**Abstract: Building a FHIR-Based Data Integration Platform for Pediatric and AYA Cancer Research**

Ali Hashmi | IBM Consulting – Federal | ahashmi@us.ibm.com

**Introduction**

Pediatric and adolescent/young adult (AYA) cancer research faces critical challenges due to fragmented data silos, inconsistent standards, and labor-intensive harmonization processes. These obstacles impede the pace of scientific discovery and the delivery of precision care to young cancer patients. To address these barriers, we propose the development of an innovative, standards-driven data integration platform leveraging the Fast Healthcare Interoperability Resources (FHIR) framework. By extending oncology-specific FHIR profiles (e.g., mCODE) with pediatric- and AYA-specific elements, this platform will enable automated, real-time extraction, harmonization, and federated analysis of diverse clinical, molecular, imaging, and patient-reported data across institutions.

**Project Objectives**

* **Establish a scalable, interoperable data ecosystem** for pediatric and AYA cancer research.
* **Automate data extraction and harmonization** from EHRs and research databases using FHIR servers and APIs.
* **Enable secure, federated analytics** to support collaborative research while preserving patient privacy.
* **Accelerate discovery** of biomarkers, risk stratification, and personalized therapies through integrated data modalities.

**Development Plan**

**Phase 1: Requirements Analysis & Landscape Review**

* **Survey existing FHIR implementation guides** (e.g., mCODE, Genomics Reporting, Imaging) and identify gaps for pediatric/AYA oncology.
* **Engage stakeholders** (clinicians, informaticians, patient advocates) to define key pediatric/AYA-specific data elements.
* **Select representative sample datasets** (synthetic and/or de-identified) for initial testing.

**Phase 2: Platform Architecture & MVP Development**

* **Deploy FHIR servers** at participating sites to facilitate standards-based data exchange.
* **Extend mCODE profiles** to include pediatric/AYA-specific elements (e.g., developmental stage, long-term survivorship data).
* **Develop automated extraction APIs** for EHR, molecular, imaging, and patient-reported outcomes data.
* **Implement federated analytics tools** to enable privacy-preserving multi-site research queries.
* **Iteratively test and refine** the platform using sample datasets.

**Phase 3: Demonstration & Evaluation**

* **Conduct hands-on workshops** with sample datasets, demonstrating automated data harmonization and federated queries.
* **Compare platform capabilities** against existing solutions and implementation guides.
* **Collect feedback** from workshop participants to inform further development.

**Preliminary Results**

Includes activities expected to be conducted during the Pediatric/AYA Cancer Data Jamboree:

* **Landscape review completed:** Current FHIR oncology profiles (e.g., mCODE) lack several pediatric/AYA-relevant data elements (see Figure 1).
* **Prototype FHIR server deployed:** Successfully ingested synthetic pediatric oncology records, mapped to extended mCODE profiles.
* **Automated extraction pipeline:** Demonstrated real-time harmonization of clinical and molecular data from EHR extracts (see Figure 2).
* **Federated query proof-of-concept:** Enabled cross-site cohort identification without sharing raw patient data.

**Figure 1: Gap Analysis of mCODE vs. Pediatric/AYA Data Needs**

|  |  |  |
| --- | --- | --- |
| Data Element | mCODE Support | Pediatric/AYA Gap |
| Age at Diagnosis | Yes |  |
| Development Stage | No | Add to Profile |
| Survivorship Data | Limited | Extend Rquired |
| Genomic Vairants | Yes |  |
| Patient-Reported | Limited | Expand PROMs |

**Figure 2: Data Flow Diagram**

Other Sites

Site FHIR Server (mCode + Pediatric/AYA Extensions)

EHR/Research DBs

Federated Analytics Engine

Researcher/Clinician

Automated Extraction APIs

**Next Steps**

* **Expand pilot testing** to include real-world, multi-institutional datasets.
* **Refine and validate** pediatric/AYA-specific FHIR extensions with domain experts.
* **Develop user-friendly tools** for cohort discovery, longitudinal tracking, and data visualization.
* **Publish implementation guides** and open-source code to promote adoption and community feedback.
* **Evaluate impact** on research efficiency, data quality, and collaborative study outcomes.

**Conclusion**

This project lays the groundwork for a transformative, interoperable data ecosystem in pediatric and AYA cancer research. By leveraging and extending FHIR standards, automating data harmonization, and enabling federated analytics, we aim to accelerate the pace of discovery and the delivery of precision care. The platform will serve as a scalable model for other rare disease domains, fostering a new era of collaborative, data-driven research.

*Workshop participants will gain hands-on experience with the platform, contribute to its ongoing development, and help shape the future of pediatric and AYA cancer research.*