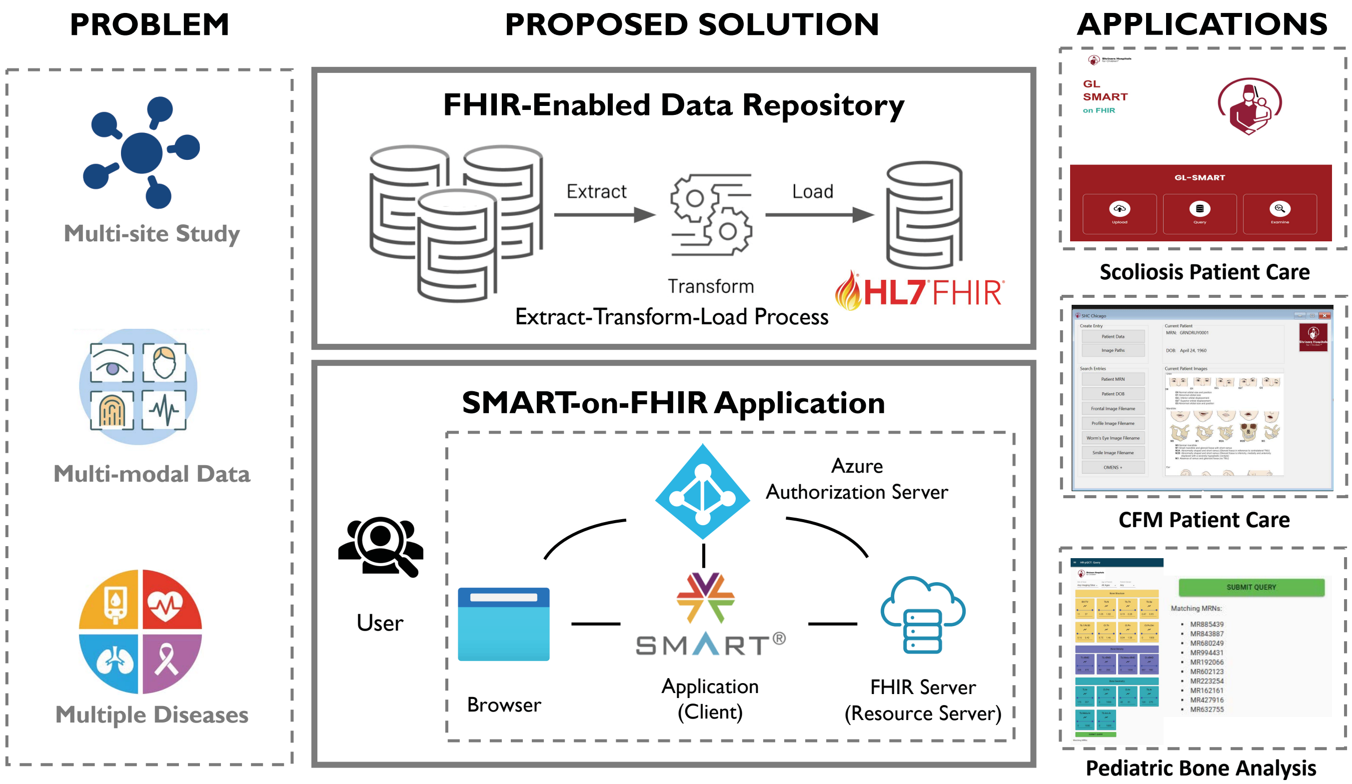

Figure 1: System diagram of the generalizable clinical infrastructure for pediatric healthcare information exchange. First, we generate a FHIR-enabled ETL pipeline for multi-site and multi-modal data harmonization to a standardized research data repository. Second, we developed a SMART-on-FHIR application to facilitate data access through a web interface for both clinical and research purposes. Lastly, we demonstrated the generalizability of our framework in improving pediatric patient care on three different clinical case studies (i.e., scoliosis, craniofacial, and rare metabolic/orthopedic conditions).

In multi-site clinical applications, Kiourtis et al. [11] proposed a solution to the general problem of multi-site healthcare interoperability by transforming healthcare datasets and aligning the data with an ontology to generate possible transformations between the data and FHIR resources. These FHIR-based data mapping mechanisms are able to take full advantage of a data-intensive environment without losing the real-world complexity of health when dealing with heterogeneous data from multiple sites. Similarly, Garza et al. [7] developed a systematic mapping approach for leveraging the FHIR standard in multi-center clinical trials for supporting the data collection needs of multi-site clinical research studies. Additionally, Ulrich et al. [18] extended existing FHIR data element resources to facilitate the translation of metadata information for annotating and mediating purposes in the North German Tumor Bank of Colorectal Cancer. In our study, we seek a general solution from a technology infrastructure standpoint that requires an upfront selection of FHIR resources for multi-site data harmonization.

In multi-modal clinical data integration, Hong et al. [9] developed a framework to integrate structured and unstructured healthcare data into an interoperable format based on clinical natural language processing tools. In addition, Pfaff et al. [16] created an open-source clinical asset mapping application to convert multiple data models to the corresponding FHIR value sets to allow input of different modalities. Our ETL process implemented a FHIR-enabled data mapping framework from multi-modal heterogeneous data to corresponding object-oriented programming (OOP) based objects in FHIR resources. With increasingly standardized data elements, multi-site and multi-modality integration can be enhanced to enable data analytic and decision support with greater success.

2.2 Data Interoperability

FHIR is a new HL7 health standard that streamlines and standardizes healthcare communication by taking a resource-centric approach rather than a document-centric approach that enables data and system interoperability. As a two-part system providing the potential for data interoperability, FHIR consists of standardized healthcare data models and an application programming interface (API) for model interaction. FHIR Resources define a logical, consistent, and comprehensive way to represent and provide a Representational State Transfer (REST)ful API for interacting with and modifying healthcare data elements. Data resources are available as online services capable of reading and writing data in real-time.

Recent studies [5, 10, 14, 19] tried to solve the interoperability complexity among EHRs caused by heterogeneities in data representation and health care processes. FHIR data exchange is based on modern RESTful API standards, allowing developers to achieve data interoperability and create applications that can connect to a healthcare information system and exchange data using the same protocols as other modern websites. As a result, Mukhiya et al. [14] developed a resource server based on GraphQL and FHIR to resolve data level heterogeneity and establish communication between two heterogeneous EHR systems.

Specifically, Substitutable Medical Applications, Reusable Technologies (SMART)-on-FHIR is a widely-used open framework of web standards that enables the definition of healthcare applications based on FHIR data elements [13]. Duke Health created a custom SMART-on-FHIR server to handle API management activities and retrieve EHR data in FHIR-compatible formats [5]. This framework allowed researchers to integrate several compatible apps into the provider- and patient-facing Epic-based EHR workflows. Similarly, Wesley et al. [19] developed a novel SMART-on-FHIR application to conveniently collect patient-reported outcomes (PROs) from patients to enable seamless integration of PRO data. Specifically, one recent study [10] highlighted the lack of SMART-on-FHIR applications in spine disease studies and clinical practice to enable digital innovations to be seamlessly integrated with EHRs and a user interface.

SC has initiated the development of a FHIR-based infrastructure to modernize its data harmonization in anticipation of future informatics and improved patient care. Researchers can push data to a private FHIR-compliant research database via a user-facing application hosted on a private Microsoft Azure cloud network managed by SC Research Programs. Additionally, patient cohort curation is facilitated through database querying of FHIR resources used to manage data flow within the research repository, allowing clients to quickly identify patient cohorts for clinical research studies. Our application aims to create a standardized data research repository with infrastructure based on multi-site and multi-modal clinical data to eliminate the potential data heterogeneity and interoperability problems within the SC healthcare system.

3 FHIR RESEARCH REPOSITORY

3.1 Local Data Description

3.1.1 Patient Information. In this study, we included a patient cohort that was diagnosed with idiopathic scoliosis at SC-Greenville or SC-Lexington between 1/1/2010 and 1/28/2021 (see Figure 2). The study population was drawn from a consecutive case review, including all eligible pediatric patients of both genders, all ethnic backgrounds, and females of childbearing potential, in order to minimize bias in our study. 700 (including 537 idiopathic patients) and 318 (including 250 idiopathic patients) individual idiopathic scoliosis patient data were collected from SC-Greenville and SC-Lexington, respectively. To satisfy the Institutional Review Boards (IRB) requirement, we restricted the use of identifiable patient information and connected the proposed research data repository to the SC patient identity source within the SC EHR system using unique patient identifiers.

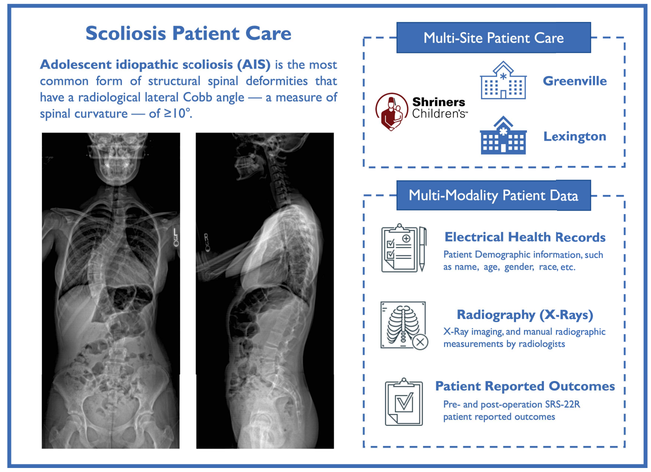

Figure 2: Description of patient data collected by SC for scoliosis patient care. Pre- and post- operative patient information along with patient EHRs are used to monitor patient rehabilitation.

Table 1: Description of Multi-modal Clinical Data Source in SC-Greenville and SC-Lexington

Multi-modal Data	Data Type	Number
Demographics	EHR	4,258
Procedural Data	Numerical and categorical	130,180
SRS-22r PROs	Numerical	1,320
Radiography	X-ray Imaging	2,224
Radio. Measurements	Numerical and categorical	45,280

3.1.2 Hospital Visit Information. We extracted hospital visit information for eligible patients, a pediatric population where multiple hospital visits are common including pre and post-operative visits. At least 200 data points pertaining to procedure-specific parameters, demographics, complications, and manually measured and human-curated radiographic measurements were present for each patient.

3.1.3 Patient Reported Outcomes. SRS-22r questionnaire [1] is a patient-reported outcome questionnaire that consists of 22 questions. These 22 questions cover different domains of topics, including function, pain, self image, mental health and satisfaction. Between 6/1/16 and 2/24/21, a total of 1,320 questionnaires were collected from 566 patients from both clinical sites (see Table 1).

3.1.4 Radiographic Measurements. Both SC sites have participated in the SSS SPP registry for quality improvement. This registry aims to collect comparative practice information that can be used to identify the most effective clinical procedures and enhance the quality of care. During each hospital visit, the registry collects approximately 40 radiographic measurements, including thoracic curve, lumbar curve, T1 tilt angle, etc. These radiographic parameters are manually calculated and verified from radiographs by clinical workers. A patient may have 2-7 separate visits, ranging from 80 to 280 radiographic measurements taken over the course of treatment.

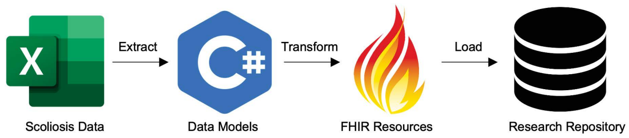

Figure 3: Overview of scoliosis data ETL process. Data is first extracted from Excel files to object-oriented models written in C#. The models are then used to transform data elements into FHIR resources which are loaded into the FHIR-enabled research repository.

3.2 FHIR Resources and Data Standardization

In this multi-site data-sharing framework, metadata mapping and data standardization are the core tasks before the transformation. This ETL process could serve as an effective pipeline for the harmonization and standardization of scoliosis patient data to create an FHIR-enabled data repository, as shown in Figure 3. All resource definitions and standardization are performed using the HL7 FHIR v4.0.1. *Patient*, *Encounter*, *ImagingStudy*, *QuestionnaireResponse*, and *Observation* resources are used in FHIR, as shown in Figure 4.

Our first step of FHIR integration is to identify the base resources covering essential information of scoliosis patient data. Specifically, we conducted an analysis between the scoliosis domain-specific data (i.e., SRS-22r questionnaire and radiographic measurements) and existing FHIR resources. Scoliosis patient data are classified into four meaningful categories: patients, visits, SRS-22r patient reported outcomes, and radiographic measurements. For every data element in each category, we investigated if there is a match of FHIR resources, and then compared details with value sets in data elements in the two domains. The detailed mapping of data elements and attributes from SC scoliosis patient data to FHIR is illustrated as follows.

3.2.1 Patient and Encounter. The *Patient* resource enables information to be organized according to SC patients using a patient-specific medical record number (MRN) to connect repository data to the rest of the EHR in the SC clinical healthcare information system. *Encounter* resources organize data into meaningful periods. As a result, we are able to integrate patient and hospital visit information into FHIR resources using precisely matched FHIR attributes and create an SRS-22r questionnaire using FHIR profiles.

3.2.2 Questionnaire, QuestionnaireResponse, and Observation. Within the scope of FHIR resources, a *Questionnaire* is an organized collection of questions intended to guide the collection of end-user responses. The *Questionnaire* defines the questions to be asked, their order and grouping, any intervening instructional text, and the constraints on the permissible responses. *Questionnaires* allow for precise control over the order, presentation, phraseology, and grouping of data, facilitating the collection of coherent, consistent data. We mapped the SRS-22r questionnaire into the *Questionnaire* resource to define a structured set of questions in FHIR. Specifically, we hierarchically organized the SRS-22r patient questionnaire into five different groups (Function, Pain, Self-image, Mental Health, Satisfaction with management), each containing related questions. The detailed mapping of SRS-22r questionnaire elements and attributes to their FHIR *Questionnaire* resource counterparts is shown in Figure 5. The responses to each question in the SRS-22r questionnaire are stored in the *Observation* derived from the corresponding *QuestionnaireResponse* resource.

3.2.3 ImagingStudy and Observation. Additionally, as the existing code system is insufficient to support detailed definitions of radiographic measurements, we implemented the FHIR *Observation* resource and a newly proposed code system (see Table. 2) for transferring main radiographic measurements based on the SSS dictionary. The *ImagingStudy* and *Observation* resources each define scoliosis imaging studies and different types of radiographic measurements, respectively.

4 SMART-ON-FHIR APPLICATION

Large healthcare organizations that collect vast amounts of data require an efficient system to transform the data into a consistent, usable format. To address the challenges with multi-modal, multi-site data harmonization at SC, we developed a web-based application using ASP.NET Model-View-Controller (MVC) technology that can effectively communicate with a FHIR-enabled data repository, as shown in Figure 6.

The MVC architectural design pattern allowed us to separate the application into three main components:

- `Model`: Models represent data structures in the application. Consistent and logical data structures allow the organization of heterogeneous data into homogeneous objects written in C#, a powerful object oriented programming language.
- `View`: Views represent the user interface in the application. Views allow users to perform CRUD operations on the data to and from the FHIR-enabled data repository.
- `Controller`: Controllers handle user interaction and work with models to transport data. Controller logic is also written in C#, enabling the application to use secure libraries for data transport to and from user views and the FHIR-enabled data repository.

The separation of application components into data, program logic, and user interface logic enables the application to be lightweight, easily testable, and require minimal code modifications when updates are required. The choice of using Microsoft's ASP.NET MVC technology based on C# allows seamless integration of the Microsoft .NET platform FHIR SDK, Hl7.Fhir.R4, and the ability to easily publish the web application on a cloud provider such as Azure which is used by SC. Uploading a compressed file containing the application to an Azure App Service Plan resource is the only prerequisite for launching the application.

Models were used to make data input granular, scalable and easily transferable to multiple endpoints. Models were developed by creating standard classes in OOP, reflecting the benefits of the OOP concepts of encapsulation and inheritance. Encapsulation refers to the security of data in a class from being manipulated in undesirable ways or accessed via external classes. The models are constructed so that the properties of each model, such as the MRN of a *Patient*, cannot be modified by any external part of the application logic, ensuring data security. Inheritance enables one class to inherit the attributes and methods of another class, which is useful for creating derivatives of existing models without creating new classes.

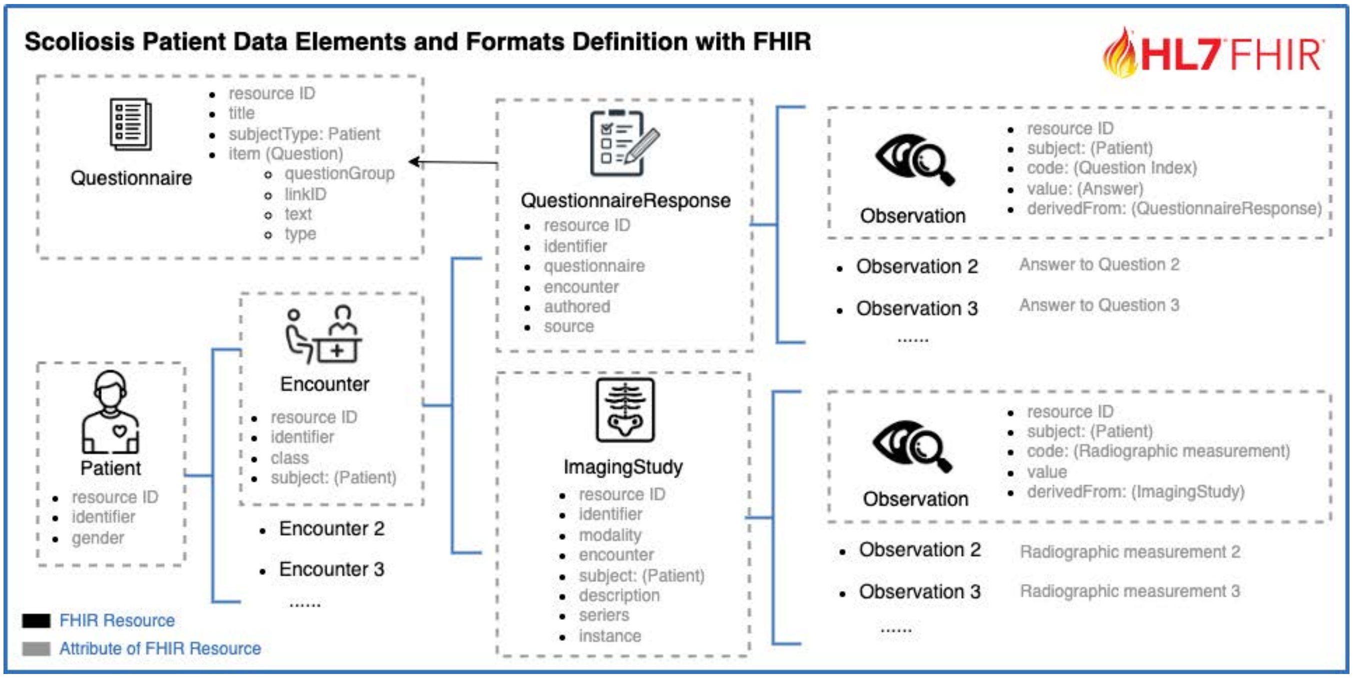

Figure 4: Standard scoliosis data elements as FHIR resources and their relationships. All resource definitions and standardization are performed using HL7 FHIR. (1) *Patient* resource for patient demographics information, (2) *Encounter* resource for each visit, (3) *Questionnaire* for SRS-22r patient reported survey, (4) *Observation* derived from *QuestionnaireResponses* for PROs, (5) *ImagingStudy* for spinal X-ray imaging, and (6) *Observation* derived from *ImagingStudy* for radiographic measurements.

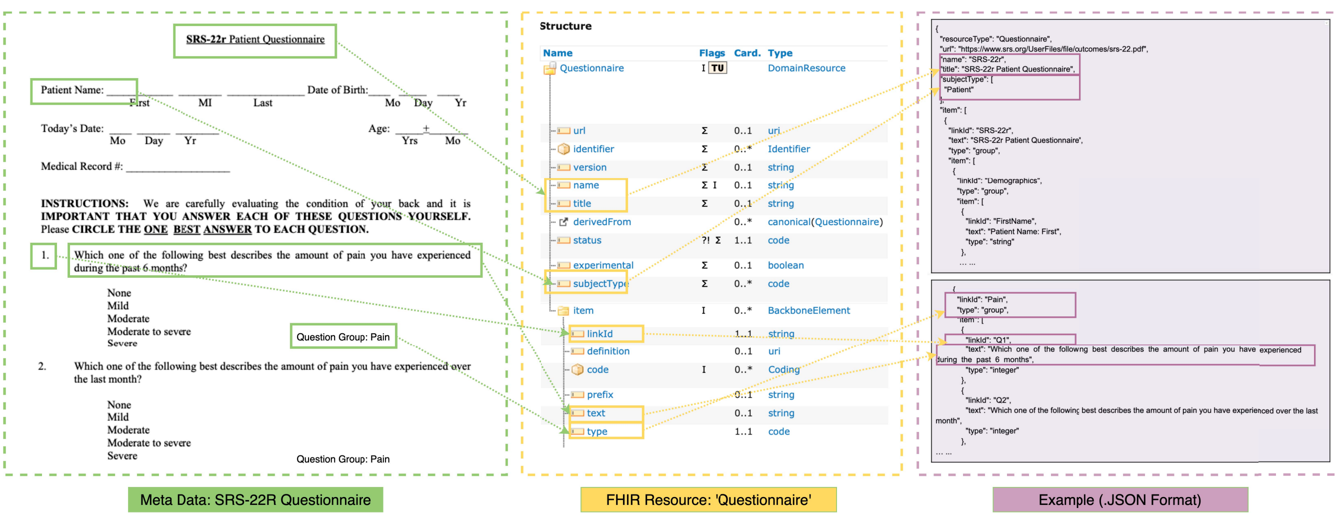

Figure 5: Transformation of the SRS-22r questionnaire into a FHIR *Questionnaire* resource. We mapped each element and attribute from meta data (Green) to their FHIR *Questionnaire* resource counterparts (Yellow) and provided an example of an SRS-22r *Questionnaire* in FHIR using JSON format (Purple).

To organize the data in an effective and logical manner, we created six models:

- `Basic Data Entry Model`: used to store basic information including MRN, date the entry was made into the system, visit date and visit type.
- `Lateral Radiographs Model`: used to store radiographic information from lateral radiographs, primarily Cobb measurements.
- `Posterior Anterior Radiographs Model`: used to store the radiographic information for the posterior and

Table 2: Coding System of Main Scoliosis Radiographic Measurements for *Observation* Resource in FHIR

	Variable Name	Description	Data Type	Acceptable Values
1	Upper_Thoracic_Curve	Cobb measurement of upper thoracic curve	Numeric	Decimals allowed but not required
2	Upper_Thoracic_Levels_Measured_Upper	Top vertebrae measured for curve	Category	List all vertebrae (C1 - S2)
3	Thoracic_Curve	Cobb measurement of thoracic curve	Numeric	Decimals allowed but not required
4	Thoracic_Levels_Measured_Upper	Top vertebrae measured for curve	Category	List all vertebrae (C1 - S2)
5	Lumbar_Curve	Cobb measurement of Lumbar curve	Numeric	Decimals allowed but not required
6	Lumbar_Levels_Measured_Upper	Top vertebrae measured for curve	Category	List all vertebrae (C1 - S2)
7	Coronal_C7_to_Central Sacral Vertical Line (CSVL)	Distance (mm) between C7 Plumbline and CSVL	Numeric	Decimals allowed but not required
8	Thoracic_Apical_to_C7_Plumb	Distance (mm) from center of thoracic apex to C7 Plumbline	Numeric	Decimals allowed but not required
9	Thoracic_Apical_Translation_to_CSVL	Distance (mm) from center of thoracic apex to CSVL	Numeric	Decimals allowed but not required
10	ThL-Lumbar_Apical_Translation	Distance (mm) from center of Thoracolumbar-Lumbar apex to CSVL	Numeric	Decimals allowed but not required
11	T1_Tilt_Angle	Tilt angle of bottom of T1 Vertebrae	Numeric	Decimals allowed but not required
12	T1_Tilt_Angle_Direction	Direction of T1 Tilt Angle	Numeric	Decimals allowed but not required
13	EIV_Angulation	Angle of bottom of end instrumented vertebrae	Numeric	Decimals allowed but not required
14	EIV_Translation	Distance (mm) from center of end instrumented vertebrae to CSVL	Numeric	Decimals allowed but not required
15	EIV_Disc_Angulation	Angle between bottom of end instrumented vertebrae and top of vertebrae directly beneath it	Numeric	Decimals allowed but not required
16	Classification_of_CSVL	Classification of Lumbar curve in relation to CSVL	Category	A; B; C;
17	Risser	Risser grade, reflection of skeletal maturity	Category	0; 1; 2; 3; 4; 5
18	Tri-Radiate	Status of tri-radiate cartilage, reflection of skeletal maturity	Category	Open; Closed

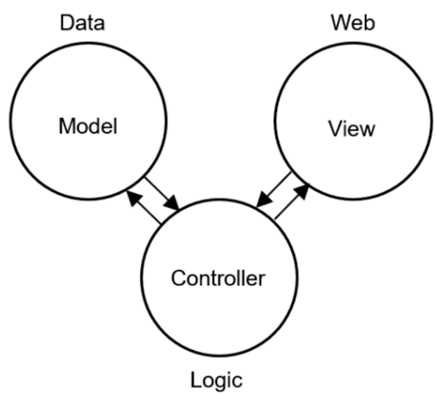

Figure 6: The MVC architecture enables logical separation of design components in our application. Models represent data structures, such as radiographs, which are converted into FHIR resources and transferred via the FHIR API to the FHIR-enabled research repository. Views represent the user facing web based portion of the application where data can be uploaded, viewed, and retrieved to and from the repository. Controllers handle the flow of data from models to the repository and the views.

anterior radiographs, such as Cobb measurements and curve directions.

- `Question Model`: used to store responses to a PRO questionnaire.
- `Radiographic Information Model`: used to store information such as whether the radiographic measures were originally completed on plain film or computer.
- `Radiographic Trunk Shape Model`: Used to store the trunk shift value.

Each model represents a unique data structure that isolates logical components in the SC Excel files where the data had previously been stored. To convert the models to FHIR, the Hl7.Fhir.Model package available for the C# programming language was used. The mappings from models to the FHIR resources used were:

(1) Basic Data Entry Model to FHIR *Patient* and FHIR *Encounter*
(2) Question Model to FHIR *Questionnaire*
(3) Radiograph Models to FHIR *ImagingStudy*

Specifically, *Patient* resources are used to store demographic and administrative related information about an individual. This was the logical FHIR resource to store each patient's MRN. *Encounter* resources are used to record information about an interaction between patients and providers. For the scoliosis data, the visit date and type, such as pre-operative, were recorded using *Encounter* resources. *Questionnaire* resources are designed to capture end-user responses from patients, providers and others in healthcare systems. Communicating data that is used in medical forms and defining specific aspects of the data such as what questions were asked, the order of the questions and answer choices is enabled by *Questionnaire* resources in flat lists or hierarchical structures. The results of *Questionnaire* resources are communicated via *QuestionnaireResponse* resources. *ImagingStudy* resources provide information on DICOM imaging studies, in this case scoliosis X-rays in DICOM format. Each study consists of a set of series that include a set of Service-Object Pair Instances (SOP Instances - images or other data) acquired or produced in a defined manner and of one modality (i.e, X-ray).

Controllers are the liaison between the front-end and back-end of the application and are connected directly to the views. The upload view enables clinicians to upload an Excel file using the web interface. The upload controller extracts the data from the uploaded Excel file, converts it into the appropriate model, then converts the model into the appropriate FHIR resource and finally sends all the FHIR resources to the FHIR-enabled data repository using the FHIR API in the Hl7.Fhir.Rest package via POST requests. The query controller contains the logic used to retrieve data for the FHIR-enabled data repository based on values that clinicians can enter directly in the query view web page. The controller sends the information to the FHIR-enabled research repository, which returns the appropriate filtered data in CSV format. The examine view can be used to quickly retrieve data for a single patient or group of patients and be viewed immediately in the web browser. The examine controller performs similar operations as the query controller, but its purpose is to provide immediate access to data that

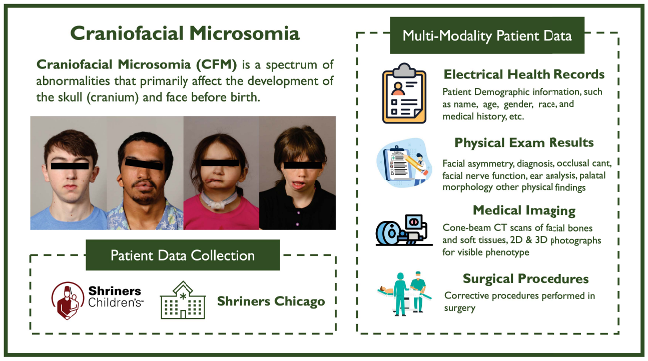

Figure 7: Data description of patient data collected by SC Chicago for CFM patient care. EHRs containing patient demographic information, diagnostic information including facial symmetry, medical imaging, and surgical procedure notes were part of the multi-modal data.

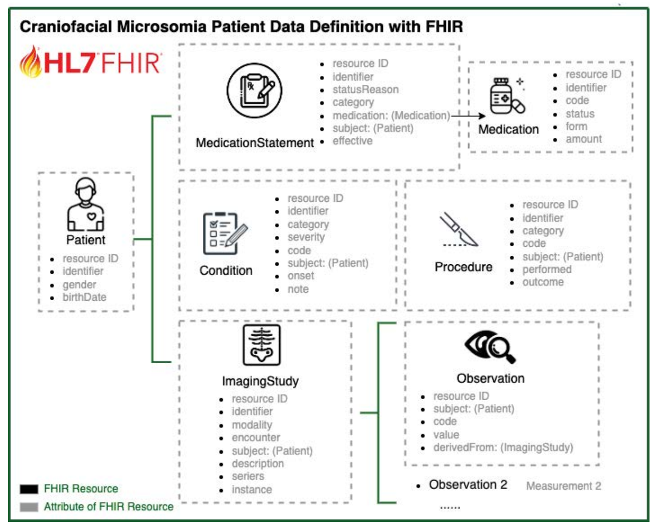

Figure 8: Standardized CFM data structure using FHIR resources. All resource definitions and standardization are performed using the HL7 FHIR.

is immediately required and can be viewed immediately as opposed to being stored in CSV format.

5 SYSTEM GENERALIZABILITY

The proposed framework has shown the ability to harmonize multi-site and multi-modal data at scoliosis patient care. To demonstrate the generalizability across multiple clinical use-cases across varying SC sites, we extended our application to craniofacial microsomia (CFM) and pediatric bone data normalization and querying.

5.1 Craniofacial Microsomia

CFM is an abnormality that affects skull development leading to part of the face being abnormally small and is the second most common craniofacial anomaly [4]. Despite the prevalence of the disease, the pathogenesis is not well understood [6] and has valuable applications in care optimization. Understanding the pathogenesis of the disease and how particular factors affect CFM phenotyping would provide insight into prognosis and treatment options by elucidating potential drug targets.

SC has collected multi-modal data at SC Chicago consisting of EHRs, disease severity measures, medical imaging, and surgical procedure data, as shown in Figure 7. These data were spread across multiple computer systems and cloud storage solutions severely limiting their access and use to establish meaningful insights. Using the same process as with scoliosis data, we were successfully able to leverage our general application framework to address these clinical data management challenges.

Despite the disparate nature of the existing CFM data, extending our application to organize and standardize all of it required only creating five new models to represent the different treatment areas:

- `Person Model`: used to store basic patient information including gender and date of birth.
- `Condition Occurrence Model`: used to store information about the condition of the subject.
- `Drug Exposure Model`: used to store information regarding medication treatments.
- `Observation Model`: used to store information during observations such as date, time and provider.
- `Procedure Model`: used to store procedure information such as date, time and provider.

Simply by creating new models and modifying small portions of the existing controller logic, a new research repository for CFM data was created, thereby avoiding the time-consuming development of new application logic that would have been necessary if building a novel application for each use case.

As in the scoliosis ETL process, CFM data was first extracted from the provided format and then transformed into models and FHIR resources (see Figure 8). The following mappings were used to convert the CFM models to FHIR resources:

(1) Person Model to FHIR *Patient*
(2) Condition Occurrence Model to FHIR *Condition*
(3) Drug Exposure Model to FHIR *MedicationStatement*
(4) Observation Model to FHIR *Observation*
(5) Procedure Model to FHIR *Procedure*

5.2 Pediatric Bone Analysis

The proposed framework was then utilized to facilitate the generation and analysis of High Resolution peripheral Quantitative Computed Tomography (HR-pQCT) data cohorts. HR-pQCT employs high resolution computerized tomography to quantify several specific 3-dimensional measurements of bone quality, geometry, and strength. The pediatric bone analysis data analyzed were generated on the HR-pQCT scanner and originally stored in PDF format. By extracting quality and structure properties using custom Python scripts and storing them using the same ETL process as scoliosis and CFM data, we were able to create an FHIR-enabled research repository for pediatric bone analysis with the same benefits.

The pediatric bone analysis required creating three new models:

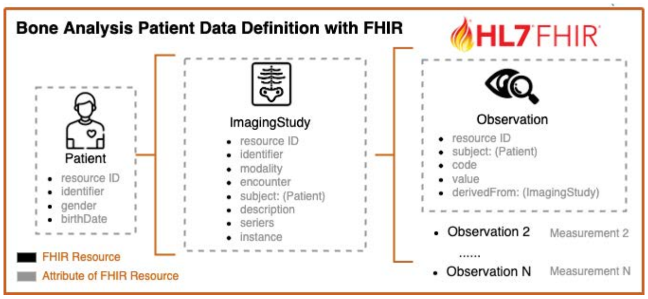

Figure 9: Standardized pediatric bone analysis data elements using FHIR resources and their relationships. All resource definitions and standardization utilise HL7 FHIR standards.

- `Person Model`: used to store basic information regarding the subject including gender and date of birth.
- `Image Model`: used to store information about the imaging study.
- `Observation Model`: used to store quantitative values derived from the imaging study.

The following mappings were used to convert the pediatric bone analysis models to FHIR resources, as shown in Figure 9:

(1) Person Model to *Patient* Resource
(2) Image Model to *ImagingStudy* Resource
(3) Observation Model to *Observation* Resource

5.2.1 System Generalization. The CFM implementation shows the generalizability of our application across different areas that must incorporate data harmonization and interoperability to unlock the potential of existing data to advance clinical research. The pediatric bone analysis case study provides yet another example of the value of our application framework. By not having to develop new core application logic for each unique clinical data challenge, significant time was saved while achieving all user objectives.

Our framework can be easily customized and applied to other clinical informatics data challenges. Once data is extracted from its initial format, such as an Excel file or PDF, the design of the application allows data elements to be immediately isolated into logical segments. Improving the value of other data sources, such as orthopedics datasets, would simply require defining the appropriate data models and appropriate FHIR resources.

Our application is also designed to meet the ISO standards for software quality including:

- `Functionality`: our application satisfies the stated needs of the end users. It is a suitable application for data aggregation and provides a secure infrastructure using robust authentication and authorization mechanisms built into ASP.NET.
- `Reliability`: the level of performance of the application can be maintained in its current state without any modifications and uptime is only dependent on a reliable web server.
- `Usability`: there is no new technology for users to learn who are familiar with using a web browser. The application is also operable on standard web browsers such as Google Chrome, Microsoft Edge, Mozilla Firefox, and Apple Safari.
- `Efficiency`: the software does not require any abnormal amount of computing resources to perform.
- `Maintainability`: there is no need to make modifications to the current application unless additional data inputs are desired. As demonstrated, in the case of adding new data, the application simply requires additional models and FHIR resource mappings - the remainder of the application logic is unchanged.
- `Portability`: the application is environment agnostic and can be run locally on Linux, MacOS or Windows.

Finally, our application design can also be generalized across various programming languages and infrastructures. We chose to use the ASP.NET framework given our familiarity with the C# programming language and the readily available HL7 FHIR packages for the framework. However, other programming languages such as Java and Python have their own MVC technology, and developers can use the design of our system to create an application based on their language of preference. Transferability of our approach to cloud infrastructure for hosting is also provider agnostic. Our application can easily be deployed to a cloud provider other than Azure simply by uploading a compressed file containing the application. Developers preferring Amazon Web Services (AWS), for example, would employ the same mechanism to deploy the application. Instead of using an Azure App Service Plan to deploy resources to, users may use an AWS Elastic Beanstalk instance.

6 DISCUSSION

In the healthcare industry, the lack of clinical data harmonization and interoperability in healthcare information exchange presents significant challenges. In this study, we have developed and implemented a systematic mapping approach for leveraging the HL7 FHIR standard across real-world multi-site clinical datasets. The mapping of study data elements from three diverse clinical case studies to FHIR resources provides insight into the utility of the health information exchange standard for supporting the data collection needs of multi-site clinical research studies.

We created a generalizable infrastructure for multi-site and multi-modal data standardization that consists of a back-end FHIR-enabled research repository accessed via a front-end web application interface. For providers at SC with the appropriate data permissions, all of a patient's data can be accessed via the patient's MRN. Additionally, aggregate patient data can be collected for clinical research purposes by creating cohort definitions based on specified radiographic and PRO values. A researcher looking for data to test their hypothesis no longer needs to consider the time cost of collecting, organizing, and standardizing the data from disparate sources.

We have released the pediatric bone analysis web application (https://apps.smarthealthit.org/app/hr-pqct-data-accession) in the SMART-on-FHIR application gallery for demonstration and promotion on real-world adoption. Local deployment examples for scoliosis and CFM patient care applications are also available in: https://github.gatech.edu/ahornback6/acmbcbdemo. We are in the process of conducting a usability study of our applications with physicians and clinical researchers to evaluate implementation success by recruiting at least five domain experts. We will use the System Usability Scale (SUS) survey to measure general usability

[2]. Specifically, we will ask users to rate the usability and helpfulness of our application when seeing patients on a five-point scale (5 indicating strong agreement, and 1 indicating strong disagreement). We will record the responses and use them to guide future iterations. If possible, we will expand to recruit 20 domain experts for more definitive quantitative studies [3].

Our goal was to provide a streamlined and integrated system for clinical data management and access that does not disrupt existing workflows. Among the challenges presented by this objective was ensuring that our user interface was simple to use and that all components were clearly labeled. Throughout the application development process, expert feedback from clinical experts and researchers was gathered. This insight was used to guide feature implementation and create a user interface that could be integrated seamlessly into the existing clinical workflow. Additionally, we discovered the value of simple search functions for assisting with basic data quality tasks, which we incorporated into our final user work.

With the incorporation of additional clinical sites and the collection of data from additional modalities, inconsistencies in variable definitions and formats will have a greater negative impact on data standardization and integration. Additionally, future work incorporating machine learning will face new challenges. While machine learning models are capable of extracting critical features from clinical data such as radiographic measurements, the majority of studies attribute their high accuracy to a large-scale, high-quality training datasets with low data heterogeneity. This is challenging to achieve in multi-site and multi-modal clinical environments.

Our future works include integrating machine learning models to leverage these large clinical datasets of rare pediatric pathologies to provide data-driven clinical support and insights into patient prognosis and optimal treatment plans. We believe our innovative solution to common data interoperability and access challenges establishes a platform for big data analysis to improve pediatric patient care. A specific future application of our approach involves developing a clinical tool to aid in surgical planning and patient rehabilitation by using machine learning models to analyze patient radiographic and demographic data to improve patient reported outcomes with the goal of decreasing post-surgical scoliosis complications.

7 CONCLUSIONS

In this study, we designed and implemented a secure and interoperable data harmonization system to improve data value across multiple clinical sites and data types. We developed a FHIR-enabled data repository by converting existing multi-modal clinical data in heterogeneous formats to standardized FHIR resources. We also improve data value by allowing easy and secure access to clinical users via easy-to-use application interfaces. Our data access and standardization solution enables healthcare providers to upload existing data easily, to query individual patient data, and to retrieve curated patient cohorts to facilitate clinical study recruitment. Finally, we demonstrated the generalizability of our technology architecture across three disparate real-world clinical case studies. The value of our work is that it provides a solution for large international healthcare systems to generate value and clinical insights from their data across clinical domains with the ultimate goal to improve pediatric patient care.

REFERENCES

[1] Marc A Asher, Sue Min Lai, R Chris Glattes, Douglas C Burton, Ahmet Alanay, and Juan Bago. 2006. Refinement of the SRS-22 health-related quality of life questionnaire function domain. *Spine* 31, 5 (2006), 593–597.

[2] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[3] JM Christian Bastien. 2010. Usability testing: a review of some methodological and technical aspects of the method. *International journal of medical informatics* 79, 4 (2010), e18–e23.

[4] Craig B Birgfeld and Carrie Heike. 2012. Craniofacial microsomia. In *Seminars in plastic surgery*, Vol. 26. Thieme Medical Publishers, 091–104.

[5] Richard A Bloomfield Jr, Felipe Polo-Wood, Joshua C Mandel, and Kenneth D Mandl. 2017. Opening the Duke electronic health record to apps: implementing SMART on FHIR. *International journal of medical informatics* 99 (2017), 1–10.

[6] Q Chen, Y Zhao, G Shen, and J Dai. 2018. Etiology and pathogenesis of hemifacial microsomia. *Journal of Dental Research* 97, 12 (2018), 1297–1305.

[7] Maryam Y Garza, Michael Rutherford, Sahiti Myneni, Susan Fenton, Anita Walden, Umit Topaloglu, Eric Eisenstein, Karan R Kumar, Kanecia O Zimmerman, Mitra Rocca, et al. 2020. Evaluating the Coverage of the HL7® FHIR® Standard to Support eSource Data Exchange Implementations for use in Multi-Site Clinical Research Studies. In *AMIA Annual Symposium Proceedings*, Vol. 2020. American Medical Informatics Association, 472.

[8] Alexander Hoerbst and Elske Ammenwerth. 2010. Electronic health records. *Methods of information in medicine* 49, 04 (2010), 320–336.

[9] Na Hong, Andrew Wen, Feichen Shen, Sunghwan Sohn, Sijia Liu, Hongfang Liu, and Guoqian Jiang. 2018. Integrating structured and unstructured EHR data using an FHIR-based type system: a case study with medication data. *AMIA Summits on Translational Science Proceedings* 2018 (2018), 74.

[10] Aditya V Karhade, Joseph H Schwab, Guilherme Del Fiol, and Kensaku Kawamoto. 2021. SMART on FHIR in spine: integrating clinical prediction models into electronic health records for precision medicine at the point of care. *The Spine Journal* 21, 10 (2021), 1649–1651.

[11] Athanasios Kiourtis, Argyro Mavrogiorgou, Andreas Menychtas, Ilias Maglogiannis, and Dimosthenis Kyriazis. 2019. Structurally mapping healthcare data to HL7 FHIR through ontology alignment. *Journal of medical systems* 43, 3 (2019), 62.

[12] Moritz Lehne, Julian Sass, Andrea Essenwanger, Josef Schepers, and Sylvia Thun. 2019. Why digital medicine depends on interoperability. *NPJ digital medicine* 2, 1 (2019), 1–5.

[13] Joshua C Mandel, David A Kreda, Kenneth D Mandl, Isaac S Kohane, and Rachel B Ramoni. 2016. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association* 23, 5 (2016), 899–908.

[14] Suresh Kumar Mukhiya and Yngve Lamo. 2021. An HL7 FHIR and GraphQL approach for interoperability between heterogeneous Electronic Health Record systems. *Health Informatics Journal* 27, 3 (2021), 14604582211043920.

[15] D Munro. 2014. Setting healthcare interop on fire. *Forbes* (2014).

[16] Emily Rose Pfaff, James Champion, Robert Louis Bradford, Marshall Clark, Hao Xu, Karamarie Fecho, Ashok Krishnamurthy, Steven Cox, Christopher G Chute, Casey Overby Taylor, et al. 2019. Fast healthcare interoperability resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. *JMIR medical informatics* 7, 4 (2019), e15199.

[17] Rishi Saripalle, Christopher Runyan, and Mitchell Russell. 2019. Using HL7 FHIR to achieve interoperability in patient health record. *Journal of biomedical informatics* 94 (2019), 103188.

[18] Hannes Ulrich, Ann-Kristin Kock, Petra Duhm-Harbeck, Jens K Habermann, and Josef Ingenerf. 2016. Metadata repository for improved data sharing and reuse based on HL7 FHIR.. In *MIE*. 162–166.

[19] Deliya B Wesley, Joseph Blumenthal, Shrenikkumar Shah, Robin A Littlejohn, Zoe Pruitt, Ram Dixit, Chun-Ju Hsiao, Christine Dymek, and Raj M Ratwani. 2021. A novel application of SMART on FHIR architecture for interoperable and scalable integration of patient-reported outcome data with electronic health records. *Journal of the American Medical Informatics Association* 28, 10 (2021), 2220–2225.

[20] Claudia Williams, Farzad Mostashari, Kory Mertz, Emily Hogin, and Parmeeth Atwal. 2012. From the Office of the National Coordinator: the strategy for advancing the exchange of health information. *Health affairs* 31, 3 (2012), 527–536.