The Office of the National Coordinator for Health Information Technology

Sync for Genes

ENABLING CLINICAL GENOMICS FOR PRECISION MEDICINE VIA HL7® FAST HEALTHCARE INTEROPERABILITY RESOURCES®

Gil Alterovitz[1,*], Jonathan Brown[2], Menzies Chan[3], Jeff Daniels[4], Erich Haberman[4], Elaine Johanson[5], Andrew LeBeau[3], Christine Lo[2], Martin Maiers[6], Bob Milius[6], Matthew Tiller[7], Jeremy Warner[8], Grant Wood[9], Alan Zhu[10]

[1] Harvard Medical School/Boston Children's Hospital and Massachusetts Institute of Technology. [2] Counsyl. [3] Illumina. [4] Foundation Medicine, Inc. [5] Food and Drug Administration. [6] National Bone Marrow Program. [7] ESAC. [8] Vanderbilt University Medical Center. [9] Intermountain. [10] PrecisionFDA/DNA Nexus.

November 2017

TABLE OF CONTENTS

Introduction **5**

Precision Medicine Initiative and the *All of Us* Research Program **5**

Background **6**

Precision Medicine Task Force **6**

Accelerating FHIR Genomic Standards to Support Clinical Care and Research, Including PMI **7**

Large-scale Data Storage and Retrieval 9

Approach **10**

Sync for Genes **10**

Genomics APIs 13

FHIR and Sync for Genes **13**

FHIR Genomics **14**

HL7 Domain Analysis Model: Clinical Sequencing **16**

Pilots and Sync for Genes Participation **16**

Family Health History Genetics Use Case: Counsyl/Intermountain Healthcare Sync for Genes Pilot **18**

Use Case Description 18

Pilot Description 18

Sync for Genes Pilot Work 19

Pilot Results 19

Pilot Next steps 20

NGS Sequencing Solutions Use Case: Illumina Sync for Genes Pilot **20**

Use Case Description 20

Pilot Description 20

Sync for Genes Pilot Work 21

Pilot Results 21

Pilot Next Steps 22

Sequencing Quality and Regulatory Genomics Use Case: FDA Sync for Genes Pilot **22**

Use Case Description 22

Pilot Description 23

Sync for Genes Pilot Work 23

Pilot Results 23
Pilot Next Steps 25
Somatic/Tumor Testing Use Case: Foundation Medicine, Inc. and Vanderbilt University Medical Center Sync for Genes Pilot 25
Use Case Description 25
Pilot Description 26
Sync for Genes Pilot Work 26
Pilot Results 27
Pilot Next Steps 27
Tissue Matching Use Case: National Marrow Donor Program/Be the Match Sync for Genes Pilot 28
Use Case Description 28
Pilot Description 28
Sync for Genes Pilot Work 29
Pilot Results 29
Pilot Next Steps 29
General FHIR Recommendations 29
FHIR Resource IDs 29
FHIR Documentation 30
Codeable Concepts 30
Sequence resource 30
Observation for Genetics 30
DiagnosticReport Genetics Profile 31
FamilyMemberHistory Genetics Profile 32
FHIR Searchable Parameters 32
DAM Use Case Development 33
Discussion and Conclusion 34
General discussion 34
General pilot feedback 34
Themes Explored During Pilot 35
FHIR Search 35
Value Sets 35
FHIR Server Functionality 35

Future work **36**

Introduction

The *All of Us* Research Program (*All of Us*),[1] part of the Precision Medicine Initiative that was launched in January 2015, aims to gather health data from one million or more people in the United States and make those data available to researchers with the goal of accelerating the pace of research and improving health. Sync for Science[2] (S4S) and Sync for Genes[3] are two technology-based efforts that collaborate with government and industry to help deliver on goals of *All of Us*. Both use Health Level Seven International (HL7®)'s Fast Healthcare Interoperability Resources (FHIR®) specification[4] as the basis for data exchange. S4S's goal is to standardize an easier way for patients to share medical records with researchers by building on the FHIR Argonaut Project.[5] Sync for Genes' goal is to standardize the sharing of genomic information by developing and building on FHIR Genomics[6] and associated technologies. Sync for Genes enables sharing of genomic information in a consistent and usable manner. Sync for Genes is the first step toward integrating clinical genomics into point-of-care and expanding from S4S's core clinical data to genomics by leveraging FHIR Genomics. This report summarizes the background and rationale for Sync for Genes, describes the five pilot use cases, and reports on the results and possible next steps.

Precision Medicine Initiative and the *All of Us* Research Program

The *All of Us* Research Program (formerly known as the PMI Cohort Program), run by the National Institutes of Health (NIH), is the cornerstone of the PMI. *All of Us* will help achieve the PMI's goal — better understanding the complexity of disease, disease prevention, and the effectiveness of treatment at the individual level — by creating a longitudinal national research cohort of health data from at least one million or more U.S. volunteers. The volunteers will complete online surveys and be asked to provide access to data in their health care providers' EHR systems, which the program will gather in coordination with a wide range of health systems and health information technology (health IT) developers.

Many *All of Us* volunteers will provide biological samples (i.e., specimens), which will be stored for use at a later date. Examples of utilization of DNA extracted from the specimens could potentially include targeted Single Nucleotide Polymorphisms (SNP) analysis, genome-wide SNP analysis to determine ancestry and to facilitate Genome-Wide Association Studies (GWAS), whole-exome sequencing (WES), and whole-genome sequencing (WGS). In addition to facilitating such information collection, Sync for Genes technologies can enable capture of existing and future clinical-grade genomic data that can complement such data. This is especially true for tumor analysis, but holds true for any clinical genomic test — including direct-to-consumer activities.

---

[1] http://allofus.nih.gov/
[2] http://syncfor.science/
[3] http://www.sync4genes.org/
[4] http://hl7.org/fhir/
[5] http://argonautwiki.hl7.org/
[6] https://hl7.org/fhir/STU3/genomics.html

All of Us will build a comprehensive research data set and make it broadly accessible available to researchers of all types, including citizen scientists, in order to fulfill the PMI's goal. The data set will contain biological, environmental, and lifestyle data in order to enable a variety of scientific opportunities, including to[7]:

- Develop ways to measure risk for a range of diseases based on environmental exposures, genetic factors and interactions between the two;
- Identify the causes of individual differences in response to commonly used drugs (commonly referred to as pharmacogenomics);
- Discover biological markers that signal increased or decreased risk of developing common diseases;
- Use mobile health (mHealth) technologies to correlate activity, physiological measures and environmental exposures with health outcomes;
- Develop new disease classifications and relationships;
- Empower study participants with data and information to improve their own health; and
- Create a platform to enable trials of targeted therapies

Sync for Genes seeks to enable clinical genomics for precision medicine via FHIR. This vision is being facilitated by the design, build, test, and piloting of genomic implementations based upon the balloted Release 3 of FHIR.[8]

Background

Precision Medicine Task Force

The Office of the National Coordinator for Health Information Technology (ONC) convened the Precision Medicine Task Force[9] in 2015 under the Health IT Standards Committee (HITSC) in order to advance data standards, address relevant privacy policies, and advance innovations in health IT that support precision medicine. The task force was charged with (following list quoted directly from[10]):

- Identifying opportunities for innovative collaboration around pilots and testing of standards that support health IT interoperability for precision medicine;
- Recommending existing standards that are currently ready to support the PMI;
- Identifying emerging standards and reference implementations that may require further pilot testing in order to support the PMI; and
- Identifying gaps in available data standards related to the PMI.

[7] https://allofus.nih.gov/about/scientific-opportunities
[8] http://hl7.org/fhir/STU3/index.html
[9] https://www.healthit.gov/FACAS/health-it-standards-committee/hitsc-workgroups/precision-medicine-task-force
[10] https://www.healthit.gov/sites/faca/files/PMTF_Transmittal_Letter_2015-09-25_v2.pdf

On September 22, 2015, the Task Force presented its recommendations regarding standards to support the PMI. The HITSC approved the final recommendations, which were, specifically (following list quoted directly from[11]):

- Readily applicable standards for the PMI;
- Promising standards for PMI;
- Standards gaps for PMI; and
- Accelerators to advance standards

The recommendations identified gaps and current efforts in order to bolster the PMI's goals with regards to genomic data. The task force recommended[12] that information about precision medicine should be added to the *Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap* (Interoperability Roadmap). [13]

Sync for Genes addresses a selection of the proposed items raised by the Task Force. Sync for Genes supports the ONC, NIH, and the Food and Drug Administration (FDA) toward fulfilling the Task Force's recommendation to "[i]dentify opportunities for ONC to support our federal partners' PMI efforts and related health IT/interoperability challenges, including National Cancer Institute, Food and Drug Administration, National Institutes of Health, and Department of Veterans Affairs." Sync for Genes seeks to fulfill another recommendation — to "[i]dentify opportunities for ONC to collaborate with industry and pilot the use of standards to enable data donation and patient access through application programming interfaces (APIs) using standards such as FHIR and OAuth 2.0" — by utilizing FHIR Genomics and its leveraging of the SMART on FHIR Genomics[14] framework as well as using the underlying security/privacy framework common to SMART on FHIR[15].

Accelerating FHIR Genomic Standards to Support Clinical Care and Research, Including the PMI

Sync for Genes builds upon the S4S pilot program and uses the HL7 FHIR Genomics standard for reporting clinical genetic and genomic observations, developed by the HL7 Clinical Genomics Work Group (CGWG). Genetic and genomic data, including data from next generation sequencing (NGS) laboratories, should be shared in a standardized way to support interoperability. In fact, the first funding opportunity announcement (FOA) for the healthcare provider organization enrollment centers[16] under the *All of Us* Research Program discussed need for standards, citing work discussed here —

---

[11] https://www.healthit.gov/FACAS/sites/faca/files/HITSC_Precision_Medicine_2015-12-10.pdf
[12] http://www.healthcareitnews.com/news/onc-task-force-recommends-adding-precision-medicine-initiative-interoperability-roadmap
[13] https://www.healthit.gov/policy-researchers-implementers/interoperability
[14] Alterovitz G, Warner J, Zhang P, Chen Y, Ullman-Cullere M, et al. SMART on FHIR Genomics: Facilitating standardized clinico-genomic apps. J Am Med Inform Assoc 2015;0:1–6. doi: 10.1093/jamia/ocv045
[15] Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association. Available online 17 February 2016.
[16] https://grants.nih.gov/grants/guide/rfa-files/RFA-PM-16-002.html

including both SMART on FHIR Genomics and the HL7 Domain Analysis Model (DAM) for clinical sequencing.

ONC seeks to expedite and support the use of standards identified in the Task Force recommendations, such as HL7 FHIR, to enable and improve patient's ability to seamlessly share their genomic information. To that end, Sync for Genes was designed to enable sharing and interpretation of genomic information by developing and testing enhancements to the CGWG suite of standards.

Figure 1 below depicts a national standard based on FHIR for clinical care and research. All data exchange among the entities is enabled by FHIR, with two labeled as Substitutable Medical Apps, Reusable Technology "(SMART) on FHIR" because they can utilize that framework for mobile applications (apps). The large arrows represent data exchange between two entities; while small arrows in the text (e.g., "Clinical Data → Argonaut") represent that the type of data being exchanged is under the purview of the named project.

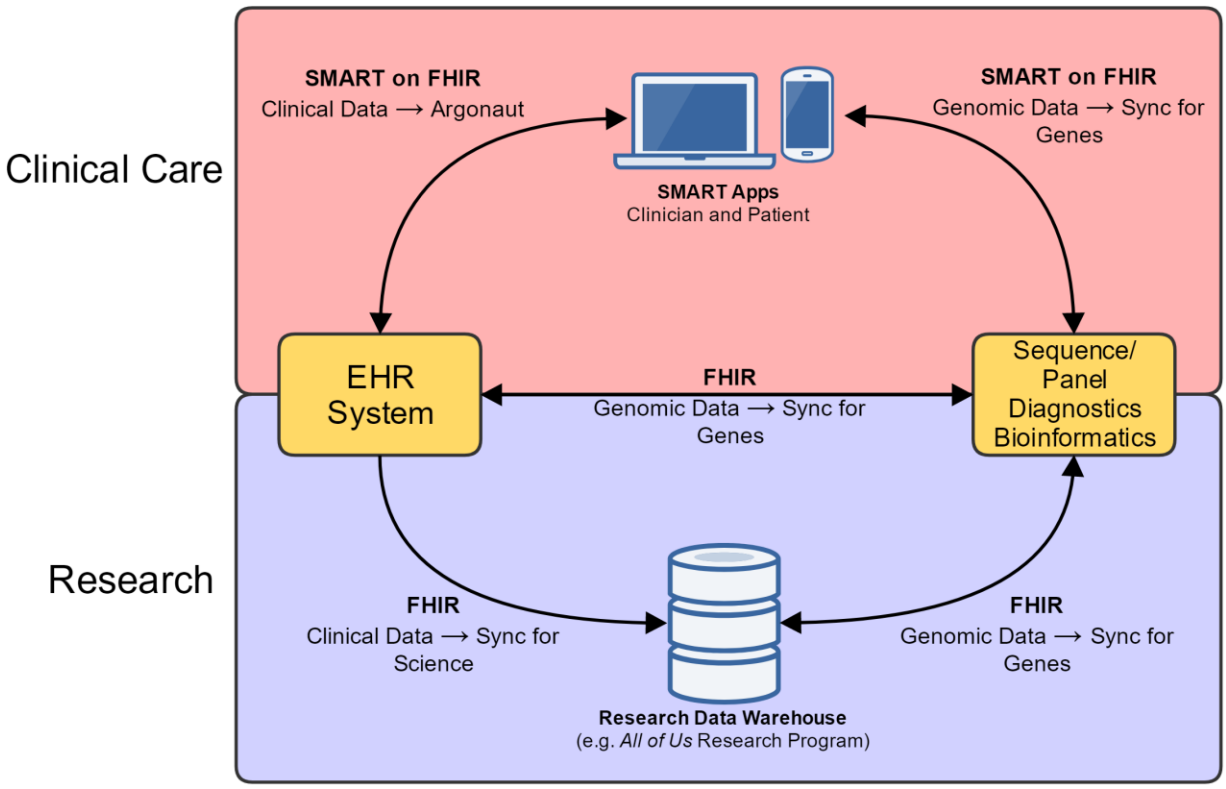

Figure 1: Enabling a single national standard for clinical care and research via FHIR

Both the Interoperability Roadmap[17] and the 2015 Edition Certification Criteria[18] (2015 Edition) encouraged a robust, application programming interface (API)-based, and interoperable health care

[17] https://www.healthit.gov/sites/default/files/hie-interoperability/nationwide-interoperability-roadmap-final-version-1.0.pdf

[18] https://www.healthit.gov/policy-researchers-implementers/2015-edition-final-rule

ecosystem. Sync for Genes aims to advance that ecosystem as well as the efforts of the PMI by creating a foundation for widespread use of genomic data to be utilized not only for research, but also at the point-of-care, leading to informed decision-making.

Large-Scale Data Storage and Retrieval

An existing precedent for networked large-scale data storage and retrieval, shown in Figure 2, involves medical images stored on a Picture Archive and Computer/Communication System (PACS).[19,20] Medical image files, such as DICOM images, are large and are not easily stored directly on an EHR system. Typically, a PACS is used to store image files that are then referenced by the EHR. The EHR holds metadata about the image, but not the image itself. A similar system could be used for the large amounts of data generated by NGS. A Genomic Archive and Computer/Communication System, as shown in Figure 3, stores sequence data generated from a sequencing laboratory. A GACS acts in the same manner as a PACS — to store big data that is not suitable to store directly in an EHR.

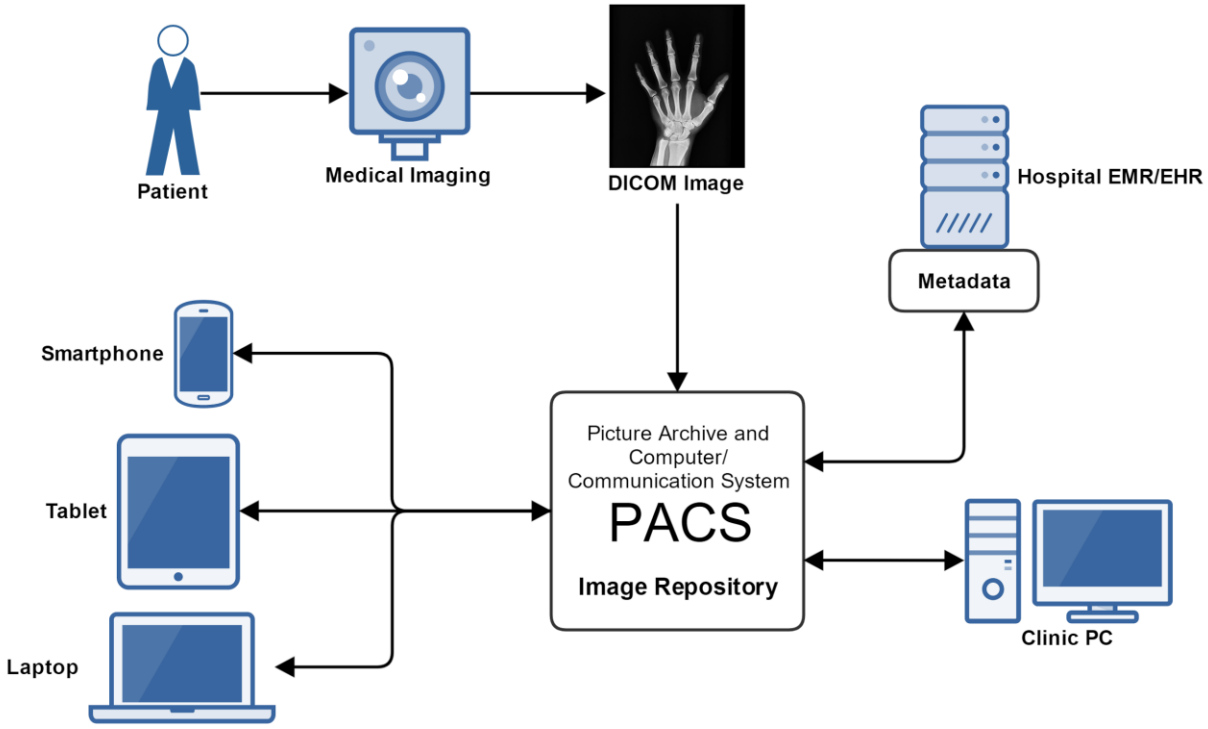

Figure 2: Picture Archive and Computer/Communication System

[19] Huang, H. K. (2004). PACS and Imaging Informatics: Basic Principles and Applications. New Jersey: John Wiley & Sons.

[20] Smith, W.L. Radiology 101: The Basics & Fundamentals of Imaging. Wolters Kluwer Health, 2013 p. 17.

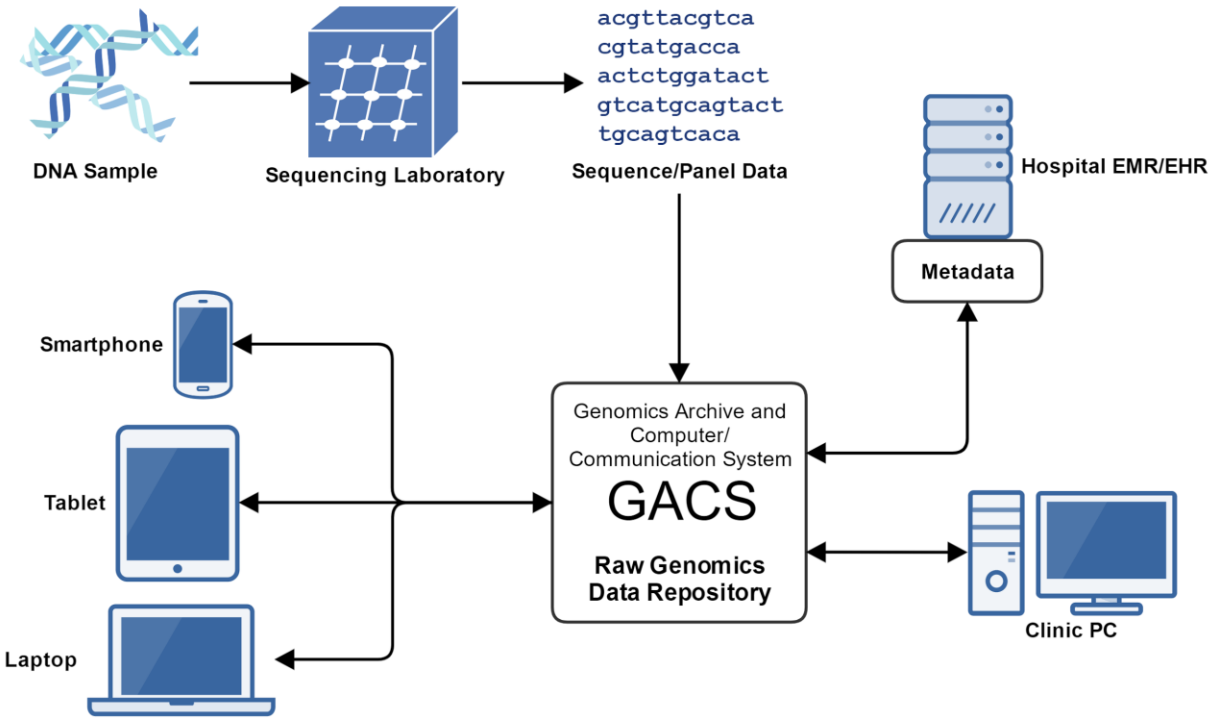

Figure 3: Genomics Archive and Computer/Communication System

Approach

Sync for Genes

It is critical that genomic data, including data from NGS laboratories, be shared in a consistent and usable way. Sync for Genes was established to expedite the use of standards to enable and improve patients' ability to seamlessly share their genomics information.

Developing a standard API for clinical genomics necessitates a set of underlying guidelines for its development. These help ensure that various stakeholder needs are considered and that the standard is adopted. Clinical genomics presents specific needs due to its need for integration with existing clinical systems, government-associated regulations, diverse stakeholders, evolving nature, big data potential, and various use cases. Below is the list of the seven developed guidelines on standard development for enabling clinical genomics at point of care and facilitating precision medicine. The standard development should:

1. Enable a layered-based adoption model for standard specification implementation.

Traditional laboratories moving into genetics have focused on communicating simple categorical laboratory results and interpretative information. With time, some of these laboratories may need to provide more provenance and underlying raw data for potential re-interpretation. On the other hand, NGS solutions have focused on the raw, underlying data — and increasingly want to add interpretive information for clinical use cases. The standard should take both of

these considerations into account by enabling a layered-based approach for adoption of the standard so that some immediate tasks can be accomplished with minimal adoption and invested work is not lost as more of the standard is adopted. In addition, the standard should ideally use a layered-based methodology that can be approached from multiple sides (e.g., raw or interpretive sides).

2. Enable a layered-based adoption model for platform investment.

Different organizations will approach storing clinical genomic information differently. With simple genetic tests, results can be stored in EHRs. As the next layer, with NGS-based tests, a GACS (Genomics Archive Computer/Communication System) has been recommended as a solution with cache of interpretive information in the EHR: "Advances in clinical genomics would lead to a new category of computer system. Much like a PACS was needed to store large amounts of imaging data; health systems will need a GACS. A GACS integrates with the EHR, yet stores the data."[21]

3. Be developed in alignment with the American National Standards (ANS) process for accreditation.

The ANS process includes "essential requirements for due process" and benchmarks on: openness, lack of dominance, balance, coordination and harmonization, notification of standards development and coordination, consideration of views and objections, evidence of consensus and consensus body vote, and appeals.[22]

4. Facilitate 2015 Edition Certification Criteria-based principles, including enabling patient access to their information via APIs.

The ONC-developed criteria for certifying health IT requiring API access via patient applications (apps) to a Common Clinical Data Set (§170.316(g)(7), (g)(8) and (g)(9), 2015 Edition of Health IT Certification Criteria). The Medicare and Medicaid Electronic Health Record Incentive Programs relates to the 2015 Edition and includes API criteria for its objectives in "Patient Electronic Access to Health Information and "Coordination of Care Through Patient Engagement."[23,24]

5. Promote clinical laboratory interoperability.

The American Medical Informatics Association (AMIA) has been "developing criteria meant to describe the qualities of desirable standards within the context of health IT interoperability.

---

[21] https://www.healthcare-informatics.com/blogs/david-raths/fhir-s-key-role-precision-medicine-initiative

[22] ANSI Essential Requirements: Due process requirements for American National Standards, Jan 2017

[23] https://www.federalregister.gov/documents/2016/11/14/2016-26515/medicare-program-hospital-outpatient-prospective-payment-and-ambulatory-surgical-center-payment

[24] https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/Downloads/OPPSOverview_Stage2.pdf

These criteria are still in draft […]".[25,26] An approach for promoting laboratory interoperability based on such principles was presented on behalf of AMIA at the multi-agency workshop sponsored by the Centers for Disease Control and Prevention (CDC), the FDA, the National Library of Medicine of the National Institutes of Health (NLM), ONC, and the Centers for Medicare & Medicaid Services (CMS).[27] Among others, this includes: 1) "the specification of standards using conventional tools and methods of the present web-economy, that can leverage prevalent information technology and communications infrastructure;" 2) "[health] IT standards should support human readability, simplicity, [and] parsimony;" 3) "The evolution of [health] IT standards that are modular and substitutable, having clear boundaries for use;" 4) "The incorporation of the FAIR data principles (findable, accessible, interoperable and reusable) to optimize the use of resources and data;" and 5) "Technology-enabled approaches that encourage patients to review and contribute directly to their record."

6. Facilitate implementations that follow Clinical Laboratory Improvement Amendments (CLIA) and Health Insurance Portability and Accountability Act (HIPAA) regulations and guidelines.

While a standard itself would not likely enforce these regulations directly — use of constraints, documentation, and examples may help facilitate workflows that enable laboratories to more easily implement these via standard. The CLIA of 1988 regulates clinical laboratory testing using human specimens. These apply to clinical genomic laboratories and include regulations on communication of such information, its display, and retention. Several of these are covered in: 42 CFR 493.1291 and 42 CFR 493.1241. The "CLIA Program and HIPAA Privacy Rule; Patients' Access to Test Reports" includes requirements on patient access to their laboratory data and privacy (79 FR 7289).

7. Facilitate data sharing as per GA4GH framework: "Framework for Responsible Sharing of Genomic and Health-Related Data"

This framework[28] focuses on data sharing and is built on a set of foundational principles for responsible sharing of genomic and health-related data: 1) respect individuals, families, and communities; 2) advance research and scientific knowledge; 3) promote health, well-being, and the fair distribution of benefits; and 4) foster trust, integrity, and reciprocity. The core elements of responsible data sharing presented are: 1) transparency; 2) accountability; 3) engagement; 4) data quality and security; 5) privacy, data protection, and confidentiality; 6) risk-benefit analysis; 7) recognition and attribution; 8) sustainability; 9) education and training; as well as 10) accessibility and dissemination.

---

[25] https://www.amia.org/news-and-publications/press-release/amia-supports-multi-agency-effort-improve-lab-data-interoperability

[26] https://www.amia.org/news-and-publications/press-release/amia-unveils-2017-policy-priorities-defines-pillars-new-policy

[27] https://www.amia.org/sites/default/files/Perspectives-on-Lab-Interoperability-Alterovitz-11082016.pdf

[28] https://genomicsandhealth.org/files/public/Framework for Responsible Sharing of Genomic and Health-Related Data - Version 10 September 2014.pdf

Genomics Application Programming Interfaces (APIs)

A number of APIs have been recently independently reviewed in the published paper "A Review on Genomics APIs."[29] The paper reviews three genomic API's: a) Google Genomics; b) SMART on FHIR Genomics; and c) 23andMe. The paper presents a comparative view of the APIs against the following eight criteria: 1) Input data to API (clinical and/or genomic features); 2) Location of data (vendor and/or other locations); 3) API response (JSON and/or XML); 4) Ability to import data; 5) Range search for variants in a given individual; 6) Identify risk for a disease in an individual; 7) Availability of reference applications using the API, and 8) Authentication. The Google Genomics API is based on **Global Alliance for Genomics and Health** (GA4GH). The specification of the SMART on FHIR Genomics evaluated in this review paper is a precursor to FHIR Genomics from Release 3 of FHIR. Based on Table 1 of the paper, the Google API supported four criteria fully, three partially, and one was not supported. The SMART on FHIR Genomics API supported all eight criteria. The 23andme API was found to support four criteria fully, two partially, and two were not supported.

FHIR and Sync for Genes

FHIR is being developed by the HL7, an American National Standards Institute (ANSI)-accredited software development organization, and is designed to satisfy the criteria under the 2015 Edition. It also enables incremental adoption, described below. In this project, FHIR Genomics from Release 3 was piloted and evaluated as a national standard for clinical genomics information exchange. FHIR Genomics was designed to meet current clinical genomics use cases utilized at health care organizations today, with special focus on clinical sequencing.[30] It integrates clinical genomic information, resulting in better patient care via point of care applications, patient facing applications, and clinical research as well as analysis.[31,32]

Sync for Genes complements S4S's efforts in sharing 2015 Edition datasets from EHRs by enabling the sharing of clinical genomic information for research programs such as *All of Us*. As the first step toward integrating clinical genomics into the point-of-care, Sync for Genes leverages the HL7 FHIR standard to communicate information from clinical genomic laboratories in a format for universal use across medicine.

As shown in Figure 4, Sync for Genes has combined participation from laboratories, providers, government, developers, patients, and coordinators. The information from this project will be used to improve patient care by piloting and further developing the FHIR Genomics standard.

---

[29] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4669666/

[30] Alterovitz, Ullman-Cullere, et. al. "Domain Analysis Model: Clinical Sequencing," HL7, 2017.

[31] Alterovitz, Warner, et. al. "SMART on FHIR Genomics: facilitating standardized clinico-genomic apps", JAMIA, 2016.

[32] Rimer, Harper, et. al., Improving Cancer-Related Outcomes with Connected Health: A Report to the President of the United States from the President's Cancer Panel, 2016.

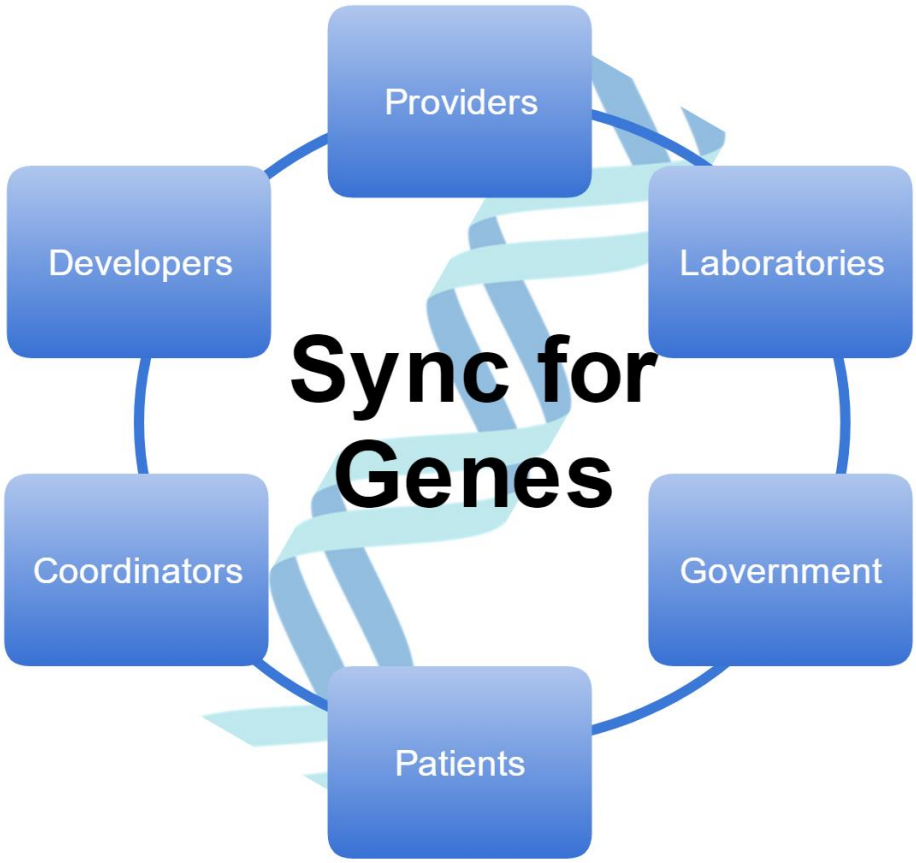

Figure 4: Stakeholders for Sync for Genes

The Sync for Genes pilot process to date has involved several steps. First, Sync for Genes solicited applications for pilots. During this time, and in coordination with stakeholders, FHIR Genomics for the Release 3 of FHIR was finalized. Sync for Genes selected five national pilots based on use cases in the CGWG's Domain Analysis Model: Clinical Sequencing[33] (DAM; discussed in section 3.3) produced by the CGWG. Pilots created/tested FHIR Genomics implementations for conformance and throughout the pilot period, provided feedback on specific issues seen. The information collected throughout the process was then disseminated and used to influence the standard. Feedback on suggested changes was presented at the Sprint 2017 HL7 Working Group Meeting (WGM) meeting (May 6-12, 2017; Madrid, Spain) and in this report, produced as "open source" via Creative Commons Attribution.[34]

FHIR Genomics

FHIR Genomics serves as basis for the specification for the Sync for Genes data format. FHIR Genomics-associated resources and profiles include:

---

[33] http://www.hl7.org/implement/standards/product_brief.cfm?product_id=446

34 http://creativecommons.org/licenses/by/4.0/

- Sequence resource — describes NGS data primarily by identifying sequence variants, but can also contain raw sequence data
- Observation-genetics profile — interprets variants from sequence resource
- DiagnosticReport-genetics profile — describes a genetic test report, referencing test details and genetic observations
- ProcedureRequest-genetics profile — orders a test to detect sequence variants
- FamilyMemberHistory-genetics profile — pedigree representation and related genetic observations are referenced here.
- Human Leukocyte Antigen (HLA)-genotyping-results profile — utilizes Minimum Information for Reporting Next-generation sequence Genotyping (MIRING)[35,36] principles to aid immunogenetics and histocompatibility research.

FHIR Genomics is designed so that traditional laboratories can start at a low level of raw information/high level of abstraction, as represented in Figure 5 at the top of the pyramid, and adopt, slowly going down the pyramid, to add in more raw underlying information. At the same time, it is designed so that NGS laboratories can move up from the bottom of the pyramid's raw reads, slowly moving to the top, adding in more interpretative/abstract information at each step. In level 1, laboratory codes return if a result is positive or negative, without its underlying provenance. Additional information and provenance is needed for many use cases where variant, gene name, and other information (as encoded in the Observation Genetics profile) may be needed to give more information on the underlying genetic test. This is captured in level 2. Adding other profiles to capture relationships for level 3, such as family history, laboratory orders, diagnostic reports with set of observations, is done with the additional profiles of FHIR Genomics. Level 4 involves adding sequence information and its associated quality metrics. Finally, for additional provenance, access to repositories (e.g., for raw reads) is provided via pointers to external repositories and platforms (e.g., GA4GH) in level 5.

A particular organization's initial use case can be simple to implement regardless of its initial placement on the pyramid. For some NGS laboratories, for example, it may make sense to start with raw reads (e.g., level 5) and slowly add more interpretive information (moving from level 5 toward level 1) as the laboratory moves toward more clinical/diagnostic uses. Complexity in the use case arises when all levels of abstraction and results are synthesized into interrelating FHIR resources and profiles. FHIR Genomics was designed specifically to allow for this range of complexity.

In Figure 5 the right half of the image, to the right of the pyramid, shows the FHIR Genomics representations and their relationships as they relate to the pyramid. The right-hand arrows indicate that the levels can incrementally be adopted going up or down in order to convey the full range of FHIR Genomics complexity.

[35] https://www.ncbi.nlm.nih.gov/pubmed/26319908
[36] http://igdawg.org/miring.html

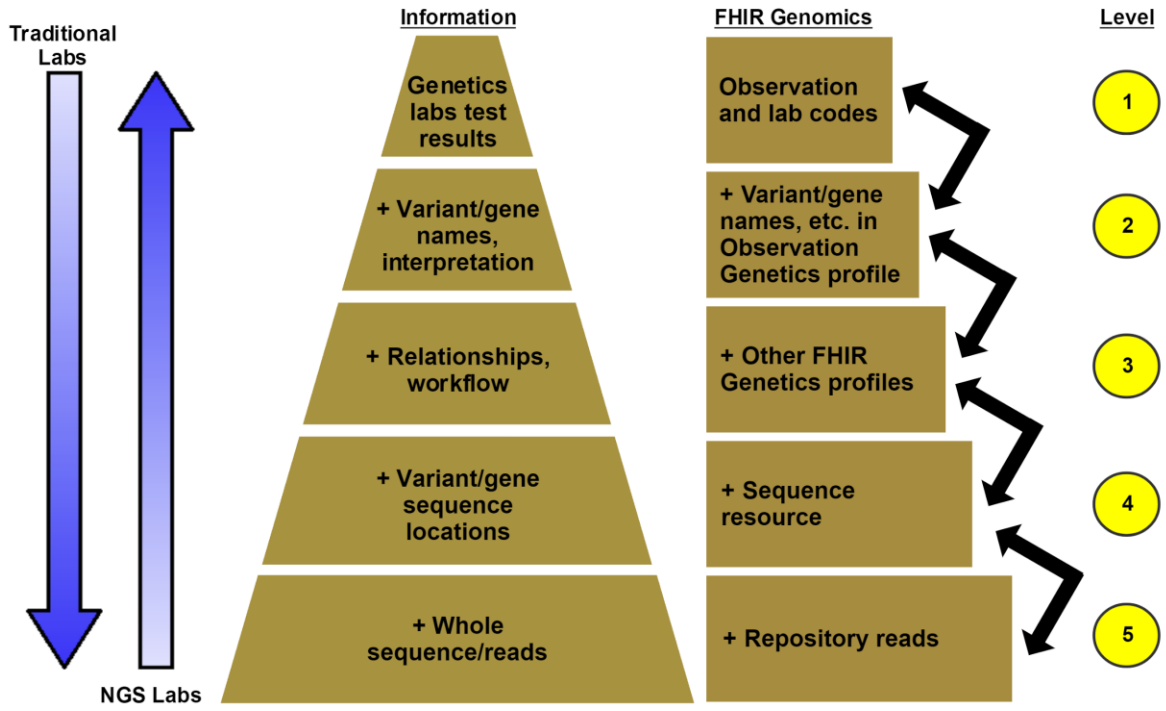

Figure 5: Incremental adoption model of Sync for Genes

HL7 Domain Analysis Model: Clinical Sequencing

In February of 2017 the HL7 Domain Analysis Model: Clinical Sequencing was published[37,38] by the CGWG as a means to establish and disseminate the clinical genomics use cases. The DAM was intended to be utilized as the basis for the development of FHIR Genomics and, together with the latter, for the development and execution of Sync for Genes. The DAM outlines multiple scenarios centered on sequencing in clinical and translational settings to which FHIR Genomics can be applied. These scenarios serve as a basis for Sync for Genes use cases, described below.

Pilots and Sync for Genes Participation

This section describes the pilots individually, listed in alphabetical order by use case. Each subsection below includes the pilot's: use case, description, work conducted for this project, specific results, and next steps. The names of the subsections below are composed of the use case name followed by the pilot name. Overall results and next steps for Sync for Genes are covered in a later section.

[37] https://www.healthcare-informatics.com/news-item/interoperability/hl7-publishes-domain-analysis-model-clinical-sequencing

[38] http://www.hl7.org/documentcenter/public_temp_177978DE-1C23-BA17-0C54664CAE99ACD3/pressreleases/HL7_PRESS_20170410.pdf

The Sync for Genes pilot phase use cases were chosen to address the core scenarios from the DAM and to fully exercise its FHIR-based representation of genetic analyses and data. Naturally, representing the sequence data itself in a standard way, as in the "NGS Sequencing Solutions" use case below, is of vital importance to many scenarios. Sync for Genes also standardizes the metadata that gives context to the sequence, as seen in the "Family Health History Genetics" and "Sequencing Quality and Regulatory Genomics" use cases. The "Somatic/Tumor Testing" and "Tissue Matching" use cases have additional complexity in that they synthesize references to sequence data with other relevant information to produce results that are themselves more complex. The five use cases and associated pilots are illustrated in Figure 6 below.

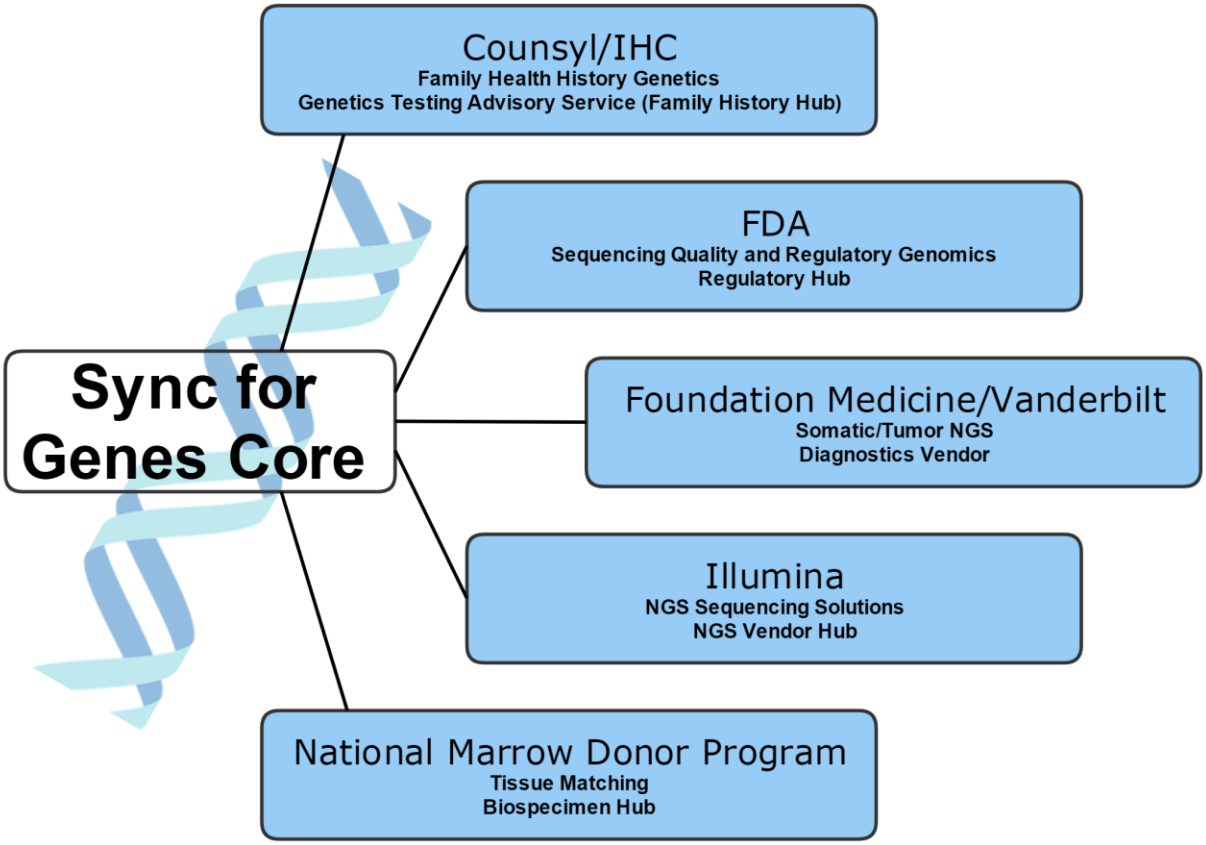

Figure 6: National pilots for Sync for Genes

Different pilots focused on specific technical implementations of FHIR, as shown in Figure 7 below and further described in the sections for each pilot. Work was done with the pilots to map from their internal format, if any, to FHIR Genomics format. There were three different platform implementations tested: read/write, read-only, and push servers. Counsyl and Intermountain Healthcare and the National Marrow Donor Program (NMDP) pilots focused on read/write servers. The FDA pilot is a read-only publically open server — with no authentication required. For Foundation Medicine, Inc. and Vanderbilt, the goal was to have a module that pushes FHIR out in a way that can be captured by another FHIR server receiver (which could ultimately include SMART-type features). A similar approach is planned within Illumina's BaseSpace, using an ephemeral applet to accomplish this task. All pilots were

successful in creating functioning FHIR-conformant code for their use cases. This is quite an accomplishment given the short implementation period.

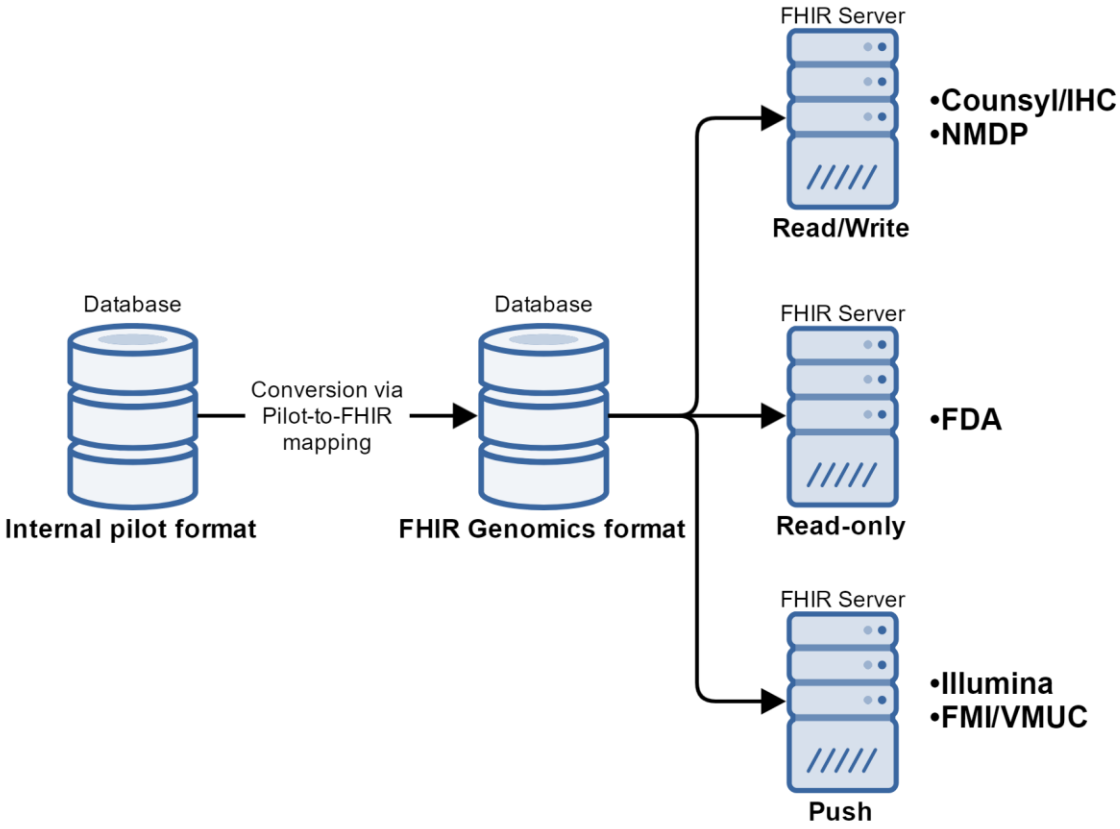

Figure 7: Flowchart of pilot approaches for Sync for Genes

Family Health History Genetics Use Case: Counsyl/Intermountain Healthcare Sync for Genes Pilot

Use Case Description

Genetic sequences, when coupled with other clinical methods and data such as patient demographics, laboratory test results, family history, pathology reports, and/or pharmacy reports, can lead to better decision-making through the utilization of family history tools, risk assessment tools, and drug dosage calculators. In some cases, clinicians translate (i.e., manually reenter) genetic data into tools for decision-making, but in other cases, patient genetic data from the EHR is automatically incorporated into clinical decision making tools. A standard data format can assist in data exchange for both of these cases, especially when the process is automated.

Access to personal and family history data adds important context and insight to genetic testing results. This use case focuses on making such history and associated genetic data available to patients, clinicians, and researchers.

Pilot Description

Counsyl is a health technology company that offers DNA screening for diseases that can impact men, women, and their children. To date, Counsyl has analyzed DNA for over 650,000 patients. Counsyl focuses on software that supports test-administration and interpretation of results, for both patients and medical providers. From this software platform, Counsyl has performed over 60,000 telehealth sessions.

Often, appropriate testing is based on personal and family health history. Counsyl's software and family-history application gathers information from patients that may be used for risk analysis, insurance reimbursement, and ultimately guiding appropriate genetic-testing. While family history is helpful in determining appropriate use of genetic testing, both the phenotypic data (personal/family history) and genotype data (genetic testing) has potential value to the patient, their family members, and the broader scientific community. Counsyl has found no standardized method to report this information to medical practices and their EHRs.

Sync for Genes Pilot Work

In this pilot, Counsyl and Intermountain Healthcare, with the Sync for Genes project, worked to demonstrate how genotype and phenotype data can be structured and shared using the FHIR protocol. Demonstrating and defining a standard framework by leveraging FHIR, has significant advantages for representation, interoperability, and data analytics. Specifically, the goal was to demonstrate an endpoint that provides both genomic and family history information, in order to share clinical genetic information with anyone who follows this standard.

The advantages of establishing a protocol to share genomic and clinical data include, but are not limited to:

- Enable patient-driven sharing with family members, providers, and researchers;
- Effective delivery to electronic health records and maintenance of the quality of data;
- Set a standard for collaboration across laboratories and institutions; and
- Streamline insurance reimbursement.

Pilot Results

The Counsyl FHIR server was designed and tested at fhir.counsyl.com behind a firewall. The architecture is a push-pull model. The destination (clinic) is pull-based and will pull the health care data it requires from a FHIR server. The source (laboratory) is push-based. The laboratory will push data to a FHIR server as data are generated so the destination can pull the data when needed. This is similar to the push/pull combination.[39]

Counsyl mapped their internal schema to the following resources and profiles: Patient, DiagnosticReport-genetics, FamilyMemberHistory-genetics, and Observation-genetics. The FamilyMemberHistory-genetics profile was used to capture the interrelationship between family members rather than just the relation to the patient. There was the desire to store different sharing of family history-related events separately. This motivated linking the FamilyMemberHistory Genetics profiles to DiagnosticReport Genetics Profile by reference to enable searching, rather than as contained

[39] https://www.hl7.org/fhir/pushpull.html

resource which prevents queries to the contained data. For identifiers, there will be distinct IDs on the Counsyl side, clinic side, and in FHIR. Counsyl will keep mappings from their internal ID to the FHIR server's ID. Likewise, the clinic would keep a mapping from their ID to one the FHIR server's ID.

Pilot Next steps

Both Intermountain Healthcare and Counsyl are interested and motivated to incorporate FHIR standards in their capturing and reporting of patient family history and genomic results. Counsyl, as a DNA testing and health tech company, is motivated to continue to make its data useful to its partners. Intermountain Healthcare is an ideal beneficiary of practical data in its management of millions of patients. The collective understanding and insights will drive personalized medicine and the overall quality of care the population. In addition, adding SMART apps that help with auxiliary reporting (i.e., risk calculation) can help drive faster adoption of FHIR.

NGS Sequencing Solutions Use Case: Illumina Sync for Genes Pilot

Use Case Description

NGS is the basis for many of the scenarios in the DAM. There are numerous examples of applications for precision medicine, NGS, and FHIR to work together. As a first step in an NGS approach, the clinician could use NGS platform[40] results to determine any information or data needed to investigate the case at hand. Just one example would be to look at the general category of rare disease diagnosis. From that point onward, NGS results can further be used to determine any information or data that would be useful for treatment decisions in the future as well. As an example, a clinician might be inclined not just to order somatic testing for an immediate treatment of cancer, but also to order a germline sequencing to determine if there are any genetic risks that may affect the patient later. A clinician may order a somatic and germline test together and receive one overall diagnostic report.

To achieve accurate and meaningful diagnoses among the other varied applications of precision medicine, there needs to be a way to determine the variations in the patient's genome compared with standard references by current knowledge. In this case, relevant variants would be stored and linked to the reference. A genome sequence can give additional insight to drug toxicity and metabolisms of various drugs, which can be derived from association with variation to the reference sequence.

NGS can yield sequence information which becomes associated with health, based on future scientific findings. Thus, new identification and classification of variants could yield future health-related information for a patient. In this scenario, the current sequence, including both sequences of clinical relevance and unknown relevance, is stored and can be reviewed further in the future.

Pilot Description

Illumina enables customers to read and understand genetic variations. Illumina has created BaseSpace Suite, a comprehensive, analytics-driven informatics solution that is integrated with Illumina sequencing instruments. BaseSpace Suite provides capabilities across the full NGS workflow, from sample intake and

[40] For background about NGS process see https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf

tracking (using BaseSpace Clarity LIMS), to genomic data management and secondary analysis (BaseSpace Sequence Hub), and ultimately to tertiary analysis (BaseSpace Variant Interpreter [Beta], BaseSpace Cohort Analyzer, and BaseSpace Correlation Engine), where the impact of genetic variation can be assessed and correlated to phenotypic information.

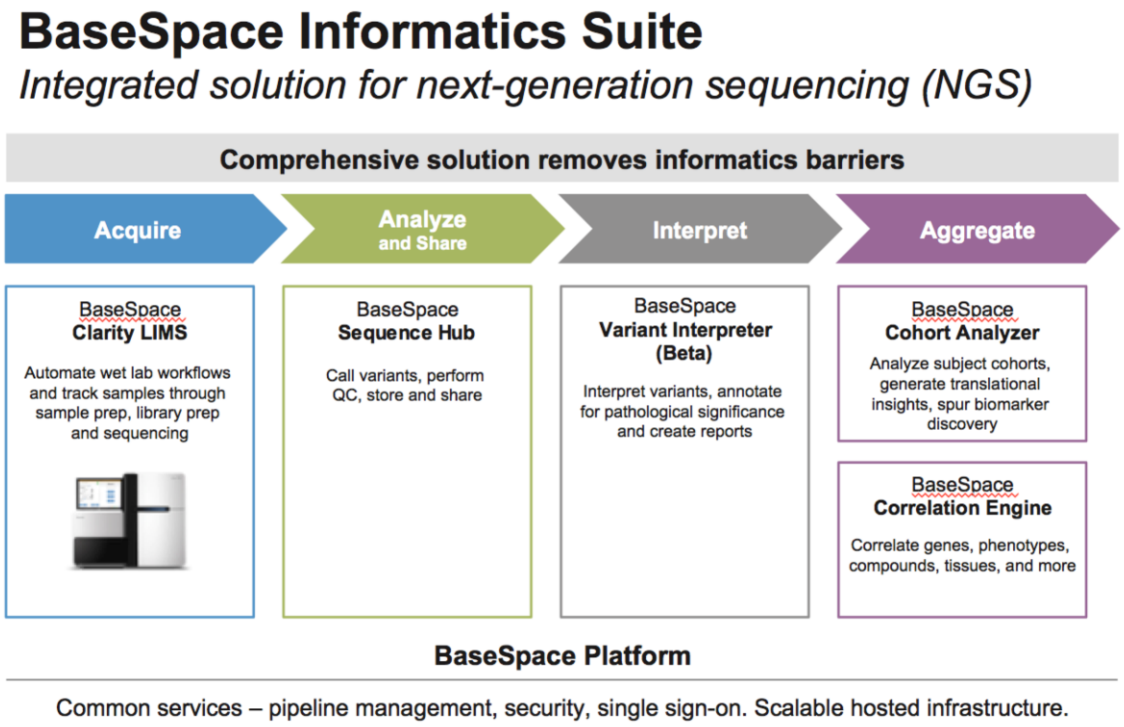

Figure 8: Illumina BaseSpace workflow (image courtesy of Illumina)

Critical to deriving knowledge from genomic data is to combine data from sequencing instruments with other information including phenotypic data and information from personal sensors (e.g., fitness tracking devices and apps). Combining data provides the opportunities for greater levels of insights, for example to support clinical research programs.

Sync for Genes Pilot Work

To support interchange of data with multiple sources of phenotypic data, Illumina used the FHIR Genomics standard to support data exchange between BaseSpace Suite apps and phenotypic data sources. The goal of this was to implement FHIR Genomics as part of BaseSpace Suite Informatics Platform, making it accessible to all apps within BaseSpace Suite, as well as providing sources of phenotypic data access to genomic data from any of the BaseSpace Suite apps. This work is not for use for diagnostic purposes.

Pilot Results

Illumina has built an application within BaseSpace Sequence Hub, a cloud-deployed sequencing analysis platform, that will take Variant Call Format (VCF)-derived information, convert it to the current draft FHIR Sequence Resource, and send it via a POST API call to a user-defined public FHIR server. Existing frameworks were used to configure a BaseSpace Sequence Hub application that performs a conversion

of VCF to FHIR Genomics. This allows genomic data to be communicated to another FHIR Genomics-compliant server/application.

Separately, an internal audience was engaged to consider the use cases for VCFs and the FHIR standard, so as to inform the working group on situations Illumina wanted to see supported by the standard.

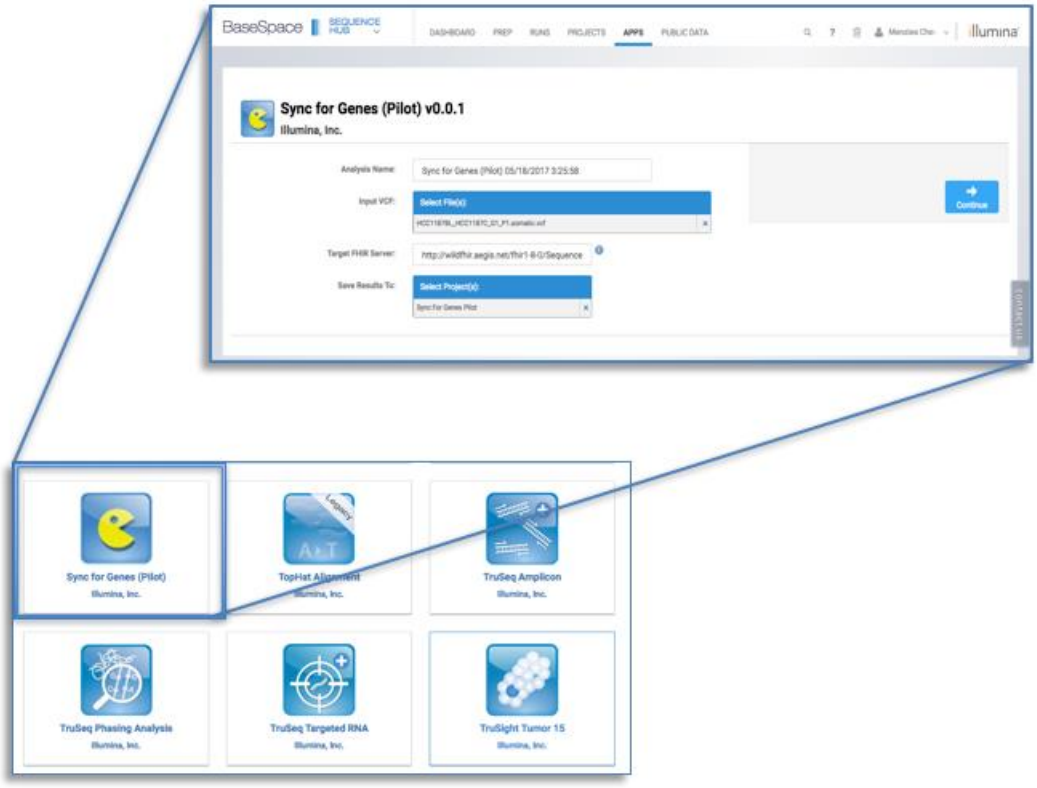

Figure 9: Illumina-created app and BaseSpace app gallery (image courtesy of Illumina)

Pilot Next Steps

Illumina built the initial implementation as a proof-of-concept for exporting VCFs from BaseSpace Sequence Hub in FHIR Genomics format. Further iterations may include:

- Authentication into an external server;
- Additional resources/profiles being created and sent; and
- Implementing the standard as a bidirectional service for working with both genomic and phenotypic data, to be utilized across the full set of products that make up BaseSpace Suite.

Sequencing Quality and Regulatory Genomics Use Case: FDA Sync for Genes Pilot

Use Case Description

One of the challenges to routinely using NGS data in a clinical setting is an incomplete understanding of the diagnostic regulations surrounding NGS data analysis. It is critical to understand the limits of

accuracy, precision, and clinical validity of NGS. In response, the FDA has developed precisionFDA,[41,42] a public research portal that allows the research community to characterize, test, and improve NGS data processing techniques and to share the results.

Adding Sync for Genes capabilities to the precisionFDA portal would not only provide a method to exchange NGS quality and regulatory data based on FHIR Genomics, but because of its open and public nature, it would also promote the adoption of such data exchange standards in the community at-large.

Pilot Description

The FDA pilot made a public FHIR Genomics-based API and deployed it as part of precisionFDA, allowing access to a community platform for NGS assay evaluation and regulatory science exploration. The precisionFDA site allows participants to benchmark the analytical validity of a genomic test, pipeline, or other variant-detecting methodology that produces a VCF file. This is done by comparing a user-provided VCF file against reference VCF material such as the National Institute of Standards and Technology (NIST)/Genome-in-a-bottle reference call sets or the Illumina Platinum Genome call sets.

Such comparisons assess the level of agreement between a test VCF and a benchmark VCF by counting the number of true positive, false positive, and false negative entries, and computing statistics such as sensitivity, positive predictive value, and f-measure. Users on precisionFDA have the option to publish the results of a comparison, so that the public can access it and see the calculated measurements. By accessing public comparisons, the greater community can explore and understand the performance of NGS pipelines.

Sync for Genes Pilot Work

To make these results more accessible, the FDA implemented a FHIR Genomics endpoint via precisionFDA, so that results of public comparisons can be exported as Sequence FHIR resources using the FHIR API.

Pilot Results

The FDA's implementation was built on their existing platform. On the precisionFDA site, users with accounts can run VCF comparisons to assess the quality of their pipelines and genetic tests and choose to publish them for other site users to see. Comparisons that are published for the entire precisionFDA community to see are deemed "public" comparisons. In the scope of this pilot, precisionFDA made all its public comparison data available for access without the need for a precisionFDA account.

The precisionFDA site runs Ruby on Rails (Rails) in the cloud on Amazon EC2 instances. The code is open source and available online.[43] For the Sync for Genes pilot, they created additional Rails routes (corresponding to URLs) to expose all public comparison data. The Rails framework makes it easy to take a URL plus parameters, and run a database query, manipulate the returned, and render it in a format all

[41] https://precision.fda.gov/
[42] https://blogs.fda.gov/fdavoice/index.php/2015/12/fda-launches-precisionfda-to-harness-the-power-of-scientific-collaboration/
[43] https://github.com/FDA/precisionFDA

relevant to the provided parameters and path. The overall approach of the site itself follows the Model View Controller (MVC) that Rails ascribes to. In this case, the comparison database records are the models, the schema formatting logic act in the controllers, and the resulting JSON/XML objects are the views.

The implementation made use of the FHIR Sequence,[44] Bundle,[45] and CapabilityStatement[46] resources. Sequence objects encapsulate individual comparison result records, Bundle objects are used to return query results with possible multiple Sequence objects, and the CapabilityStatement are used to broadcast what data and formats the FHIR-related paths on precisionFDA support.

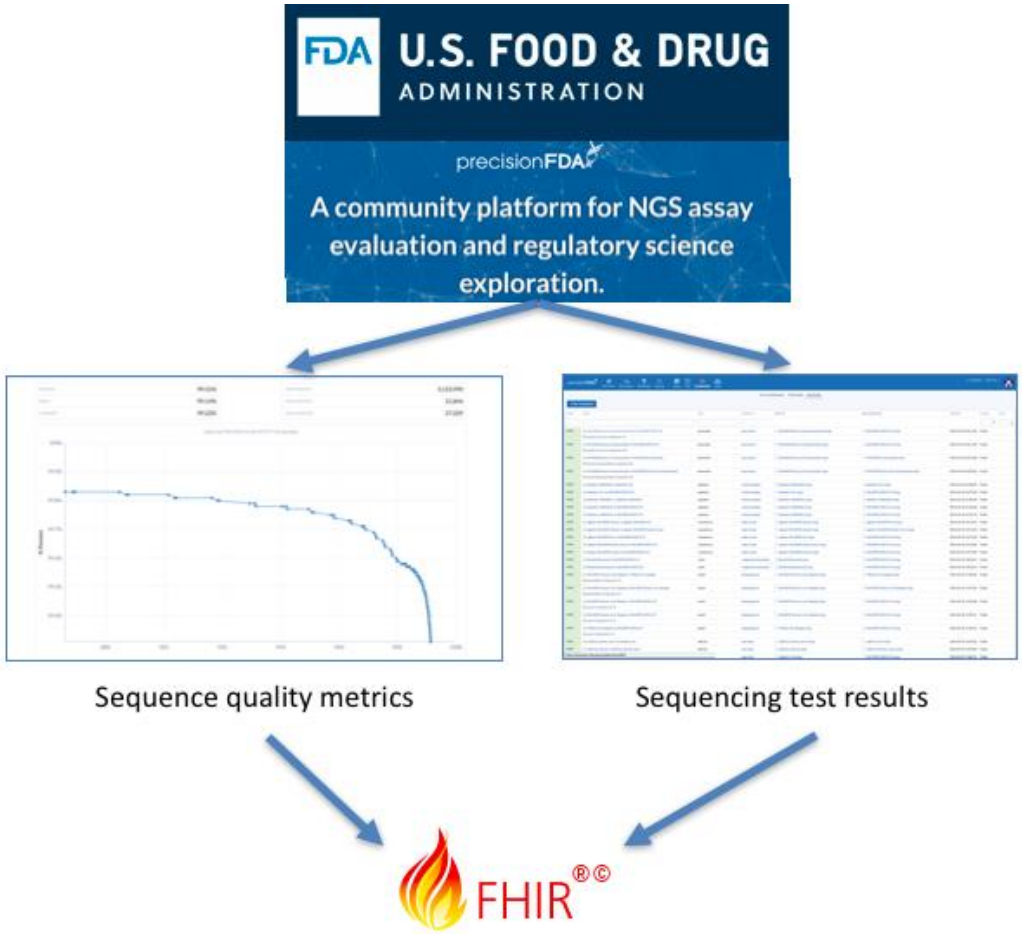

Figure 10: precisionFDA pilot overview

Please see Figure 10 for an illustration of the pilot. There were three main components in implementing the FDA pilot, none of which proved significantly difficult, according to the pilot site. These steps simply

[44] https://www.hl7.org/fhir/STU3/sequence.html
[45] https://www.hl7.org/fhir/STU3/bundle.html
[46] https://www.hl7.org/fhir/STU3/capabilitystatement.html

required some coordination to draft the schemas and expected behavior by comparing with test servers sites.

- **Retrofitting precisionFDA data into the FHIR Sequence object, and defining new FHIR schema fields.** The pilot site needed to identify what fields in the FHIR Sequence object made sense for the VCF comparison data that precisionFDA has and what new fields were required in order to properly describe any data fields that were not accounted for. This may require future work since the VCF comparison framework on precisionFDA may be subject to change as comparison tools improve and diversify.
- **Exposing the data.** There was just a bit of technical work to expose the relevant data according to the agreed upon schema. The existing architecture made it easy to fetch the public comparison data and manipulate it properly to expose it as FHIR Sequence objects with both JSON and XML support.
- **Implementing basic FHIR functionality.** Since the pilot implemented the Sync for Genes pilot on the existing precisionFDA site, some basic FHIR routes had to be manually created. This included implementing a path for the capability statement, implementing bundles for listing all possible results or subsets of results, which also requires implementing basic pagination and querying ability for the various sequence objects. As with the schema definitions, this may be altered in the future as the comparison data becomes richer and demands that more query-able fields be supported.

When implementing the FDA pilot, the pilot found that FHIR was accommodating in that it had low minimal requirements in terms of required data fields. This allowed a straightforward implementation without worrying about bloating with too much data either unnecessary or irrelevant to the precisionFDA system. For systems just onboarding FHIR Genomics, this low barrier to entry is key in allowing for smoothly and easily adopting specification conformance.

Pilot Next Steps

There are a number of possibilities for future extensions to the FHIR workflows implemented for the pilot. PrecisionFDA revolves around sharing and comparing data, and so do the next steps, which are to:

- Host and publish additional public comparison data as the community continues to grow;
- Host additional genomic reference standard data in addition to the VCF comparison data that is currently available;
- Host genetic testing quality control data;
- Further implement FHIR Genomics as required to support new data; and
- Make additional comparison data, genomic reference standard data, and testing quality control data available via FHIR.

Somatic/Tumor Testing Use Case: Foundation Medicine, Inc. and Vanderbilt University Medical Center Sync for Genes Pilot

Use Case Description

Somatic genetic testing investigates a patient's acquired genetic variants, as opposed to hereditary "germline" variants, which are inherited from parents. Somatic mutations are specific to a particular tumor and are not found in other tissues, and may have active "driver" effects or passive "passenger"

effects. Additionally, the same mutations identified in different cancers may have different clinical implications, so it is important to understand the genetic variants present in a specific tumor and correctly assess pathology. A tumor biopsy can be sequenced to understand the genetic variations specific to a particular case.

Pilot Description

Foundation Medicine, Inc. (FMI) and Vanderbilt University Medical Center (VUMC) have a long-standing relationship as provider and consumer of NGS tumor panel testing, respectively. Initially, VUMC received NGS reports from FMI via standard fax, as is typical in the medical industry. This resulted in the conversion of the original full-color NGS reports to black and white scanned artifacts. An assessment of the faxed NGS reports demonstrated poor information fidelity and lack of provider notification, which prevented optimal utility of molecular profiling at VUMC. In close collaboration, FMI and VUMC custom developed a method to package a post-processed laboratory report into an extensible markup language (XML) file that included demographics, ordering information, and the FMI-designated actionable variants that recapitulate the first page of the NGS report. The process of this development and successful implementation was described in the *Journal of Oncology Practice*.[47] Furthermore, VUMC subsequently demonstrated that structured data provided by FMI could be used to fulfill operational, research, and hypothesis generation needs.[48]

Although this collaboration has been successful, the customized nature of the XML files limits generalizability. FMI has the capability to transmit standardized genomic file formats (e.g., BAM); however, these are used primarily for research and electronic transfer of clinical reports remains a customized solution.

Sync for Genes Pilot Work

In support of the Sync for Genes effort, VUMC and FMI proposed to transform the existing custom XML data structure into a format compliant with the latest FHIR specification Release 3 API. The transformed data can be stored on either FMI or VUMC servers and can be accessed on-demand using standard FHIR technologies. These data can then also be exposed to local instantiations of SMART on FHIR apps, e.g., SMART Precision Cancer Medicine,[49] an app for display of cancer genomic data that was highlighted in the 2016 President's Cancer Panel report.[50] As a demonstration of principle, this collaboration illustrates how clinical reporting of cancer genomics data can be standardized across laboratories and institutions.

---

[47] Rioth MJ, Staggs DB, Hackett L, Haberman E, Tod M, Levy M, et al. Implementing and Improving Automated Electronic Tumor Molecular Profiling. J Oncol Pract Am Soc Clin Oncol. 2016 Mar;12(3):e332-337.

[48] Rioth MJ, Thota R, Staggs DB, Johnson DB, Warner JL. Pragmatic precision oncology: the secondary uses of clinical tumor molecular profiling. J Am Med Inform Assoc JAMIA. 2016 Jul;23:773–6.

[49] Warner JL, Rioth MJ, Mandl KD, Mandel JC, Kreda DA, Kohane IS, et al. SMART precision cancer medicine: a FHIR-based app to provide genomic information at the point of care. J Am Med Inform Assoc JAMIA. 2016 Jul;23:701–10.

[50] https://PresCancerPanel.cancer.gov/report/connectedhealth&sa=D&ust=1495117584691000&usg=AFQjCNGLvvu0iVgtc3IxE7kPmClpdViaAA

Pilot Results

The pilot created FHIR Genomics R3 data based on FMI customized files. The FMI-VUMC pilot used a RESTful framework to put and retrieve information from several FHIR R3 servers. The data being transferred to the test server were FHIR JSON objects mapped from FM XML report files. All data elements defined in the FMI data dictionary were successfully mapped to existing FHIR Resources or metadata profiles. Extensions were required to create explicit relationships between certain resources (e.g., to attach a literature reference to a specific genomic observation).

The approach taken in the pilot was assuming that there were no existing resources on the server that explained clinical/personal/phenotypic information about the patient (e.g., Condition resource and Patient resource). The pilot made the assumption that information on associated resources were not present on the server as well (e.g., Practitioner resource and Organization resource). To circumvent these assumptions in the future, it would be beneficial to share FHIR resource IDs between the provider and consumer so that existing resources could be updated seamlessly. The resources/profiles used were: Organization, Practitioner, Patient, DiagnosticReport, Condition, ProcedureRequest, Specimen, Provenance, Observation genetics profile, and Sequence resource.

Applications developed during the duration of this pilot included a web app called FM-to-FHIR, which was implemented using an AngularJS front end and a Django backend. FM-to-FHIR enables user-specified FMI XML files to be mapped to FHIR resources behind the scenes in JSON format and uploaded to the user-specified server. Another application developed was a program that maps all of the FMI XML files in a directory to FHIR resources, and writes the outputs as Bundle resources in JSON format.

Arguably, one of the most important contributions of this pilot was the refinement of FHIR Genomics through successful mappings of various types of genomic alterations. These alterations included substitutions, small insertions and deletions (indels), copy number changes, and rearrangements. The current Sequence resource structure allowed for a seamless mapping of substitutions and indels. Copy number changes were best represented by the structureVariant element of the Sequence resource from the January 2017 STU3 candidate build (Release 1.8.0) and is being currently evaluated for future versions of FHIR. Rearrangements were also mapped using the structureVariant element, and both regions of the rearrangement were related to each other on the Observation resource level. This pilot's mappings helped provide context for a discussion on how to best represent rearrangements in FHIR.

Pilot Next Steps

Once FMI migrates from their custom XML to the new FHIR JSON format, the next steps are to:

- Transmit protected health information (PHI) from the FMI server to a clinical-grade VUMC genomics server;
- Transmit the PHI information from the VUMC genomics server to the VUMC clinical system(s);
- Either reroute the data from the VUMC genomics server to and Epic EHR clinical system, if Epic builds a module to accept externally-sourced genomic data, as VUMC is scheduled to go-live with an Epic EHR system in late 2017; or
- Reroute the data to an LIS system at VUMC, which is a Cerner-based EHR system.

Ultimately, if Epic and/or Cerner build FHIR Genomics capabilities into their native FHIR servers, the next logical step would be to transfer relevant data from the FMI servers directly to the Epic and/or Cerner FHIR servers, making the data potentially available for clinical care and research studies such as *All of Us*.

Tissue Matching Use Case: National Marrow Donor Program/Be the Match Sync for Genes Pilot

Use Case Description

Potential organ and tissue transplants must be screened to ensure compatibility with the patient, with the most important biomarker being Human Leukocyte Antigen (HLA). Working with the HLA community, the National Marrow Donor Program (NMDP) helped develop a set of guidelines, Minimum Information for Reporting Immunogenomics NGS Genotyping (MIRING),[51] which identifies eight reporting principles for NGS-based genotyping of immunogenomics data. MIRING data is reported in an XML-based technical specification called Histoimmunogenetics Markup Language (HML).[52] HML, while used in the HLA community, is not widely interoperable with other healthcare systems, such as EHRs. Encapsulating MIRING principles, as informed by HML, into FHIR advances interoperability in health care.

Pilot Description

The National Marrow Donor Program® (NMDP) is a non-profit organization that enables patients to receive the marrow or umbilical cord blood transplants they need. NMDP operates the Be The Match Registry®. NMDP's partnerships with international and cooperative registries provide doctors with access to nearly 27 million potential donors and more than 680,000 cord blood units worldwide.

Each year, NMDP receives HLA typing for over 500,000 new potential donors for the Be The Match Registry®. These laboratory tests are done by two contract laboratories that must comply with standards NMDP establishes. Other laboratories, transplant centers, registries, and other partners typically comply with standards that NMDP promotes to maintain interoperability with the network.

NMDP is the central hub of a network of over 500 partners, including donor centers, collection centers, apheresis centers, transplant centers, testing laboratories, recruitment groups, member cord blood banks, international donor centers and cooperative registries, with immunogenetic data (including HLA) being exchanged between partners. NMDP Bioinformatics Research[53] has a history developing and promoting software and data standards for sharing immunogenetics biomarker data such as HLA and killer-cell immunoglobulin-like receptors (KIR). NMDP has developed the HML, MIRING, and Genotype List String,[54] and promoted their adoption by NMDP laboratories, vendors, and partners.

---

[51] Mack SJ, Milius RP, Gifford BD, et al (2015) Hum Immunol. 76(12):954-62. doi: 10.1016/j.humimm.2015.09.011

[52] Milius RP, Heuer M, Valiga D, et al (2015) Hum Immunol. 76(12):963-74. doi: 10.1016/j.humimm.2015.08.001

[53] https://bioinformatics.bethematchclinical.org/

[54] Milius RP, Mack SJ, Hollenbach JA, et al (2013) Tissue Antigens. 82(2):106-12. doi: 10.1111/tan.12150.

Sync for Genes Pilot Work

NMDP now wants to move toward seamless interoperability with health care systems, integrating research results, as well as empowering donors and patients. To make this possible, NMDP has embraced open source programming and open standards for developing its next generation of tools (HML 2.0) and data standards by leveraging FHIR and the genomics standards for creating HLA typing reports using FHIR Genomics resources and profiles. NMDP has created and posted to external FHIR servers individual resource instances and transaction bundles containing HLA data, are developing terminology services for HLA nomenclature, and will be deploying its own FHIR-compliant development and testing server (using a customized HAPI[55]-based implementation).

Ultimately, NMDP believes that these standards-based endeavors will make it possible for it to achieve its vision of exchanging patient/donor immunogenetic data, with consent, directly with EHRs, typing laboratories, as well as other health care and research systems.

Pilot Results

NMDP followed the HLA reporting bundle strategy described section 7 (HLA genotyping results Profile) in the genomics implementation guidance document in FHIR. Reports were transaction bundles that are informed by MIRING principals and the HML. They educated their partners through their Data Standards Hackathon (DaSH), which NMDP organizes several times a year, on how to use FHIR and the HLA reporting bundle. As part the DaSH events, their partners upload HLA typing reports to the NMDP HAPI-based FHIR server.

The client was either a simple python script, curl command line, or an AngularJS web based client submitting an HLA report bundle to the FHIR development server. Resources used included Sequence, Specimen, Organization, Patient, Bundle, and the clinical genomics profiles (Observation, DiagnosticReport, DiagnosticRequest, HLA Genotyping Results).

Pilot Next Steps

The NMDP will:

- Continue to develop HLA reporting solutions using the Clinical Genomics resources and profiles; and
- Expand other FHIR solutions in their organization beyond clinical genomics.

General FHIR Recommendations

Many recommendations and feedback regarding FHIR came up during the course of the Sync for Genes pilots. They are discussed in the subsections below.

FHIR Resource IDs

[55] http://hapifhir.io/

It was felt that ways to refer to specific resources/profiles (e.g., system-specific ID vs. FHIR ID) may introduce some redundancy and was suggested that they may be better defined and/or unified. This is a general FHIR item (not genomic-specific).

FHIR Documentation

Documentation was seen as comprehensive by some pilots, but needing additional use case examples and/or different visual formatting by others. For example, links within the FHIR build documentation[56] that point internally to other parts of the document could, in general, be improved to enable easier navigation across documentation. It was suggested that a "Required flag" on all required properties be highlighted in a separate column on the description of FHIR resources overall. This would allow people to easily see it and to more easily use utilities to autogenerate a schema or Object Relational Mapping (ORM) to easily annotate in their code, facilitating use of an API framework to easily autogenerate a resource object schema in an automated fashion. This may help FHIR be adopted by more developers in general. Currently, this is captured in the cardinality field of the resource description, but this data point can be difficult to parse out. One solution would be to add explanation of how to extract cardinality information automatically via the open source FHIR schema and/or an automatically generated required field column in the FHIR build process.

Codeable Concepts

Additional details on codeable concepts[57] would be useful for implementers in some cases, e.g., where detailed descriptions or expansions of data are not listed. A codeable concept is essentially a value set or ontology reference. Expanding such codeable concept definitions and expansions of properties to the primitive data level (e.g., Boolean, string, integer, decimal) would make for more robust resource definitions. For example, this applies to several of the FDA benchmark fields and coordination with FDA on this can help in this effort.

In addition, derived specimens often used in NGS (such as via Formalin-Fixed Paraffin-Embedded [FFPE] block, or cell block) need to be captured by the FHIR Specimen resource. This could be done by augmenting value sets used for capturing specimen type and/or processing procedure information.

Sequence Resource

The "referenceSeq" section[58] of the current specification can be used to define the sequence's test space. FHIR would benefit from having official guidance around this aspect.

For the sequence benchmarking fields and newly recommended Receiver Operator Characteristic (ROC) quality metric, there is currently no established standard terminology associated with the elements. While not essential to performance, in the future it may be useful to define a standard value set, for these to further optimize the resource's structure.

Observation Genetics Profile

---

[56] https://hl7.org/fhir/STU3/genomics.html

[57] http://hl7.org/fhir/STU3/datatypes.html - codeableconcept

[58] http://hl7.org/fhir/STU3/sequence-definitions.html#Sequence.referenceSeq

There was the idea that it would be helpful to organize information in the Observation Genetics profile by hierarchically structured elements, much like how Sequence is organized. For example, phaseSet can have ID and Sequence reference hierarchically underneath it. Finally, there was a recommendation to add specific extensions to observation to store classification information (currently done via laboratory codes, which require an extra step to learn, parse, and encode). At a minimum, this would contain variant, disease (interpretation), and classification (e.g., benign, unknown, pathogenic) information.

For genomic testing involving trio analysis,[59] it was noted that there could be guidance around recording inheritance as used in a trio analysis, such as "de novo," "accumulative," "inherited from mother," "inherited from father", "inherited from both." This could, for example, be added to interpreted variant information under the Observation Genetics profile.

Another item examined by several pilots was the organization of variants via the interpretation model in order to separate out raw information into a sequence resource and interpreted, optimally searchable information into an Observation Genetics profile. Consensus was expressed on the current design. For example, it was noted that a Sequence resource has coordinates that are tied to a specific genome build. The variant ID location (e.g., Reference SNP ID [Rs ID]) and other annotations (e.g., pertaining to a gene context) typically remain stable even between genome builds. For example, Rs ID 12345 will have different genomic coordinates between two different human genome assemblies (e.g., GRCh37 and GRCh38), but they will refer to the same variant. This justified having sequence data and interpreted information reside in two different locations within FHIR. Thus, the ID is kept in the Observation Genetics profile and the genomic coordinates are stored in Sequence. Addressing the modeling of variants with respect to searchability, it was felt that denormalization by including the variant ID in the Observation Genetics profile also helps facilitate optimized searching.

DiagnosticReport Genetics Profile

Literature references are a core component of cancer molecular profile reports. It was suggested that FHIR should support incorporation of such references into a diagnostic report. While DiagnosticReport currently does allow for attachments, those were felt to support the diagnosis itself rather than supporting information about the underlying genetics.

Cancer molecular profile reports often make treatment recommendations based on the published literature and assign a level of certainty to such recommendations. These are not clinical decision support, but rather they are provided as a prompt to the ordering clinician to make use of the results. Right now, the manner in which these recommendations are made is highly variable between laboratories, but, should be standardized and ideally be mapped to one or more of the professional society recommendations.[60] Based on interactions with FHIR Infrastructure, HL7 Connect-a-thon, and CGWG, it was determined that a path forward would be to add this information to the DiagnosticsReport Genetics profile with extensions to link to the references/recommendations.

---

[59] An example of a trio analysis https://www.bcm.edu/research/medical-genetics-labs/test_detail.cfm?testcode=1600

[60] Li et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists, (2017).

Looking toward the future, Sync for Genes pilot input be used to examine FHIR Genomics support for and explanation of facilitating implementations, such as for CLIA[61] or College of American Pathologists (CAP)[62]. For example, laboratories need to communicate certain minimum elements that CLIA requires and receive acknowledgements back that data was received. Other minimum suggested elements to transmit might have certain domain-level specificity.[63] These can be defined in documentation for existing elements, defined as new elements, and/or defined as domain-specific profiles on top of the standard genetics profiles.

FamilyMemberHistory Genetics Profile

The pilots determined that additional ancestry codes should be able to be captured for each family member. Currently, the FamilyMemberHistory Genetics Profile has a reference to an Observation resource for this.

While Logical Observation Identifiers Names and Codes (LOINC) have codes for demographic ethnicity-type information, this type of information is generally captured separately as administrative information in the Patient resource. The Sync for Genes pilots thought that there should be new category for specifically capturing genetic ancestry and associated codes for the type of ancestry (including percentage and source of information — patient survey vs. genetic testing). This new category would be used in the Observation Genetics profile.

FHIR Searchable Parameters

A list of queries (Table 1 below) based on pilot input for the Sync for Genes use cases was established. These queries may be useful to add to future publications of the DAM as a basis for future development of FHIR. FHIR can optimize search parameters by designating elements as searchable. This provides a mechanism to make elements that should be searchable and readily accessible by developers. To that end, it was recommended that commonly used fields be exposed as named extensions in standard profiles, as is done currently, rather than as entries within an Observation's component element. This enables the structure of FHIR to be leveraged and to encode of relationships. This also obviates the concern whereby use of different coding systems (e.g., Systematized Nomenclature of Medicine [SNOMED] vs. LOINC) may obfuscate the underlying semantics. Table 1 summarizes the types of queries that the pilots would find useful.

Table 1: Standard queries for use cases from pilots

Search Categories	Desired Search Capability
Basic Queries	Find patients by condition and affected status
	Find patients by phenotypical attribute

[61] https://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html
[62] http://www.cap.org/web/home
[63] https://www.ncbi.nlm.nih.gov/pubmed/27814769

Search Categories	Desired Search Capability
	Find patients by molecular data type available
	Find patients by specific molecular marker (genomic region, position, gene, SNP) Find mutation by mutation type and position
	Find patients based on mode of inheritance
	Search for patient by ID
	Search for diagnostic reports with a given variant
	Given diagnostic report, return associated family history information
	Search for variants given gene and variant classification
	Search for patients based on genetic ancestry
Boolean Queries	Find all patients with a set of variants
	(Support AND, OR, NOT, and nested Boolean queries)
Specimen-based Queries	Find specimens collected from specific specimen body sites
	Find specimens from this patient's pedigree
	Find specimens based on patient disease status
	Find specimens based on specimen disease status
	Find specimens based on technology platform
	Find specimens based on instrument identifier
Time-based Queries	Find specimens collected within an absolute date range
	Find specimens collected within a relative date range

DAM Use Case Development

The DAM is the basis for the FHIR Genomics design. A couple of use cases need to be further developed for next version of DAM, which is in currently undergoing development and scope expansion after its initial release in February 2017. Thus, as the DAM is further expanded in scope these use cases will be further defined. Some areas are currently are under active development in the CGWG. This includes cytogenetics/Fluorescence In Situ Hybridization (FISH) and structural variant analysis where the exact sequence or change in sequence is not known. FHIR should enable capturing these use cases ideally in discrete International System for Human Cytogenetic Nomenclature (ISCN)-compliant fields. Regarding rearrangements, sometimes even if these are detected at the sequence (DNA) level, the partner gene is unknown and/or the breakpoint is unclear. The Sync for Genes pilots' input on this matter is to use the Sequence resource (with pointers from Observations that are related to each other) and Observation.

Discussion and Conclusion

General discussion

Pilots found it invaluable that FHIR Genomics allows for a highly flexible range of possibilities and use cases in terms of conformance and adoption. There is no "minimal" set of requirements overall, only that the data published has to conform to the FHIR schemas, which have elements/values that can be enforced as required via cardinality/constraints. This and the "levels" system (described previously) make it easy for any system to start onboarding FHIR Genomics in manageable steps. For future work, additional time for interactions between the pilots would help the pilots learn more from each other as well.

General pilot feedback

Pilot feedback was gathered about Sync for Genes as well as thoughts about FHIR Genomics before and after the program. Several pilots had similar views on the importance of FHIR, and specifically FHIR Genomics, as a key toward clinical genomics interoperability. For one pilot, the thoughts about implementation of FHIR Genomics are more or less the same before and after the program: "FHIR itself is a great advancement towards the interoperability of health care and resources, and consequently a great advancement towards improved patient health care. As sequencing technology has advanced rapidly, FHIR Genomics is a necessary response to the massive reduction in sequencing costs and massive increase in sequencing data. Clinical genetics tests will only increase in popularity and adoption, and therefore the need to lay the groundwork for adopting clean, standardized formats for genomic health[ ]care data are critical in the current time. Leveraging the already-in-place HL7/FHIR framework, FHIR Genomics is a natural but important step towards removing the barriers between effective research and clinical health care."

Another echoed similar thoughts saying it "believe that the development of the FHIR Genomics standard is a crucially-important development for supporting the adoption of genomics for precision medicine initiatives, as well as for basic biological research. The ability to easily combine genomic and phenotypic data are paramount for advancing the utilization of genomic data. [The pilot] believed this before the Sync for Genes program, and having been involved in the pilot program, is excited to see the initial realization of this achievement with the success of the program."

In addition, there was the comment that "to be truly useful in an actual clinical setting[,] there needs to be shared patient information between the vendor and the hospital/center requesting the genomic

information... [Namely], it would be highly beneficial if the vendor returned the results of the genomic testing with actual patient information included (e.g., the patient ID in the EHR) so that we can seamlessly link/associate the genetic information with the patient's clinical/phenotypic information."

In terms of testing, the thinking was that "the work [Sequence resource and profiles] is solid by having the input by many in the workgroup and continues to evolve. The work we did as part of the Sync for Genes pilot was a natural outcome of the work [the pilot] was already doing. This was a great opportunity to more formally vet the standards in real work. The Clinical Genomics landscape is both broad and deep and these standards need to be tested through pilots such as this."

Themes Explored During Pilot

During the course of this Sync for Genes Pilot, the pilots explored different aspects of FHIR Genomics for their implementations. The recommendations and finding are discussed here.

FHIR Queries

FHIR's query capability generated great interest. It was discussed that FHIR implementers do not have to make all available data elements searchable. The FHIR capability statement[64] specifies which elements are searchable. And since all FHIR communication, including URL information, should be secured with TLS/SSL[65] there is minimal risk to data leakage by using search parameters in the URL. Some pilots were, of course, interested in seeing search examples in the documentation, and while there are examples available in the core FHIR specification, additional examples for various FHIR Genomics use cases could be useful. Lastly, the returned results from a query and their format were of interest. Results are returned as a Bundle resource[66] that contains a limited number of results per "page" and can further results can be requested as desired. One pilot did see a need for the Bundle to contain just pointers to other resources rather than the resources themselves because of the potentially large volume of data being returned. FHIR's search capability does not do this currently, but it is worth looking into, possibly by designing certain resource elements as references.

Value Sets

As touched upon earlier in this document, pilots had interest in value sets and how to define and utilize them in FHIR. The CGWG can coordinate with future pilots/implementers to define meaningful value sets for use cases and workflows. During this pilot phase, value sets related to publication references, ancestry, and specimen type were discussed. Value sets can be specified for use within FHIR, and FHIR plans on allowing value sets to be registered at the FHIR website as a way for a group to maintain a value set.

FHIR Server Functionality

The functionality of FHIR servers, in general, interested the pilots. Contained resources,[67] in which one resource is contained within another resource as opposed to being referenced, typically should not be

---

[64] https://www.hl7.org/fhir/STU3/capabilitystatement.html
[65] https://www.hl7.org/fhir/security.html
[66] https://www.hl7.org/fhir/STU3/bundle.html
[67] https://www.hl7.org/fhir/STU3/references.html#contained

used when the resource can be identified on its own (i.e., it is a fully formed resource) or is expected to be searchable independently. Searching on contained resources introduces special issues for consideration[68] that need to be taken into account.

Details related to storing data on the FHIR server were explored. The order of operations required when multiple resources reference each other can get complicated, but a resource can always be modified to reference a newly-created resource. It should be noted that when a new resource is created, it is assigned an ID internal to the FHIR server that cannot be modified, but it also has a separate element called "identifier" that is designed to hold its business identifier (e.g., patient ID).

Future work

As computers and electronic records became ubiquitous, more clinical information became available in electronic form and available for physicians and empowered patients. Eventually, information stored based on clinical encounters exceeded the amount of such information obtained specifically for research. Likewise, it is expected that clinical sequencing will exponential increase in the coming years and this data to empower physicians and patients. And, similarly, the number of patient sequences done for *clinical* utility will likely outpace the number of sequenced for *research* in the coming years. Thus, standards for communicating this information from laboratories for use at point of care, and facilitating its potential secondary use for research, will become increasingly important in the years ahead.

FHIR Genomics has come a long way since initial development for FHIR DSTU2. The Sync for Genes program and associated pilots have provided a real-world test bed for the standard development and clarified the vision for next steps. There are several important areas for future development and pilot testing to facilitate clinical genomics at point-of-care and use in research studies such as *All of Us*. Recommended next steps for the Sync for Genes mission in this area include these eight areas for development:

1. Facilitate piloting through testing with the DAM workflows and business rules in context.

> All of the pilots' implementations in this phase of Sync for Genes were successful proofs of concept. For example, FHIR Genomics support on the FDA's official precisionFDA site has been deployed for public use and the Sync for Genes app has been used by Illumina in BaseSpace. The next step to facilitate real-world use case implementation is to pilot full workflows/business rules as in the DAM, while incorporating the findings from the pilot phase into the next round of development of Sync for Genes, FHIR Genomics, and the pilots individually. Using workflows and business rules will give vital context and guidance for many future implementation decisions, including FHIR specification/services content (such as terminology), privacy and security, and even the system architectures that would enable the workflows. These could also address CLIA-related issues in workflow at laboratories and in information flow between laboratories and other stakeholders.
>
> In some cases, workflows and business rules in any future Sync for Genes activity need to account for the problem of patient matching. Patient matching can be an issue when

[68] https://www.hl7.org/fhir/search.html#containedType

transmitting or searching for data about a particular patient. Data systems in different laboratories, clinics, or repositories will assign an identifier to a patient, but often this identifier is not known to any outside system. FHIR itself defines a mechanism to perform patient matching but leaves the matching implementation details open so that any suitable specification can be used.

2. Facilitate piloting that test of additional use cases from the DAM that *integrate* multiple pilot sites/stakeholders in a shared use case.

During the pilot program, a subset of the HL7 DAM Clinical Sequencing use cases was tested. Certain other use cases are known to be supported or under development outside the pilot program (e.g., pharmacogenomics). Yet, others may need to be explicitly tested to foster adoption (e.g., newborn testing). Of special interest would be use cases that would enable a number of pilots to work together on a common overall workflow involving each of them. Such work would also enable pilots to synergize and enable critical mass to develop to foster a burgeoning, self-reinforcing ecosystem. Other areas for potential further development include structural variants and cytogenetics.

3. Facilitate development of a FHIR Genomics standardized code examples portfolio for each of the DAM use cases.

The DAM use cases have been closely followed by stakeholders looking to developing clinical genomics support. The use cases were also critical to the development of FHIR Genomics. However, the DAM itself does not contain any FHIR code, as it is standard-agnostic. Facilitating the development of a portfolio of standard code examples in FHIR for each use case would be a boon for those seeking to implement these use cases with FHIR. This may include development of associated terminologies to aid in use case implementations. Indeed, FHIR defines how to use terminologies and terminology services – which would be useful for certain use cases – such as HLA. Thus, further development of the DAM can aid FHIR – and FHIR implementations can help in refining the DAM.

4. Facilitate development and testing of apps and services that integrate genomic data seamlessly with clinical information.

This part facilitates the flow described in left/right arrow going into SMART-based apps in Figure 1. This may involve an EHR (e.g., EHR in combination with GACS) or in combination with a laboratory-based FHIR server. For information to be generated and used in a consistent manner, both sides of the equation (consumer and producer) need to be tested via apps/services that are created and tested for clinical genomics. SMART enables context to be loaded directly into the app, thereby enhancing workflow integration. This also reduces duplication of information entry and storage as well as potential for error.

5. Facilitate EHR integration of genomic information from laboratories via FHIR.

NGS solutions and health IT developers represent two stakeholders that evolved in different marketplaces. Each has technologies that are complementary. Facilitating integration via pilot combinations that integrate a laboratory and a health IT developer would be tremendously

useful for the field. For example, in Figure 1, the middle blue arrow going into EHR, representing the transfer of genomic data, is a GACS to handle NGS data. The EHR can represent an EHR in combination with GACS. Pilots that would help enable this framework and test standard use and scope within different platforms include: 1) pilots that test FHIR Genomics for GACS integration with EHRs and/or 2) pilots that link laboratories with cloud-based system vendors.

6. Facilitate testing apps with pilot sites to test contribution of genomic information into research data warehouses via FHIR

This could test the capturing of genomic information from NGS/sequencing-based organizations' FHIR servers and/or in combination with EHR (e.g., EHR in combination with GACS). Please see left/right arrow going into research data warehouse in Figure 1. This would help test use of FHIR for genomics patient-based contributions via PMI workflows and examine standard needs and usage for dealing with additional components including the new "Federal Policy for the Protection of Human Subjects Final Rule" and its 2018 timeline.

7. Facilitate piloting that includes a protected health information (PHI) focus via pilot testing and guidance on CLIA/HIPAA-related issues and privacy/security via SMART on FHIR.

The privacy and security of genomic data is of vital importance. These have begun to be explored by a number of the pilots. Yet, additional considerations with respect to PHI and workflow integration with CLIA/HIPAA-related guidance for pilots may be beneficial for adoption. While FHIR provides some security guidance, it stops short of specifying technologies because that is outside of FHIR's implementation. SMART on FHIR defines security and privacy protocols and practices and S4S has utilized the SMART on FHIR privacy and security practices. The PMI has provided the *Data Security Principles Implementation Guide*, *Data Security Policy Principles and Framework*, and *Privacy and Trust Principles*. Drawing on SMART on FHIR and associated technologies in the context of pilot use cases and workflows can help in guiding pilots in this arena as well.

8. Facilitate incentives for incremental adoption via the Sync for Genes levels metrics for FHIR Genomics support

Sync for Genes outlined five levels of FHIR adoption for different types of laboratories (see Figure 5). Support by each individual laboratory for each level can be done incrementally, as one moves between levels from top to bottom or from bottom to top. Work can then be done stepwise – with each level enabling new functionality. In this way, one does not to have to support the entire standard from the beginning, while at the same time not losing work when adding support for new levels.

Incentives need not be monetary. For example, incentives can include voluntary programs for self-identification via logos to signify conformance at specific levels, with such support leading to enhanced marketability of health IT products and to joining as a part of a larger ecosystem. Validation scripts for FHIR Genomics conformance were already developed as part of Sync for Genes. Involvement of government agencies or departments (e.g., CDC, ONC, CMS, FDA, NIH, DoD, VA.) to test and deploy FHIR to enable intra- and inter-agency interoperability can help incent other players to use this technology for interfacing as well.

All of Us,Sync for Science, and S4S are marks of the U.S. Department of Health and Human Services.