Journal of the American Medical Informatics Association Advance Access published July 21, 2015

SMART on FHIR Genomics: Facilitating standardized clinico-genomic apps

RECEIVED 20 December 2014
REVISED 10 April 2015
ACCEPTED 18 April 2015

Gil Alterovitz[1,2,3,*], Jeremy Warner[4,5,*], Peijin Zhang[6,*], Yishen Chen[7], Mollie Ullman-Cullere[8], David Kreda[2], Isaac S. Kohane[1,2,3]

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.
OXFORD UNIVERSITY PRESS

ABSTRACT

Background Supporting clinical decision support for personalized medicine will require linking genome and phenome variants to a patient's electronic health record (EHR), at times on a vast scale. Clinico-genomic data standards will be needed to unify how genomic variant data are accessed from different sequencing systems.

Methods A specification for the basis of a clinic-genomic standard, building upon the current Health Level Seven International Fast Healthcare Interoperability Resources (FHIR®) standard, was developed. An FHIR application protocol interface (API) layer was attached to proprietary sequencing platforms and EHRs in order to expose gene variant data for presentation to the end-user. Three representative apps based on the SMART platform were built to test end-to-end feasibility, including integration of genomic and clinical data.

Results Successful design, deployment, and use of the API was demonstrated and adopted by HL7 Clinical Genomics Workgroup. Feasibility was shown through development of three apps by various types of users with background levels and locations.

Conclusion This prototyping work suggests that an entirely data (and web) standards-based approach could prove both effective and efficient for advancing personalized medicine.

Keywords: genomics, data sharing, clinicogenomics, data warehouse, i2b2, SMART, FHIR, HL7, standards, EMR, EHR

BRIEF COMMUNICATION

BACKGROUND

The progress of medical care and health services research increasingly depends upon combining different types, sources, and volumes of data efficiently. Specific plans to apply "big data" solutions to produce individually tailored clinical decision support (CDS) for a patient at the point of care necessitate overcoming obstacles that arise from the different ways that data are collected and coded into electronic systems. Thus, it is not just a matter of data scale but also a matter of reaching agreements on how these data are represented and accessed. Without such standards, the ability of both researchers and developers to create tools that make the best use of such valuable data is severely limited.

The development of quick and cost-efficient DNA sequencing techniques has led to a dramatic growth in the volume of human genome sequencing data.[1] Genomics data are considered to be large and unwieldy, which explains why several conflicting systems for data management and storage already exist. Most genomic data systems offer application protocol interfaces (APIs) that reflect their underlying data storage formats. For example, Illumina, Inc. offers variant APIs (in their Basespace product) that map directly to the Variant Call Format (VCF). There are obvious benefits to this arrangement when data is being used for research. In particular, researchers maintain compatibility with software operating on raw data files. Other proprietary APIs include those of GenoSpace, LLC and Seven Bridges Genomics, Inc. These APIs focus on operating on genomic data stored in a cloud. Another example that is focused on communicating genomic information in the cloud is the Global Alliance for Genomics and Health (GA4GH). Rather than a proprietary API controlled by a single company, GA4GH brought together a number of stakeholders, including Substitutable Medical Applications & Reusable Technologies (SMART) on Fast Health Interoperability Resource (FHIR) Genomics. In the clinical genomics field, a previous effort by Health Level Seven International (HL7®) involved communicating variants, in a message format, between the electronic health records (EHRs) of Partners® HealthCare and Intermountain HealthCare in a demonstration project. It can be argued that genomic information may not be suitable for the message format used by HL7. New web technologies, such as Representational State Transfer (REST)-based APIs/web services, have recently been adopted by HL7 and hold great promise in this regard. In addition, authentication and user interface issues need to be standardized. Finally, it may be that a simpler, developer-centric approach is needed for wider adoption of clinicogenomic standards, as has been the focus of SMART®.[2,3]

For application developers and clinicians, creating an abstraction layer above specific file formats offers important advantages. First, although the tools and technology of DNA sequencing continue to progress rapidly and develop divergences in what is stored within sequencing systems, gene and variant data will likely continue to be used in clinical applications. Second, an API that directly maps to sequencing files may involve unnecessary details (eg, sequence alignment), which are not used by the majority of developers and clinicians. Conflicting vendor approaches pose challenges to end-users, including physicians, caregivers, patients, and medical researchers, and developers, who must build solutions to work with genomics data across multiple formats or else forego valuable data.[4,5] It is unsurprising that both of these groups would like some form of data standardization to succeed.[6] However, data standardization may not always fit the technical needs of a rapidly evolving discipline, and, even if they do, adoption of standardization measures typically requires extensive work with an uncertain payoff. Third, an API needs to be linked to a standards organization, which creates standards that are and will be used by EHR vendors, clinicians, and government mandates (eg, Meaningful

Correspondence to Gil Alterovitz, 10 Shattuck St, #311B, Boston, Massachusetts, United States, 02115, USA; ga@alum.mit.edu

For numbered affiliations see end of article.

Use Stage 3). Finally, the API should be able to fully contextualize genomics information with other clinical data for clinical care, subsequent outcomes analysis, and discovery research. For example, this would enable capturing evidence of genetic test benefits for facilitating coverage of those tests by insurance companies and the Centers for Medicare & Medicaid Services (CMS).

To encourage adoption of genomics standards, SMART on FHIR Genomics has been conceived by extending SMART on FHIR. SMART on FHIR emerged from the combination of the SMART Platform project[2,3] and the HL7® FHIR® effort.[7] SMART on FHIR combines FHIR-compliant clinical data access from EHR systems with web standards to launch web and mobile apps from a user's EHR session, including FHIR's REST-based API for clinical data, OAuth 2.0 authentication, and HTML5 to support web and native mobile apps.[3]

METHODS

SMART on FHIR Genomics specifies genomic variant data resource definitions to support the development of clinico-genomic apps. Data can be sourced from EHR systems or separate sequencing systems. Genomic data systems that comply with SMART on FHIR Genomics will be able to field data queries and return variant data results predictably, irrespective of their internal approach to data storage. The key effort for a data provider is simply to implement a SMART on FHIR Genomics data adaptor, which creates a binding to convert between the standard SMART on FHIR Genomics format and the provider's native format.

We created the SMART on FHIR Genomics specification by building on top of a large set of existing technologies, consistent with the FHIR ideal to provide a codeable concept for any given data element without "reinventing the wheel." For genomic data nomenclature, we adopted: the systematized Nomenclature of Medicine, Clinical Terms, for diseases and qualifier values; the Human Genome Variation Society mutnomen syntax, for mutation names and locations;[8] the Consensus Coding Sequence representation, for genomic regions;[9] and the Human Genome Organization Gene Nomenclature Committee (HGNC a.k.a. HUGO) symbols and identifiers, for common gene names.[10]

For representing queries and returning structured payloads of standards-coded clinical genomics data, we provide definitions based on the initial FHIR Draft Standard for Trial Use (DSTU1). FHIR adopts contemporary web technology that is convenient both for data providers and consumers.[11] SMART on FHIR uses FHIR's profiling feature to lock down nomenclature requirements for consistent semantics in clinical data. SMART on FHIR Genomics leverages FHIR's formal mechanism for resource extension to handle multiple types of clinical genomics data, including: wrappers for genomic data files (eg, VCFs), messaging services for genomic lab results, and document services for interpretative genomic reports.

Resource and Extension Definitions

Following FHIR protocol, we propose three new FHIR resource and extension definitions for variant data: the Sequence resource, to represent patient genetic information; the SequencingLab extension, to describe the sequencing technology used for sequence generation; and the GeneticObservation extension, to link a phenotype to variant data (Table 1). These definitions are discussed further below:

- The Sequence resource represents the raw genetic information of a patient and contains information about the specific read given by a sequencer for amino acid, RNA, or DNA sequences. The resource definition is designed in an abstract, non-format-specific way, such that developers can focus on the genetics, rather than native file formats.
- The SequencingLab extension of the Procedure resource[12] is a container for holding results about the specific Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory or research laboratory used to sequence a patient's genome. It may be useful for clinical documentation (for keeping track of how each patient's genotypes were determined). Our extension definition links to GA4GH data repositories and currently allows developers a "back door" for utilizing raw sequencing files by incorporating references to GA4GH's DataSet API endpoint.
- The GeneticObservation extension of the Observation resource[13] provides information about a patient's phenotype and genotype relationship. It links a patient phenotype to a set of genotype-phenotype relationships associated with that trait. As an example use case, one GeneticObservation resource may describe cancer risk, and clinicians can query that resource to discover patient variants that affect their risk. GeneticObservation supports both germline and somatic variants.

Full specification of the proposed resource and extensions may be found in Table 1 or on the SMART on FHIR Genomics website (Table 2). Figure 1 illustrates the relationships among the defined resources in a typical use case.

RESULTS

Figure 2 depicts one type of data coordination to feed a clinico-genomic app using SMART on FHIR augmented with SMART on FHIR Genomics. In such an app, clinical data from an EHR system is combined with genomics data from sequencing systems, all via SMART-standardized API calls. An app might identify, for example, an unidentified clinical problem using variant data for that patient. A different app might tap a gene-drug interaction database to propose or contraindicate certain drug therapies.

In practice, we refined SMART on FHIR Genomics by working with outside app developers with no a priori knowledge of SMART or FHIR. One group consisted of undergraduate students from universities unaffiliated with the authors of this paper. A second group consisted of a team of professional programmers at Vanderbilt University Medical Center. Both groups successfully developed useful clinico-genomics apps within a short timeframe. Based on their testing and our own, we iteratively refined the three SMART on FHIR Genomics definitions. With additional refinements, we believe our design could be offered for eventual HL7 balloting for a later release of FHIR. In the meantime, we have been providing input to the HL7 Clinical Genomics Workgroup and FHIR Management Group.

Figure 3 depicts three clinico-genomic apps that leverage SMART on FHIR and SMART on FHIR Genomics. These apps access, calculate, and present CDS and other types of patient information in the areas of diabetes, genetic risk, and cancer diagnosis. For more information, please visit the links located in Table 2.

DISCUSSION

The medical research community has increasingly recognized that creating "omics" data standards is problematic.[6] Tenenbaum et al.[6] presented a vision of "a standards-compliant, integrated data repository for clinical and 'omics' data, among other types," but believed that achieving this vision would be very difficult. They posited, moreover, that data standards would work best if they were concealed from researchers and that software developers, medical informaticists, and data curators, perhaps, would be "better equipped to delve into data standards than would be a clinician or bench scientist, but even they are typically not experts in specialized standards."[6]

Table 1: Description of the additional FHIR-compatible specifications (a. Sequence resource, b. SequencingLab extension to the FHIR Procedure resource, c. GeneticObservation extension to the FHIR Observation resource) defined by the SMART on FHIR Genomics API

Field Name	Field Type	Cardinality	Description
a. Sequence			
GenomeBuild[1]	String	1	Assembly of the sequence
Type[2]	String	1	Type of the sequence (Protein, DNA, RNA), SNOMED-CT
Quantity[2]	Quantity	0..1	Quantity of the sequence
ReferenceSeq[1,2] (ReferenceAllele)	String	0..1	Reference of the sequence (IUPAC format)
ObservedSeq[1,2] (ObservedAllele)	String	1	Read string for the sequence (IUPAC format)
CIGAR[1]	String	0..1	CIGAR string for the sequence
Source.Sample[1,2] (GenomicSourceClass)	Code	0..1	Source of the sequence. The genomic class of the variant: Germline, Somatic, or Prenatal. Associated with LOINC answer list: 48002-0.
Source.Lab[2]	SequencingLab	0..1	Laboratory source of the sequence
Chromosome[1]	String	1	Chromosome of the sequence.The chromosome containing the genetic finding values should be 1-23, X, Y, mito, viral, bacteria.
StartPosition[1,2] (GenomicStart)	Integer	1	Start of the sequence
EndPosition[1,2] (GenomicStop)	Integer	1	End of the sequence
Species[1]	CodeableConcept	1	Species identifier (NCBI taxonomy)
PatientID[2]	Patient	1	Genetic laboratory's patient identification for this sequence
b. SequencingLab			
GeneticsLaboratory[4]	Organization	1	Sequence Laboratory organization
Repository[2]	Uri	1	Repository for this laboratory (GA4GH type)
DatasetId[2]	String	1	Dataset identification of a laboratory folder containing multiple sequences
c. GeneticObservation			
AssessedCondition[1]	Condition	1	Condition described by this observation
SourceSeq[2]	Sequence	0..*	Sequence resource linked to this observation
DNASequenceVariation[4]	String	0..*	HGVS nomenclature for cDNA variant
GeneId[4]	CodeableConcept	0..*	HGNC identifier and symbol
VariantTranscript ReferenceSequenceId[4]	CodeableConcept	0..*	cDNA reference sequence identifier either RefSeq or ENSEMBL
DNASequenceVariationType[4]	CodeableConcept	0..*	Classification of variant change using LOINC Answer List values 48019-4 or Sequence Ontology values
DNARegionName[4]	String	0..*	Gene region containing the variant, eg, Exon 19
ProteinReference SequenceId[4]	CodableConcept	0..*	Protein reference sequence identifier either RefSeq or ENSEMBL
AminoAcidChange[4]	String	0..*	HGVS nomenclature for amino acid change
AminoAcidChangeType[4]	CodableConcept	0..*	Classification of variant change using LOINC Answer List values 48019-4 or Sequence Ontology values
VariationId[4]	CodableConcept	0..*	Variant identifier in ClinVar, dbSNP, or COSMIC
AlleleName[4]	String	0..*	Common name for variant for display purposes
AllelicState[4]	CodableConcept	0..*	Level of occurrence of the DNA variation in relation to the genomic context. LOINC answer list: LOINC 53034-5
Subject[3]	Patient	1	Genetic laboratory's patient identification for this sequence
Specimen[3]	Specimen	1	Specimen source of data
Interpretation[3]	CodeableConcept	0..*	Interpretation of the effect of this observation. Uses "Observation Interpretation Codes" value set of FHIR.
Comment[3]	String	0..*	Comments on this variant

API, application protocol interface; CIGAR, Compact Idiosyncratic Gapped Alignment Report; COSMIC, Catalogue of Somatic Mutations in Cancer; FHIR, Fast Health Interoperability Resource; GA4GH, Global Alliance for Genomics and Health; HGNC, HUGO Gene Nomenclature Committee; HGVS, Human Genome Variation Society; IUPAC, International Union of Pure and Applied Chemistry; LOINC, Logical Observation Identifiers Names and Codes; NCBI, National Center for Biotechnology Information; SMART, Substitutable Medical Applications & Reusable Technologies; SNOMED-CT, Systematized Nomenclature of Medicine - Clinical Terms.

[1] Adopted from SMART into the Health Level Seven International (HL7) Clinical Genomics Standard Profile for Genetics. Where elements were constrained for integration into the observation profile, which is limited to genetic information, the FHIR Draft Standard for Trial Use 2 (DSTU2) name is in parenthesis.

[2] Sequence and Sequencing Lab Resources are undergoing HL7 review for inclusion in the FHIR Draft Standard for Trial Use 3 (DSTU3).

[3] These components are inherited from the FHIR Observation resource.

[4] These components focus on HL7 standard reporting findings based on traditional genetic testing technologies.

Table 2: SMART on FHIR Genomics and FHIR reference web links

Name	Web Link
FHIR Wiki	http://wiki.hl7.org/index.php?title=FHIR
FHIR Specification Home (DSTU1)	http://www.hl7.org/implement/standards/fhir/index.html
FHIR DSTU2 Site	http://www.hl7.org/FHIR/2015May/index.html
FHIR Continuous Integration Build	http://hl7-fhir.github.io/
SMART on FHIR Documentation	http://docs.smartplatforms.org
SMART on FHIR Genomics Documentation	http://projects.iq.harvard.edu/smartgenomics
SMART on FHIR Genomics Server Code	http://github.com/dsrcl/fhir-genomics
Health Services Platform Consortium	http://healthcaresoa.org/
Argonaut Project Charter	http://geekdoctor.blogspot.com/2014/12/the-argonaut-project-charter.html

DSTU, Draft Standard for Trial Use; FHIR, Fast Health Interoperability Resource; SMART, Substitutable Medical Applications & Reusable Technologies.

Figure 1: Example of SMART on FHIR Genomics FHIR resource elements. A patient has three Sequence results, two SequencingLab sources, and three GeneticObservation results. GeneticObservation A is related to variants in Sequence A and B. GeneticObservations B is related to the variants in Sequence C, while GeneticObservation C describes a trait affected by a somatic mutation that is not linked to the germline sequence. Sequence A was determined at SequencingLab A, while Sequences B and C come from SequencingLab B. All eight resources are linked to the patient record, which is linked to additional clinical data (see Figure 2).

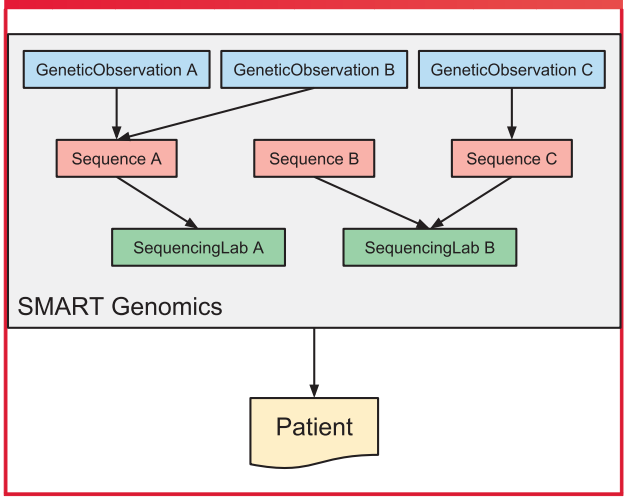
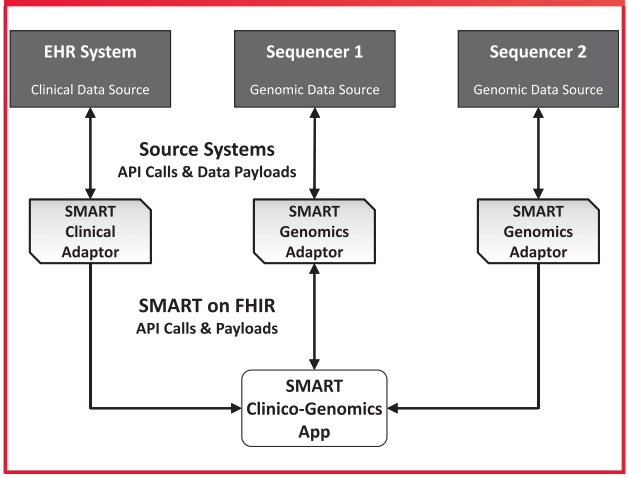


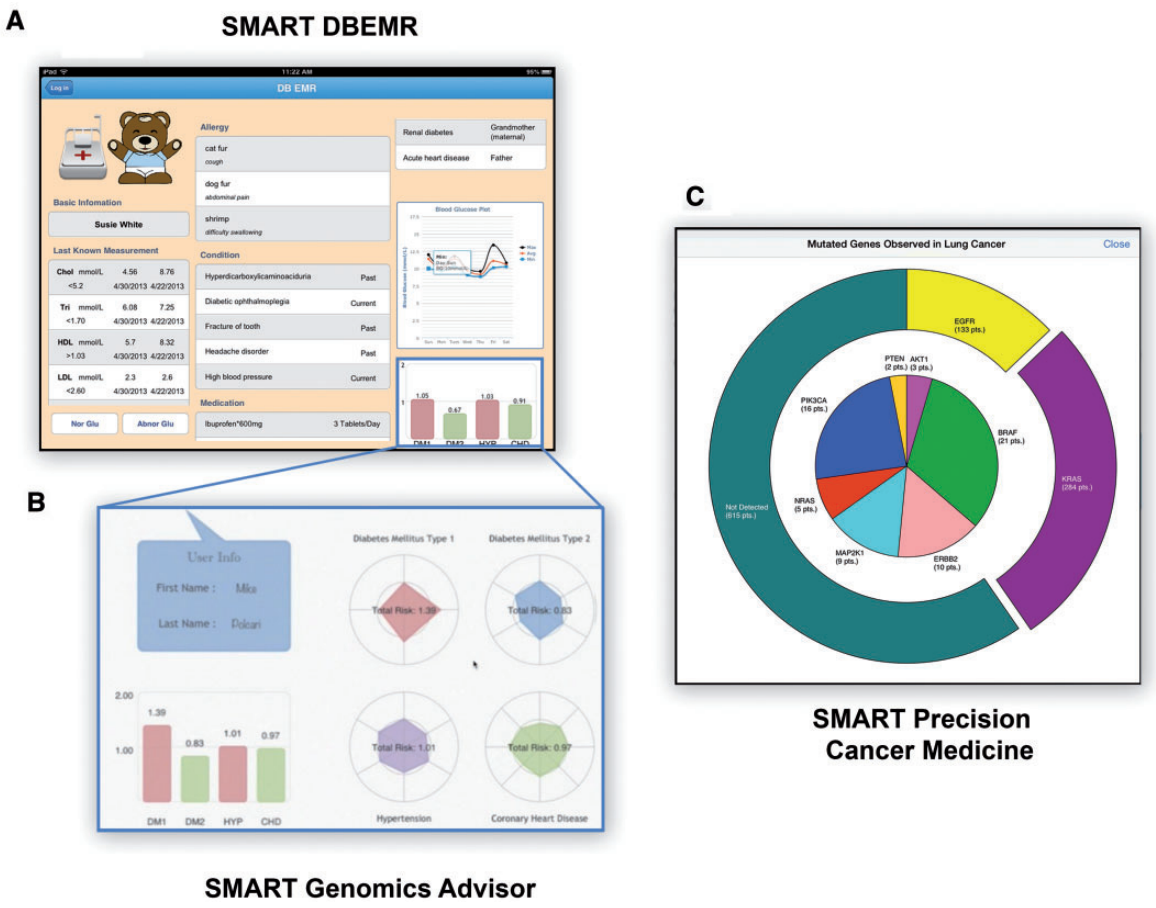
Figure 2: A combination of SMART on FHIR for clinical data and SMART on FHIR Genomics for genomic data would enable standardized support for clinico-genomic apps. Data adaptors must be individually implemented for each source system. App developers would use the SMART on FHIR API exclusively to query and receive (or send) payloads from/to these "black box" systems. In this example, two genomic variant data sources are joined with a patient's clinical data. If sequence data are directly accessible via an EHR system (a rarity today), the SMART on FHIR Genomics Adaptor would be implemented against that EHR system.



The SMART on FHIR Genomics specification demonstrates how it is possible to chip away at "the omics challenge." Our prototype apps demonstrate how a wide range of clinico-genomics use cases can be covered by our API. Moreover, the developer learning curve for SMART on FHIR Genomics proved very shallow: inexperience in working with genomic data could be overcome during a connectathon, suggesting that concerns about complexity can easily be addressed. Finally, we tested the feasibility of building SMART on FHIR Genomics data source adaptors by prototyping adaptors for Illumina, 23andMe, and the Vanderbilt Research Derivative.[14] Our preliminary efforts should encourage vendors that adaptor creation will not require a significant software engineering endeavor. Synergies are possible using enterprise clinical rules service (ECRS) for scalable clinical decision support.[15] In addition, new standards for extracting phenotypic information from unstructured EHRs [16] may enable linking genetic information to contexts outside of structured data. In short, we believe that the current specification of SMART on FHIR Genomics would hit the trifecta of sequencing data vendors, EHR vendors, and clinical app developers.

BRIEF COMMUNICATION

Figure 3: This figure depicts SMART on FHIR Genomics apps (projects.iq.harvard.edu/smartgenomics). The DB EMR (Diabetes Electronic Medical Record) app (A) integrates clinical, genomic, and sensor information. The Genomics Advisor app (B) presents estimates of disease risk based on germline Single Nucleotide Polymorphisms (SNPs) and operates alone or embedded in another app (as shown here). The Precision Cancer Medicine (PCM) app (C) integrates cancer genomic and clinical information to compare a patient to others with similar phenotypes and genotypes. PCM is a point-of-care mobile app that integrates a patient's clinical and somatic mutation data and is intended to aid physician-patient communication.



The FHIR standard was designed to create resource definitions for the most common medical scenarios. As of early 2015, there are 99 defined FHIR resources. It is our belief that genomics data are sufficiently distinct from existing FHIR resources to warrant a new resource and two extensions to the existing FHIR standard. While extensions can be created ad infinitum, a resource would have to be added to the central FHIR specification, which has a planned cap of 150 resources (personal communication, Lloyd McKenzie). Given that personalized medicine is rapidly becoming necessary for improved medical care, and given that patient sequencing data are already produced routinely by sequencing vendors, there is a compelling case for the HL7 community to engage, refine, and incorporate SMART on FHIR Genomics into FHIR. The HL7 Clinical Genomics Workgroup (of which two authors are co-chairs) recently voted to take portions of SMART on FHIR Genomics forward by incorporating those parts into a clinical genomics observation profile for FHIR DSTU2, which is currently in an HL7 ballot (see Table 1).

Finally, the timing of advancing the results of this effort could not be better. As of this writing, FHIR is getting industry attention and input as it moves from DSTU1 to balloting for DSTU2. A broad coalition of industry and academic partners have recently formed the Health Services Platform Consortium and the Argonaut Project to advance, among other things, the SMART on FHIR framework for traditional clinical data and interoperability. CMS and the Office of the National Coordinator have continued to make interoperability a key focus of policy and funding. Finally, President Barack Obama announced the Precision Medicine Initiative during the 2015 State of the Union.[17] As the requirements for this initiative become clearer, we would expect that standard data frameworks for genetic variant data would come to the forefront as important tools for meeting those requirements. In this way, SMART on FHIR Genomics could be very beneficial.

CONCLUSION

SMART on FHIR Genomics describes a set of genomic data variant resource definitions and profile extensions compatible with FHIR. Designed to be part of the SMART on FHIR framework, the SMART on FHIR Genomics specification provides developers with a unified framework to tap multiple sources of genomic and EHR clinical data and

create the type of apps that will be needed for precision medicine. Our genomics data specification has been tested and found to be capable of supporting high functioning clinico-genomics apps, and the feasibility of our endeavor has been verified by the creation of prototype adaptors with common genomics data sources. In short, we tested a way to successfully provide standardized data access to genomic data. Moreover, we made that means of standardized access part of a larger, unifying approach to EHR clinical data and app development.

Combining SMART on FHIR Genomics and SMART on FHIR should yield impressive benefits for real-world stakeholders. DNA sequencing systems can deliver more value by offering a standard way to integrate gene and variant data with EHR system data (and vice versa).[18] Standardization ensures that clinical app developers can build and deploy powerful medical apps better, faster, and cheaper. Finally, providers can hope to obtain workflow-integrated, genomics-enriched CDS for their clinical staff that will help usher in the era of precision medical care.

CONTRIBUTORS

GA, PZ, JW, TC, MU, DK, and IK contributed to the design of the study and writing of the paper. GA, PZ, TC, and DK contributed to implementation of study.

FUNDING

This work was supported by the Strategic Health IT Advanced Research Projects Award 90TR000101 from the Office of the National Coordinator of Health Information Technology.

COMPETING INTERESTS

None.

ACKNOWLEDGEMENTS

We would like to thank Rachel Ramoni, Joshua Mandel, and Kenneth Mandl for their contributions and discussions on this work. We would like to thank Xin Yang and Yao Chen (undergraduates at Southern University of Science and Technology China), Ziou Zheng (undergraduate from Xi'an Jiaotong University), as well as Ross Oreto and Daniel Carbone of Vanderbilt University, who programmed the various apps.

REFERENCES

1. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31–46.
2. Mandl KD, Kohane IS. No small change for the health information economy. *N Engl J Med.* 2009;360(13):1278–81.
3. Mandl KD, Mandel JC, Murphy SN, *et al.* The SMART Platform: early experience enabling substitutable applications for electronic health records. *J Am Med Inform Assoc.* 2012;19(4):597–603.
4. Field D, Sansone SA, Collis A, *et al.* Megascience. 'Omics data sharing. *Science.* 2009;326(5950):234–6.
5. Wooley JC, Field D, Glockner FO. Extending standards for genomics and metagenomics data: a Research Coordination Network for the Genomic Standards Consortium (RCN4GSC). *Stand Genomic Sci.* 2009;1(1):87–90.
6. Tenenbaum JD, Sansone SA, Haendel M. A sea of standards for omics data: sink or swim? *J Am Med Inform Assoc. 2014*;21(2):200–3.
7. Raths D. Trend: standards development. Catching FHIR. A new HL7 draft standard may boost web services development in healthcare. *Healthcare Informatics.* 2014;31(2):13,6.
8. den Dunnen JT, Antonarakis SE. Nomenclature for the description of human sequence variations. *Hum Genet.* 2001;109(1):121–4.
9. Pruitt KD, Harrow J, Harte RA, *et al.* The Consensus Coding Sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* 2009;19(7):1316–23.
10. Povey S, Lovering R, Bruford E, Wright M, Lush M, Wain H. The HUGO Gene Nomenclature Committee (HGNC). *Hum Genet.* 2001;109(6):678–80.
11. Bender D, Sartipi K. HL7 FHIR: An agile and RESTful approach to healthcare information exchange. In: *Proceedings of the IEEE 26th International Symposium on Computer-Based Medical Systems (CBMS 2013).* 2013:326–31.
12. Procedure Resource, HL7 FHIR DSTU1 v0082 2015. Health Level 7 International website. http://www.hl7.org/implement/standards/fhir/procedure.html. Accessed February 8, 2015.
13. Observation Resource, HL7 FHIR DSTU1 v0082 2015. Health Level 7 International website. http://www.hl7.org/implement/standards/fhir/observation.html. Accessed February 8, 2015.
14. Danciu I, Cowan JD, Basford M, *et al.* Secondary use of clinical data: The Vanderbilt approach. *J Biomed Inform.* 2014;52:28–35.
15. Goldberg HS, Paterno MD, Rocha BH, *et al.* A highly scalable, interoperable clinical decision support service. *J Am Med Inform Assoc.* 2014;21(e1):e55–62.
16. Pathak J, Bailey KR, Beebe CE, *et al.* Normalization and standardization of electronic health records for high-throughput phenotyping: the SHARPn consortium. *J Am Med Inform Assoc.* 2013;20(e2):e341–8.
17. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med.* 2015.
18. Rioth M, Staggs D, Warner J. Incorporation of externally generated next-generation tumor genotyping into clinical and research workflows: successes and lessons learned. *J Clin Oncol.* 2014;32 suppl 30:abstr 156.

AUTHOR AFFILIATIONS

[1]Children's Hospital Informatics Program, Boston, MA

[2]Center for Biomedical Informatics, Harvard Medical School, Boston, MA

[3]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA

[4]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN

[5]Department of Medicine, Division of Hematology/Oncology, Vanderbilt University, Nashville, TN

[6]Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA

[7]Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL

[8]Dana-Farber Cancer Institute, Boston, MA

*Contributed equally to this work

BRIEF COMMUNICATION