



# Enabling FAIR Health Data Sharing: A Methodology for Transforming Existing Data into HL7 FHIR Repositories

*Wasim Fathima Shah*

*Wellmark- Blue Cross Blue Shield, United States*

[Wasimfathima786@gmail.com](mailto:Wasimfathima786@gmail.com)

DOI: <https://doi.org/10.55248/gengpi.4.823.51564>

## ABSTRACT

Navigating the challenges of health data sharing, we introduced a novel method aligned with FAIR principles to enhance interoperability, employing the HL7 FHIR standard. Our Data Curation Tool, tested on datasets from two distinct institutions, offers users an intuitive interface for efficient data mapping into HL7 FHIR while preserving data integrity. The transformation not only ensures alignment with FAIR standards but also generates additional FHIR resources, achieving top-tier results in data Findability, Accessibility, Interoperability, and notable Reusability. This innovative approach optimizes existing health data for FAIR-compliant sharing.

Keywords: Health data sharing, FAIR principles, Interoperability, HL7 FHIR standard, Data Curation Tool, Data mapping, Data integrity, Findability, Accessibility, Reusability, Data transformation.

## 1. Introduction

In the evolving landscape of digital health, the emphasis on sharing health data is becoming increasingly pronounced due to its potential to foster research, streamline integrated care, and enhance personal health record management. However, the trajectory toward efficient health data sharing is riddled with technical, ethical, and regulatory impediments[1]. Addressing technical challenges necessitates focus on facets like data modeling, formatting, and storage. Despite strides in interoperability through Common Data Elements (CDEs) and Common Data Models (CDMs), there are still hurdles, notably the lack of unified governance and the emergence of isolated data silos[2,3,4,5,6,7,8]. Efforts are underway to create a harmonized landscape, but proprietary models and varied architectures continue to complicate the process[8-12]. In this context, the FAIR guiding principles have emerged as a beacon, emphasizing findability, accessibility, interoperability, and reusability of data[13]. Although not a standard, the FAIR framework is gaining momentum, especially in the biomedical realm[14-17]. While there have been significant accomplishments in aligning biomedical datasets with FAIR principles, patient data from electronic health records remains a nascent area of exploration[24,25]. Standards like HL7 FHIR are now being examined for their compatibility with FAIR, hinting at a future where health data standards inherently align with these principles[26,27]. HL7, with its longstanding commitment to health data interoperability, has been instrumental, especially with the wide adoption of its various versions and the Clinical Document Architecture (CDA)[28].

The collective feedback from both industry stakeholders and academic circles has been instrumental in refining health data standards. Drawing upon the insights from prior standards and leveraging existing web technologies and contemporary development strategies, the HL7 FHIR emerged as a transformative health data modeling and exchange paradigm[29]. Embracing a modular design, it encapsulates various health-related entities like patients and medications into discrete FHIR resources. In light of the inherent challenges in formulating an all-encompassing standard, FHIR introduced "profiles" which offer customization capabilities to tailor existing resources to specific requirements[30]. This flexibility extends from modifying cardinality constraints to introducing new fields, thereby offering comprehensive adaptability. With FHIR, clients can seamlessly interact with servers, recognizing their capabilities and responding in real-time.

The health IT sector has notably endorsed FHIR. Initiatives like the Argonaut Project[31] exemplify the endeavors to bolster FHIR adoption among commercial EHR providers[32]. Its prominence is further underscored by the patronage from tech giants such as Apple and Google, and leading EHR vendors including Epic and Cerner[33,34]. Regulatory bodies, like the Office of the National Coordinator for Health Information Technology, have even incorporated FHIR support as a certification prerequisite[35]. On a global scale, European nations have integrated FHIR into their eHealth blueprints, evidenced by the National Health Service in the UK and Germany's Medical Informatics Initiative[36,37]. The open-source EHR landscape, represented by entities like OpenEMR and Vista, has also welcomed FHIR, further bolstered by its academic acknowledgment and escalating scholarly references[38,39].

---

## 2. Objective

In the backdrop of FAIR principles' significance in health data dispersion and the escalating traction of HL7 FHIR, our study embarks on the journey to introduce a Data Curation Tool (DCT). This tool orbits around a novel methodology aimed at assimilating and transitioning pre-existing health datasets into HL7 FHIR repositories. Our overarching aspiration is to equip health data custodians with the tools to mold their data into the FAIR format, harnessing a recognized and prolific health standard. Our endeavor's unique facet is our innovative approach to utilizing FHIR resources to align with every FAIR criteria.

While the momentum for FHIR integration within health IT infrastructures is building, the process of adapting legacy data into FHIR remains intricate. The act of aligning historical health data with FHIR poses significant challenges, demanding both time and profound FHIR comprehension. Echoing the FAIR-centric blueprint we previously outlined[25], our solution offers a robust methodology and the DCT. This tool ensures that domain specialists can effortlessly convert health data into FHIR, aided by a user-centric graphical interface. Beyond direct data mappings, our approach ingeniously employs additional FHIR resources, ensuring full FAIR compliance upon data integration into the FHIR platform. We have subjected our methodology to rigorous scrutiny using authentic datasets from two distinct healthcare bodies, emphasizing data utility and adherence to FAIR metrics.

---

## 3. Methods

### 3.1 Collaborative Approach and Design Process

To develop the Data Curation Tool (DCT), we embraced an iterative methodology, fostering a collaborative environment with experts and researchers from diverse health institutions. Our key collaborators stemmed from five prominent organizations: (1) Andalusian Health Service (SAS; Virgen del Rocío University Hospital) in Spain, (2) Health Sciences Institute of Aragón (IACS) in Spain, (3) Geneva University Hospital in Switzerland, (4) the University of Porto in Portugal, and (5) Catholic University of the Sacred Heart in Italy[25]. Additionally, experts from various other institutions, including the authors' affiliations, enriched our collaborative sessions.

During the design phase, our primary focus was to glean insights from these experts. We initiated the design process by presenting mock-up graphical user interfaces (GUIs) to our collaborators, subsequently gathering their feedback through web-conferenced group interviews. It's noteworthy to mention that our diverse group of collaborators had varying familiarity with the FHIR standard. As depicted in Table 1, a significant proportion (10/22, 45%) were newcomers to FHIR. A smaller group (6/22, 27%) possessed rudimentary knowledge, recognizing FHIR as a pivotal health care data interoperability standard. Our collaboration was further enriched by the inclusion of FHIR experts, primarily software engineering professionals.

Simultaneously, an in-depth analysis of the existing data infrastructures, formats, and technological frameworks of our partner organizations was conducted. This dual approach ensured that our DCT was not only aligned with the FAIR-ification workflow[25] but also tailored to the specific needs and nuances of each collaborating institution.

### 3.2 FHIR Repository Integration and Configuration

The Data Curation Tool (DCT) is designed to seamlessly connect with an FHIR server end point. Upon successful validation of the FHIR end point, the DCT moves forward with the mapping configuration and data transformation. The FHIR's CapabilityStatement resource, which describes the behaviors and features of an FHIR server, is retrieved and processed by the DCT. This capability processing ensures the user has a comprehensive view of the available FHIR resource types, supported profiles, and their associated operations. The DCT's flexibility is evident as it is not restricted to any specific FHIR implementation guide or profiles. It can process machine-readable profile definitions, thereby enforcing the rules defined by these profiles during user mapping configuration. This approach ensures invalid configurations are eliminated from the onset[41].

For the purpose of our evaluation, we employed the onFHIR repository, an open-source solution accessible on GitHub[41]. onFHIR allows for the incorporation of any valid FHIR profile definition, facilitating rule enforcement. We set up onFHIR using publicly available FHIR profiles, ensuring that the DCT consistently produces valid FHIR resources aligned with the profile definitions supported by the connected FHIR repository[42].

### 3.3 Mapping and Data Transformation Process

Upon processing the FHIR repository's CapabilityStatement, the user can begin mapping configurations. Whether it's pointing to a CSV, Excel file, or connecting to a relational database, the DCT provides an intuitive platform for users to map FHIR resource types from their source data structures, ensuring alignment with profile definitions. A standout feature of the DCT is its reverse mapping design, which, based on user feedback, offers a more user-focused experience, emphasizing the requirements of the selected FHIR resource and profile[43].

The DCT user interface is designed for individuals with basic knowledge of health data entity types, as categorized by HL7 FHIR. This understanding aids users in selecting appropriate resource types for data tables or sheets to initiate the mapping process. Additionally, users need a foundational grasp of code systems, value sets, and coded values within their datasets. However, intricate FHIR-related details are abstracted away, with the DCT offering field names, descriptions, data types, and value set restrictions. Through this, users can accurately map source fields based on presented descriptions and constraints.

Behind the scenes, as mappings are configured through the graphical interface, the DCT generates a JSON-based mapping configuration that adheres to our proprietary internal mapping language. This decision to design an internal mapping language, as opposed to utilizing the FHIR Mapping Language[43], stemmed from our methodology's unique requirements. Our approach ensures reduced complexity in both development and configuration. Notably, the DCT also handles referential integrity between FHIR resource instances, provided the source data maintains unique identification across records. This automated referential integrity reflects in the generated FHIR resources, streamlining the entire data transformation process.

### ***3.4 Terminology Transformation and Translation***

In the realm of data transformation, the DCT not only addresses the syntactic transformation, which involves matching FHIR elements with source data columns, but also the semantic transformation, leveraging a FHIR-based terminology server. At the initiation of the data curation process, users have the option to connect to a terminology server, in addition to the FHIR end point of the repository. The inclusion of a terminology server is optional, dependent on whether the transformation requires terminology translation based on the code systems used in the source datasets and the code systems specified by the target FHIR profiles. The DCT employs FHIR resource types such as CodeSystem, ValueSet, and ConceptMap to handle code system mappings during configuration and terminology translations during data validation and transformation.

Once an FHIR element is syntactically matched with a source data field, users can further configure the terminology translation for that matching, if necessary. Some data fields in the source data may utilize proprietary or custom local code lists, while others might employ standard code systems. A common example is the gender field in the Patient resource. Many datasets employ their own custom codes to denote a patient's gender. If the target gender element under the Patient resource type mandates a different code system, like the HL7 administrative gender codes, then our methodology necessitates the mapping from custom codes to the target code system to be integrated and provided by the connected FHIR terminology server. In scenarios where the target code system is predetermined by the FHIR profile definition, the DCT ensures the adherence to specific standards, such as the International Classification of Diseases, 10th Revision (ICD-10) for the Condition resources. In cases where the source data employs the same standard like ICD-10, no additional configuration is required. However, if there's a deviation, mappings can be set up from local custom codes, Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT), or Medical Dictionary for Regulatory Activities (MedDRA) to ICD-10, or even between different versions of the International Classification of Diseases, provided the connected FHIR terminology server offers these code mappings[25].

### ***3.5 Data Integrity and Transformation Process***

Ensuring the integrity of FHIR resources requires attention to both mapping configurations and the inherent qualities of the source data. When configuring mappings, users can preview a subset of the data (e.g., initial rows) and test the created FHIR resources' compliance. The DCT then evaluates the resources against the chosen FHIR profile definitions, verifying aspects such as the presence of mandatory elements, compliance with cardinality restrictions, and the appropriate setup of value set enforcements and terminology translations.

After finalizing the mappings, a comprehensive validation process is initiated. In this phase, the DCT produces FHIR resources from the entirety of the data, validates them before any commitment to the FHIR repository, and leverages the FHIR standard validation endpoint. Users can then refine their mappings if issues arise, ensuring that only valid resources are stored in the repository. This meticulous validation, albeit potentially lengthy, guarantees the congruence of the mapping configuration with the source data's nature. Once this validation is complete, the data transformation is finalized by committing the validated data to the FHIR repository.

### ***3.6 Evaluation and Implementation***

Our methodology and the DCT underwent rigorous evaluation in terms of data transformation quality and adherence to FAIR principles. We gauged the utility of the transformed data and assessed its FAIR compliance. This assessment took place using health data sets from two distinct Spanish institutions: SAS and IACS. While both institutions operate within the same country and thus have some overlapping code systems and unstructured text fields in Spanish, they differ fundamentally in their functions and data types. SAS operates as a healthcare organization, while IACS is primarily a health research institute. Consequently, their source data structures vary both syntactically and semantically.

For each institution, a subject-matter expert, familiar with the source data's content model, undertook the data transformation using the DCT. Prior to the transformation, we equipped these experts with documentation, video guides, and training on the DCT and FAIR-ification procedures via online sessions. Although these sessions didn't focus on FHIR training explicitly, foundational concepts like different resource types, profile definitions, and code systems were addressed.

During the evaluation, various FHIR resource types, including Patient, Encounter, Observation, Condition, and MedicationStatement, were utilized to configure the mappings, in line with the requirements dictated by the profiles sourced from the FHIR repository. Additionally, the DCT autonomously generated Provenance and DocumentManifest resources as part of its backend operations[25].

### ***3.7 Methodological Evaluation of the DCT***

A comprehensive end-user assessment of the DCT was undertaken in a separate study where 16 domain experts from different European organizations appraised the tool's usability through a set of questionnaires[43]. These experts scrutinized the tool's GUI functionalities in relation to its stipulated

requirements. The feedback from this evaluation was overwhelmingly positive and indicated the tool's proficiency. In contrast, our study's primary focus was on evaluating the proposed methodology in the dimensions of data utility and adherence to FAIR principles.

### **3.8 Ethical Considerations**

Prior to the study's commencement, ethical clearance was procured from the involved health research entities, specifically, Virgen del Rocío University Hospital under the umbrella of SAS and IACS, and Aragón Health Research Institute [1269-M1-20], in accordance with local regulations. A series of protective measures were instituted to ensure participants' rights and freedoms, which included adherence to the principle of data minimization, acquiring informed consent, and disseminating informational materials. Additionally, each institution designated a data protection officer. To bolster the ethical framework, an independent ethics advisory board was constituted at the study's outset. This board was tasked with reviewing study outputs, producing reports, and delivering presentations to aid the FAIR4Health Consortium.

### **3.9 Results and Data Utility Assessment**

The concept of data utility pertains to the analytical accuracy and validity of the transformed data. To gauge data utility, an initial step involved contrasting the aggregate count of entities in the source data against those in the resultant FHIR resources. Additionally, to delve deeper into the integrity of the data transformation at the granular level of individual fields, we defined five specific criteria for data utility. These criteria, which we term as "privacy-concerned", were chosen to offer a robust assessment of data utility without infringing on the data privacy norms set by the participating organizations. This approach, endorsed by the data protection officers of the respective institutions, aimed to strike a balance between rigorous evaluation and privacy adherence. For a more precise assessment of our methodology's data utility, we disregarded any erroneous or incomplete data from the source datasets. Our methodology was designed to exclude records that failed validation from the FHIR repository, and any such exclusions were flagged to the user with an accompanying explanation, such as missing mandatory fields or unparseable numerical values.

### **3.10 FAIRness Assessment**

To gauge the FAIRness of the transformed datasets, we utilized the maturity indicators and assessment techniques from the FAIR Data Maturity Model developed by the Research Data Alliance (RDA) [44]. This model delineates a series of indicators for each FAIR principle, each bearing a unique identifier. These indicators are stratified into three categories based on their significance: essential, important, and useful. We conducted a comprehensive assessment of the transformed data against each of these indicators, registering scores for each.

In the scoring process, we took into account the innate functionalities of HL7 FHIR and the manner in which we employed FHIR resources and profiles via the DCT. Our methodology inherently satisfies certain indicators, while others are addressed indirectly through the standard and its execution. For instance, the aspect of access authentication (under the accessibility principle) is an illustration of an indicator catered to indirectly by the FHIR standard and its practical application. Conversely, our methodology ensures the resolvability and accessibility of both data and metadata, and guarantees the persistence of metadata even if the corresponding data becomes obsolete (as per the accessibility principle).

It's pivotal to note that the maturity model distinctly demarcates data and metadata when formulating the indicators. While a rigid delineation between data and metadata is often elusive, our design of the DCT views FHIR's content modeling constructs, such as profiles or the CapabilityStatement, as metadata. This is in conjunction with the inherent metadata constructs like the meta-element of the foundational FHIR resource (named "Resource") or resources like Provenance and DocumentManifest. These resources are auto-generated by the DCT during the data transformation process.

---

## **4. Discussion and Implications**

### **4.1 Key Outcomes**

This research emphasized the efficacy of the methodology and the Data Curation Tool (DCT) we developed in facilitating the FAIR-ification of real-world health data. The translation of data into HL7 FHIR, in compliance with FAIR principles for health data sharing, was made feasible through the DCT. This tool empowers domain experts to convert existing health data into FHIR through intuitive graphical interfaces. Evaluation results highlighted that the integrity and utility of the transformed data remain intact, achieving the highest FAIRness levels for the Findable, Accessible, and Interoperable principles and a commendable level for the Reusable principle.

In light of the growing embrace of HL7 FHIR as a health data interchange standard, numerous endeavors have been launched to convert data from older formats to FHIR. Our approach offers several benefits over existing efforts: Clinical Asset Mapping Program for FHIR (CAMP FHIR): CAMP FHIR is an open-source tool designed to convert Clinical Data Models (CDMs) to a specific set of FHIR resources tailored for clinical research data [8]. While CAMP FHIR employs SQL scripts and a command-line interface, the DCT offers a visual interface for mapping and is adaptable to various data formats, not just relational databases. FHIR Converter: Developed by Microsoft, the FHIR Converter can only process two types of data, HL7 version 2 and CDA, to FHIR resources via predetermined mappings [45]. In contrast, the DCT offers broader support for diverse data sources. Firely FHIR Facade: A paid plug-in, Firely FHIR Facade allows for data transformation but lacks a graphical interface for defining mappings, making it less accessible for non-technical users [46].

## 4.2 Study Limitations

The study does acknowledge certain limitations. Given the inherent inconsistencies and gaps in healthcare data, achieving harmonization can be challenging. Our approach allows flexibility in adopting FHIR profiles, which can be tailored to handle data inaccuracies and incompleteness. Moreover, while our design and development phases incorporated feedback from five different healthcare and research organizations, the evaluation was conducted using data from only two institutions, thus limiting the scope of the study.

Additionally, the study did not encompass data procured from patient telemedicine devices or similar sources. While our approach met the transformation requirements of two diverse health organizations, there's room for further enhancement to cater to intricate mapping needs of varied health research entities. The current DCT, being a standalone desktop application, might face challenges in handling large datasets. However, by employing distributed data processing techniques, the methodology can be adapted to manage substantial volumes of health data. Future endeavors will focus on refining our internal mapping language and addressing complex data operations, as well as broadening our evaluations to encompass more institutions from different regions.

## 4.3 Broader Implications

The DCT, coupled with our FHIR utilization, aids in the transition of institutions to HL7 FHIR. This not only boosts FAIRness levels but also smoothens the adoption of SMART on FHIR applications [47]. This paves the way for healthcare institutions to seamlessly integrate a plethora of third-party applications such as decision support tools, risk estimators, and graphical representations [48]. Moreover, our approach and DCT foster integration with clinical research models, promoting the secondary use of healthcare data.

## 4.4 Integration with BRIDG Model

Our methodology promises an effortless integration with the Biomedical Research Integrated Domain Group (BRIDG) Model, a collaborative effort by esteemed institutions like the Food and Drug Administration, National Cancer Institute, Clinical Data Interchange Standards Consortium, and HL7. The BRIDG Model seeks to enable protocol-driven research by offering a unified perspective on the semantics across diverse research domains, ranging from basic and preclinical research to clinical, translational, and related regulatory artifacts. Leveraging our approach, institutions can conveniently map their existing Electronic Health Record (EHR) data to the BRIDG Model by setting the BRIDG FHIR resources [51] as the target FHIR model.

## 4.5 Aligning with HL7's Vision

Further emphasizing the potential of our approach, it aligns well with the vision statement of the HL7 Vulcan Accelerator Project [52]. This project envisions an integrated landscape where research seamlessly blends into healthcare delivery. The objective is to simplify data collection and exchange, with HL7 FHIR serving as the backbone for data interchange. Furthermore, a collaboration between the Observational Health Data Sciences and Informatics and HL7 has been announced, aiming to bolster the exchange of information across clinical care and observational research domains [53,54]. The collaboration focuses on creating mappings between the Observational Medical Outcomes Partnership (OMOP) model and HL7 FHIR. Herein, our Data Curation Tool (DCT) emerges as a crucial facilitator, streamlining the mapping process between the OMOP model and custom-designed HL7 FHIR profiles tailored for clinical research studies.

---

## 5. Discussion and Implications

### 5.1 Key Outcomes

This research emphasized the efficacy of the methodology and the Data Curation Tool (DCT) we developed in facilitating the FAIR-ification of real-world health data. The translation of data into HL7 FHIR, in compliance with FAIR principles for health data sharing, was made feasible through the DCT. This tool empowers domain experts to convert existing health data into FHIR through intuitive graphical interfaces. Evaluation results highlighted that the integrity and utility of the transformed data remain intact, achieving the highest FAIRness levels for the Findable, Accessible, and Interoperable principles and a commendable level for the Reusable principle.

In light of the growing embrace of HL7 FHIR as a health data interchange standard, numerous endeavors have been launched to convert data from older formats to FHIR. Our approach offers several benefits over existing efforts:

**Clinical Asset Mapping Program for FHIR (CAMP FHIR):** CAMP FHIR is an open-source tool designed to convert Clinical Data Models (CDMs) to a specific set of FHIR resources tailored for clinical research data [8]. While CAMP FHIR employs SQL scripts and a command-line interface, the DCT offers a visual interface for mapping and is adaptable to various data formats, not just relational databases. **FHIR Converter:** Developed by Microsoft, the FHIR Converter can only process two types of data, HL7 version 2 and CDA, to FHIR resources via predetermined mappings [45]. In contrast, the DCT offers broader support for diverse data sources. **Firely FHIR Facade:** A paid plug-in, Firely FHIR Facade allows for data transformation but lacks a graphical interface for defining mappings, making it less accessible for non-technical users [46].

## 5.2 Study Limitations

The study does acknowledge certain limitations. Given the inherent inconsistencies and gaps in healthcare data, achieving harmonization can be challenging. Our approach allows flexibility in adopting FHIR profiles, which can be tailored to handle data inaccuracies and incompleteness. Moreover, while our design and development phases incorporated feedback from five different healthcare and research organizations, the evaluation was conducted using data from only two institutions, thus limiting the scope of the study.

Additionally, the study did not encompass data procured from patient telemedicine devices or similar sources. While our approach met the transformation requirements of two diverse health organizations, there's room for further enhancement to cater to intricate mapping needs of varied health research entities. The current DCT, being a standalone desktop application, might face challenges in handling large datasets. However, by employing distributed data processing techniques, the methodology can be adapted to manage substantial volumes of health data. Future endeavors will focus on refining our internal mapping language and addressing complex data operations, as well as broadening our evaluations to encompass more institutions from different regions.

## 5.3 Broader Implications

The DCT, coupled with our FHIR utilization, aids in the transition of institutions to HL7 FHIR. This not only boosts FAIRness levels but also smoothens the adoption of SMART on FHIR applications [47]. This paves the way for healthcare institutions to seamlessly integrate a plethora of third-party applications such as decision support tools, risk estimators, and graphical representations [48]. Moreover, our approach and DCT foster integration with clinical research models, promoting the secondary use of healthcare data.

Discussion and Implications

## 5.5 Integration with BRIDG Model

Our methodology promises an effortless integration with the Biomedical Research Integrated Domain Group (BRIDG) Model, a collaborative effort by esteemed institutions like the Food and Drug Administration, National Cancer Institute, Clinical Data Interchange Standards Consortium, and HL7. The BRIDG Model seeks to enable protocol-driven research by offering a unified perspective on the semantics across diverse research domains, ranging from basic and preclinical research to clinical, translational, and related regulatory artifacts. Leveraging our approach, institutions can conveniently map their existing Electronic Health Record (EHR) data to the BRIDG Model by setting the BRIDG FHIR resources [51] as the target FHIR model.

## 5.4 Aligning with HL7's Vision

Further emphasizing the potential of our approach, it aligns well with the vision statement of the HL7 Vulcan Accelerator Project [52]. This project envisions an integrated landscape where research seamlessly blends into healthcare delivery. The objective is to simplify data collection and exchange, with HL7 FHIR serving as the backbone for data interchange. Furthermore, a collaboration between the Observational Health Data Sciences and Informatics and HL7 has been announced, aiming to bolster the exchange of information across clinical care and observational research domains [53,54]. The collaboration focuses on creating mappings between the Observational Medical Outcomes Partnership (OMOP) model and HL7 FHIR. Herein, our Data Curation Tool (DCT) emerges as a crucial facilitator, streamlining the mapping process between the OMOP model and custom-designed HL7 FHIR profiles tailored for clinical research studies.

---

## 6. Concluding Remarks

In this research, we introduced a holistic approach and a bespoke DCT to harness the potential of existing health data, which often lies fragmented across multiple silos. Our goal was to ensure these data sets are readily shareable as FHIR resources, adhering to the principles of FAIR (Findable, Accessible, Interoperable, and Reusable). Our iterative, agile methodology, encompassing everything from requirement analysis to GUI development and software design, ensured the DCT evolved based on end-user feedback.

Our comprehensive evaluation utilized real-world data from two distinctive Spanish entities, SAS (a hospital) and IACS (a health research institute). By employing privacy-aware evaluation metrics, we assessed the transformed data's utility and accuracy. The results underscored the DCT's capability to adeptly convert existing health data into HL7 FHIR without compromising on data utility. Moreover, our FAIRness assessment, grounded in the FAIR Data Maturity Model by the Research Data Alliance (RDA), revealed exemplary FAIRness levels across the Findable, Accessible, and Interoperable principles. For the Reusable principle, we achieved a commendable level 3, addressing the essential and important FAIR metrics. In essence, our findings highlight the DCT's proficiency in ensuring health data is both FAIR and seamlessly transformed into HL7 FHIR. (1)

## References

- 
- [1] Vayena E. Value from health data: European opportunity to catalyse progress in digital health. *Lancet*. 2021 Feb 20;397(10275):652-653. doi: 10.1016/S0140-6736(21)00203-8. Medline: 33571452.

- [2] Kush RD, Warzel D, Kush MA, Sherman A, Navarro EA, Fitzmartin R, et al. FAIR data sharing: the roles of common data elements and harmonization. *J Biomed Inform.* 2020 Jul;107:103421. doi: 10.1016/j.jbi.2020.103421. Medline: 32407878.
- [3] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform.* 2015;216:574-578. Medline: 26262116.
- [4] Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc.* 2012;19(1):54-60. doi: 10.1136/amiajnl-2011-000376. Medline: 22037893.
- [5] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc.* 2010;17(2):124-130. doi: 10.1136/jamia.2009.000893. Medline: 20190053.
- [6] Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc.* 2014;21(4):578-582. doi: 10.1136/amiajnl-2014-002747. Medline: 24821743.
- [7] Sentinel Common Data Model. Sentinel. URL: <https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model> [accessed 2022-12-30].
- [8] Pfaff ER, Champion J, Bradford RL, Clark M, Xu H, Fecho K, et al. Fast healthcare interoperability resources (FHIR) as a meta model to integrate common data models: development of a tool and quantitative validation study. *JMIR Med Inform.* 2019 Oct 16;7(4):e15199. doi: 10.2196/15199. Medline: 31621639.
- [9] Klann JG, Phillips LC, Herrick C, Joss MA, Waghlikar KB, Murphy SN. Web services for data warehouses: OMOP and PCORnet on i2b2. *J Am Med Inform Assoc.* 2018 Oct 01;25(10):1331-1338. doi: 10.1093/jamia/ocy093. Medline: 30085008.
- [10] Klann JG, Joss MA, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: transforming i2b2 data into the OMOP common data model. *PLoS One.* 2019 Feb 19;14(2):e0212463. doi: 10.1371/journal.pone.0212463. Medline: 30779778.
- [11] Klann JG, Abend A, Raghavan VA, Mandl KD, Murphy SN. Data interchange using i2b2. *J Am Med Inform Assoc.* 2016 Sep;23(5):909-915. doi: 10.1093/jamia/ocv188. Medline: 26911824.
- [12] Reimer AP, Milinovich A. Using UMLS for electronic health data standardization and database design. *J Am Med Inform Assoc.* 2020 Oct 01;27(10):1520-1528. doi: 10.1093/jamia/ocaa176. Medline: 32940707.
- [13] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016 Mar 15;3:160018. doi: 10.1038/sdata.2016.18. Medline: 26978244.
- [14] New Models of Data Stewardship. National Institutes of Health. URL: <https://commonfund.nih.gov/data> [accessed 2022-12-30].
- [15] Big Data to Knowledge. National Institutes of Health. URL: <https://commonfund.nih.gov/bd2k> [accessed 2022-12-30].
- [16] H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. European Commission, Directorate-General for Research & Innovation. 2020. URL: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf) [accessed 2022-12-30].
- [17] Mons B, Neylon C, Velterop J, Dumontier M, da Silva Santos LO, Wilkinson MD. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Inf Serv Use.* 2017 Mar 07;37(1):49-56. doi: 10.3233/isu-170824.
- [18] van Reisen M, Stokmans M, Basajja M, Ong'ayo AO, Kirkpatrick C, Mons B. Towards the tipping point for FAIR implementation. *Data Intell.* 2020 Jan;2(1-2):264-275. doi: 10.1162/dint\_a\_00049.
- [19] Holub P, Kohlmayer F, Prasser F, Mayrhofer MT, Schlünder I, Martin GM, et al. Enhancing reuse of data and biological material in medical research: from FAIR to FAIR-health. *Biopreserv Biobank.* 2018 Apr;16(2):97-105. doi: 10.1089/bio.2017.0110. Medline: 29359962.
- [20] Wilkinson MD, Verborgh R, da Silva Santos LO, Clark T, Swertz MA, Kelpin FD, et al. Interoperability and FAIRness through a novel combination of web technologies. *PeerJ Comput Sci.* 2017 Apr 24;3:e110. doi: 10.7717/peerj-cs.110.
- [21] Wilkinson MD, Dumontier M, Sansone SA, Bonino da Silva Santos LO, Prieto M, Batista D, et al. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci Data.* 2019 Sep 20;6(1):174. doi: 10.1038/s41597-019-0184-5. Medline: 31541130.
- [22] FAIR Data Maturity Model Work Group. Research Data Alliance (RDA). URL: <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg> [accessed 2022-12-30].
- [23] FAIRplus Project. FAIRplus. URL: <https://fairplus-project.eu/> [accessed 2022-12-30].
- [24] Alvarez-Romero C, Martínez-García A, Sinaci AA, Gencturk M, Méndez E, Hernández-Pérez T, et al. FAIR4Health: findable, accessible, interoperable and reusable data to foster health research. *Open Res Europe.* 2022 May 31;2:34. doi: 10.12688/openreseurope.14349.2.

- [25] Sinaci AA, Núñez-Benjumea FJ, Gencturk M, Jauer ML, Deserno T, Chronaki C, et al. From raw data to FAIR data: the FAIRification workflow for health research. *Methods Inf Med*. 2020 Jun;59(S 01):e21-e32. doi: 10.1055/s-0040-1713684. Medline: 32620019.
- [26] FAIRness for FHIR project. Health Level 7 (HL7) International. URL: <https://confluence.hl7.org/pages/viewpage.action?pageId=91991234> [accessed 2022-12-30].
- [27] Frexia F, Mascia C, Lianas L, Delussu G, Sulis A, Meloni V, et al. openEHR is FAIR-enabling by design. *Stud Health Technol Inform*. 2021 May 27;281:113-117. doi: 10.3233/SHTI210131. Medline: 34042716.
- [28] HL7 Standards. HL7 International. URL: <http://www.hl7.org/implement/standards/index.cfm?ref=nav> [accessed 2022-12-30].
- [29] Bender D, Sartipi K. HL7 FHIR: an Agile and RESTful approach to healthcare information exchange. In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. 2013 Presented at: CBMS '13; June 20-22, 2013; Porto, Portugal p. 326-331. doi: 10.1109/cbms.2013.6627810.
- [30] Profiling FHIR. HL7 FHIR. URL: <https://www.hl7.org/fhir/profiling.html> [accessed 2022-12-30].
- [31] Precious J. The Argonaut Project. Confluence. 2022 Jan 10. URL: <https://confluence.hl7.org/display/AP/Argonaut+Project+Home> [accessed 2022-12-30].
- [32] HL7 FHIR Accelerator™ Program. HL7 International. URL: <https://www.hl7.org/about/fhir-accelerator/> [accessed 2022-12-30].
- [33] FHIR Fact Sheet. The Office of the National Coordinator for Health Information Technology (ONC). URL: <https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf> [accessed 2022-12-30].
- [34] Posnack S, Barker W. Heat Wave: The U.S. is Poised to Catch FHIR in 2019. *HealthITbuzz*. 2018 Oct 1. URL: <https://www.healthit.gov/buzz-blog/interoperability/heat-wave-the-u-s-is-poised-to-catch-fhir-in-2019> [accessed 2022-12-30].
- [35] 2015 Edition Cures Update Criteria. The Office of the National Coordinator for Health Information Technology (ONC). URL: <https://www.healthit.gov/topic/certification-ehrs/2015-edition-cures-update-test-method> [accessed 2022-12-30].
- [36] FHIR APIs. National Health Service (NHS). URL: <https://digital.nhs.uk/services/fhir-apis> [accessed 2022-12-30].
- [37] MII Core Data Set with HL7 FHIR. Medical Informatics Initiative Germany (MII). URL: <https://www.medizininformatik-initiative.de/en/medizininformatik-initiative-bundesweit-einheitliche-auswertbarkeit-von-versorgungsdaten> [accessed 2022-12-30].
- [38] Purkayastha S, Allam R, Maity P, Gichoya JW. Comparison of open-source electronic health record systems based on functional and user performance criteria. *Health Inform Res*. 2019 Apr;25(2):89-98. doi: 10.4258/hir.2019.25.2.89. Medline: 31131143.
- [39] Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The use of FHIR in digital health - a review of the scientific literature. *Stud Health Technol Inform*. 2019 Sep 03;267:52-58. doi: 10.3233/SHTI190805. Medline: 31483254.
- [40] FAIR4Health Data Curation and Validation Tool. GitHub. 2022 Sep 2. URL: <https://github.com/fair4health/data-curation-tool> [accessed 2022-12-30].
- [41] onFHIR FHIR Repository. GitHub. 2022 Dec 1. URL: <https://github.com/srdc/onfhir> [accessed 2022-12-30].
- [42] FAIR4Health Common Data Model. GitHub. 2021 May 17. URL: <https://github.com/fair4health/common-data-model> [accessed 2022-12-30].
- [43] FHIR Mapping Language. HL7 FHIR. URL: <https://hl7.org/fhir/mapping-language.html> [accessed 2022-12-30].
- [44] FAIR Data Maturity Model Work Group. Research Data Alliance (RDA). URL: <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg> [accessed 2022-12-30].
- [45] FHIR Converter. GitHub. 2022 Nov 25. URL: <https://github.com/microsoft/FHIR-Converter> [accessed 2022-12-30].
- [46] FHIR Facade. Firely. URL: <https://fire.ly/products/fhir-facade/> [accessed 2022-12-30].
- [47] SMART on FHIR. URL: <https://smarthealthit.org/> [accessed 2022-12-30].
- [48] Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc*. 2016 Sep;23(5):899-908. doi: 10.1093/jamia/ocv189. Medline: 26911829.
- [49] Biomedical Research Integrated Domain Group (BRIDG) Model. National Cancer Institute. URL: <https://bridgmodel.nci.nih.gov/about-bridg> [accessed 2022-12-30].
- [50] Becnel LB, Hastak S, Ver Hoef W, Milius RP, Slack M, Wold D, et al. BRIDG: a domain information model for translational and clinical protocol-driven research. *J Am Med Inform Assoc*. 2017 Sep 01;24(5):882-890. doi: 10.1093/jamia/ocx004. Medline: 28339791.
- [51] Hastak S. Biomedical Research and Regulation FHIR Resources. Confluence. 2023 Jan 9. URL: <https://confluence.hl7.org/pages/viewpage.action?pageId=97476291> [accessed 2022-12-30].



- 
- [52] Precious J. Vulcan Accelerator Home. Confluence. URL: <https://confluence.hl7.org/display/VA/Vulcan+Accelerator+Home> [accessed 2022-12-30].
- [53] HL7 International and OHDSI Announce Collaboration to Provide Single Common Data Model for Sharing Information in Clinical Care and Observational Research. Observational Health Data Sciences and Informatics. URL: <https://www.ohdsi.org/ohdsi-hl7-collaboration/> [accessed 2022-12-30].
- [54] Precious J. FHIR-OMOP Initiative - Homepage. Confluence. 2022 Sep 8. URL: <https://confluence.hl7.org/display/OOF/FHIR-OMOP+Initiative+-+Homepage> [accessed 2022-12-30].