A Data Modeling Process for Achieving Interoperability

Eleftheria Kouremenou[1], Athanasios Kiourtis[1], Dimosthenis Kyriazis[1]
Affiliation 1: University of Piraeus, Department of Digital Systems,
Piraeus, Greece, {elkour, kiourtis, dimos}@unipi.gr

***Abstract*—The ability of two or more systems to communicate with each other is known as interoperability. Semantic and syntactic are the two categories used in each Big Data sector. Some of the challenges of interoperability are compatibility problems, security issues, and the requirement for global standards. Specifically, there were issues with data management and exchange for 23 out of 100 health entities that took part in a survey, due to staff hesitation about technological upgrades. Interoperability considering the use of standards, such as the Fast Healthcare Interoperability Resources (FHIR) standard in healthcare, is the most effective and common technique. Independently of databases, any standard specifies how data can be transferred among them in a coherent and homogeneous way. In this paper, through an analysis of the literature, it is suggested a generic method that can be domain agnostic, for designing a data exchange and mapping model that can be understood by anyone without the need for technical knowledge, enabling interoperability through data modeling.**

Keywords*— *Big data processing; interoperability; machine learning; data modeling

I. Introduction

Data interoperability involves the utilization of standardized data formats that facilitate the integration and accessibility of diverse datasets. This concept corresponds to the letter "I" in the FAIR (Findability, Accessibility, Interoperability, and Reusability) data principles concepts [1]. Interoperability is the flexibility of systems and services that generate, exchange, and study data to share well-defined standards for the latter's content, context, and semantics (vocabulary). Moreover, it ensures that data can be exchanged effectively across various systems. It is necessary to standardize data formats, vocabularies, and languages to achieve interoperability. For example, suppose a healthcare practitioner from one establishment needs to access the "age" of a patient from a separate institution; it becomes essential that this data is uniformly communicated over the years. In the practical realm, interoperability deals with a crowd of challenges, surrounding compatibility difficulties, security threats, and an unequivocal need for universal standards.

Within the domain of healthcare, the fluid exchange of data emerges as a requirement for the realization of proficient patient care and efficacious disease control [2]. A research endeavor, transpiring between February and March of 2022 [3], revealed that a notable 23% of healthcare administrators determine the limited technological infrastructure, coupled with hesitations towards technology adoption, as a major barrier to data utilization. The leader framework that overcomes this obstacle is the Health Level 7 (HL7) Fast Healthcare Interoperability Resources (FHIR) standard as it refers to the importance of interoperability in healthcare systems [4]. This paper analyzes the HL7 FHIR architecture, focusing on FHIR profiles and related data models. Hence, it is increasingly critical for a rising number of healthcare organizations and research establishments to embrace FHIR, a nascent standard that streamlines the data exchange process amongst healthcare providers, patients, and other relevant stakeholders. FHIR demonstrates significant advantages, but its adoption has been inconsistent, highlighting the need for additional research to unlock its potential. Recent studies have highlighted both the benefits and challenges of implementing FHIR in healthcare settings, indicating that despite the standard's promise, further work is necessary to overcome extant obstacles. In this context, focusing specifically on the FHIR protocol, this paper undertakes an in-depth analysis of the benefits, technologies, and relevant research associated with interoperability in Electronic Health Records (EHRs). Thus, it is included a step-by-step guide on how one can build a FHIR model, highlighting the most important elements of a structure of a relevant model and resource accordingly. A hypothetical example is used as a use case to demonstrate how every process or can be mapped and presented by FHIR models and resources.

The rest of this paper is structured as follows. Section II includes the HL7 FHIR description, Section III provides the relative research, section IV proposes the methodology, and Section V includes our concluding remarks.

II. The Case of HL7 FHIR Standard

HL7 stands as a notable framework in the domain of EHRs, serving as a conduit for efficient data transfer. It encapsulates specific rules for semantic and syntactic interoperability, thereby ensuring a consistent vocabulary and variable structure across the EHR landscape. In 2019 [5], 83% of hospitals and 58% of clinicians used certified Application Programming Interface (API) technology from a market leader. This

979-8-3503-2887-5/23/$31.00 ©2023 IEEE

represented 99% and 95%, respectively, of all FHIR-enabled technology deployments. The findings show that health Information Technology (IT) market leaders who all support FHIR drove much of the adoption of FHIR-enabled certified API technology in 2019 [5]. HL7 International, a non-profit, ANSI-accredited standards-developing organization, is striving to craft standards and frameworks for smooth data exchange, integration, and retrieval in the healthcare field. It is under this umbrella that FHIR came into existence in 2011. Delving into the semantic aspect, HL7 includes a distinct set of variables within the code concept class, aligning with the international classifications of disease protocols, namely ICD-9 and ICD-10 [3]. These protocols assign specific alphanumeric values to each disease portraying, for instance, diabetes as FHIR-code noted c150. This semantic interoperability [6] ensures a standardized representation of diseases, easing data interchange and understanding, across diverse healthcare systems. On the flip side, syntactic interoperability within HL7 [7] is concerned with the structural aspect, proposing a specific shape for JSON imaging variables. It aims to provide a standardized structure based on common entities, interested members, procedures, and concepts prevalent in most health systems. Among the ubiquitous entities are patients - encapsulating demographic data and essential file details like birthdate, sex, name, identity, and conditions - embodying the diseases a patient has encountered with attributes like condition identity. FHIR builds upon its predecessors (HL7 versions 2 and 3, and Clinical Document Architecture), merging their strengths with modern web technologies like Representational State Transfer (REST) architecture, APIs, XML, and JSON formats, along with authorization tools like Open Authorization [8]. A FHIR data model acts as a compass in the wilderness of APIs, enabling smooth data exchanges among applications. It is a blueprint, elucidating how different entities interact with each other. The FHIR data model is depicted graphically as a network of boxes and lines, portraying terms including a 'patient' and related 'conditions', as well as the relationships between them. The essence of FHIR lies in its resources, which are fundamental building blocks defining the content and structure of exchangeable information. These resources, characterized by structured definitions of canonical Uniform Resource Locator (URL), which serves as unique identifier, can cross-reference each other, facilitating a cohesive network of data. Examples of resources within FHIR include patients, prescriptions, observations, or diagnostic reports, each symbolizing a different entity in healthcare. The resource-oriented architecture of FHIR lends itself to flexibility, scalability, and ease of implementation, embodying a novel healthcare standard that encapsulates the strong points of previous HL7 standards while bridging the gaps therein. FHIR accelerates and optimizes communication across systems, offering a platform for interoperability in health research.

III. Related Studies

The research in [9] investigates Taiwan's transition to a FHIR-based Personal Health Record (PHR), leveraging the "My Health Bank" (MHB) PHR from Taiwan National Health Insurance. The study notes the challenges that many hospitals face with Clinical Document Architecture (CDA). Tools such as ASP.NET, XPath, JSPath, and the HL7 Application Programming Interface (HAPI) server were utilized for data verification, and for converting and publishing MHB records safely to the FHIR format. With protective measures including NTP, SNTP, TLS 1.2, and OAuth 2.0. Another research is the work in [10] that presents an adept solution for the storage and transfer of semantic and structured health data by merging the FHIR protocol with SNOMED-CT vocabularies. The authors' idea, compressed in a health coaching application (namely eCoach), mirrors this concept efficiently. It is a representation of PHRs functionalities facilitating seamless two-way EHR interactions. Another research is [11] that proposes the integration of genomic data with EHRs, covering the path for personalized healthcare hesitation. With the FHIR Genomics Operations, the authors have standardized molecular data handling, focusing on precision medicine applications. These operations enhance genomics data representation, cater to significant raw data in non-FHIR formats, and address numerous genomic facets, from variation discovery to pharmacogenomic testing. In [12] an API developed by the Office of the National Coordinator, offers straightforward access to patient records within approved health IT infrastructures. Comprising the SMART/HL7 FHIR Bulk Data Access API, this research complies with legal terms and information-blocking guidelines, enhancing population health management and fostering extensive, data-driven innovation. The research in [13] showcases Duke Health's integration of SMART on FHIR tech into their EHR system. The authors highlight the significance of security, careful mobile application integration, and the need for more applications that harvest genetic data. Another study [14] has analyzed the mapping process for converting various data in FHIR format. The methodology also includes an ontology transformation into various healthcare data sets, followed by storing the results in a knowledge base as triple information. The researchers in [15] stated that successful interoperability between health systems can upgrade early diagnosis and the course of disease. In addition, it was pointed out how important is the right design and the right choice of entities and resources of a system.

IV. Proposed Methodology

Outlined below (Fig. 1) is a thorough approach to developing an interoperable model within any domain – emphasis is given to the healthcare domain for better explanation purposes, integrating stakeholders like data scientists, researchers, workers, and administrative staff.

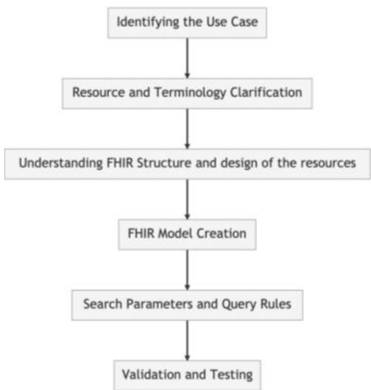

Fig. 1. Overall workflow of the data modeling process

A. *Identifying the Use Case*

The pragmatic issue that are aimed to be addressed should be found. Whether located in a research environment or a specific location, the key entities involved should be identified, such as workers, analysts, etc., and their roles and interrelations should be known.

B. *Resource and Terminology Clarification*

This step determines the FHIR resources requisite to depict the data entities in the use case. In healthcare, this might include Patient, Practitioner, Condition, Observation, and Appointment resources, among others. Moreover, the value set for each characteristic of every entity is established, aligning this with standard semantic terminologies like ICD-10, SNOMED-CT, LOINC, and RxNorm to ensure interoperability.

C. *Understanding the Standard's Structure and Design of the Resources*

This step aims to learn the common structure of the standard's resources, including elements like "id", "meta", "implicitRules", and "language". Furthermore, for each resource, it should be defined the necessary data elements, their data types, cardinality, and optionality, while reusing standard definitions when applicable. To be more easily understandable, the following example represents a fictional patient resource. It shows a patient named TE POTE, with the identifier "002". This patient is a male patient and was born on November 14, 2000. He lives at 3300 Washtenaw, Ann Arbor, MI 48104, USA, and can be reached by phone at (03) 2019 1114. Fig. 2 depicts the same information in JSON structure.

ResourceType	Identifies the type of resource. In this case, the resourceType is "Patient".
ID	A unique identifier for the resource. For this example, the id is "592269".
Meta	Contains information about the version and last updated time and date of the resource, the resource is version 1 and was last updated on 1-24-2020, at 04:35 UTC.
Text	A human-readable text of the resource. In this case, the patient's name is TE POTE, their identifier is 002, and their address is 3300 Washtenaw, Ann Arbor.
Identifier	A list of identifiers for the patient. This case includes a single identifier with the system "http://example.org" with the value "002", it is different from the ID, which is the number of the patient, it contains many different versions of the representation of the patient.
Name	The names of the patients. In this case, the use "official", the family name (surname) "POTE", and the name "TE".
Telecom	A list of the patient's contact information and phone. It contains a single contact with the system "phone", the value "(03) 2019 1114", and the use "home"".

D. *Model Creation*

This step maps the involved entities, processes, and relationships to the standard's resources and protocols. It defines constraints, primary keys, and secondary keys for each entity. It also creates a profile that contains custom restrictions, extensions, value sets, and bindings that match the entities of the specific use case.


```
  "resourceType": "Patient",
  "id": "592269",
  "meta": {
    "versionId": "1",
    "lastUpdated": "2020-01-24T00:04:35.530+00:00",
    "source": "#70bjoSfolcTqa66T"
  },
  "text": {
    "status": "generated",
    "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><div
class=\"hapiHeaderText\">TE <b>POTE </b></div><table
class=\"hapiPropertyTable\"><tbody><tr><td>Identifier</td><td>002</td></tr><tr><td>Add
ress</td><td><span>3300 Washtenaw </span><br/><span>Ann Harbor </span><span>MI
</span><span>USA </span></td></tr><tr><td>Date of birth</td><td><span>14 November
2000</span></td></tr></tbody></table></div>"
  },
  "identifier": [ {
    "system": "http://example.org",
    "value": "002"
  } ],
  "name": [ {
    "use": "official",
    "family": "POTE",
    "given": [ "TE" ]
  } ],
  "telecom": [ {
    "system": "phone",
    "value": "(03) 2019 1114",
    "use": "home"
  } ],
  "gender": "male",
  "birthDate": "2000-11-14",
  "address": [ {
    "use": "home",
    "text": "3300 Washtenaw Ann Harbor, MI 48104, USA",
    "line": [ "3300 Washtenaw" ],
    "city": "Ann Harbor",
    "state": "MI",
    "postalCode": "48104",
    "country": "USA"
  } ]
}
```


Fig. 2. JSON Schema of a Patient Resource

The following JSON schema (Fig. 3) indicates an example of an observation resource and showcases how it relates to the patient resource. The reference element is important and reveals how the resources of a FHIR system are connected. Once the two resources are bound, someone can use the subject reference element in the observation resource to retrieve the patient resource with the number "592269". The reference practitioner reveals that this observation was performed by the practitioner with ID "56789".


```
{
  "resourceType": "Observation",
  "subject": {
    "reference": "Patient/592269"
  },
  "performer": [
    {
      "reference": "Practitioner/56789"
    }
  ]
}
```


Fig. 3. Observation Reference Example

E. *Search Parameters and Query Rules*

It defines how users will query the data model. It creates search parameters that allow for effective data retrieval, facilitating the understanding of the route of a FHIR resource, comprising the FHIR version, resource type, and the resource ID element. For example, the patient with the ID "592262" can be searched with "/r4/Patient/592262", which is the link for accessing information from the model and managing FHIR resources and variables on an FHIR server. This route is useful for many actions, including HTTP GET, PUT, and POST requests, as well as retrieval, creation, updating, and deletion of resources and their attributes on the FHIR server database. An HTTP POST request to "https://fhir-server-url>/r4/Patient" with the JSON or XML representation of the Patient resource in the request body, for example, will instruct the FHIR application to create a resource, with the resource ID returned.

F. Validation and Testing

It includes validation tools like the HAPI FHIR server that follows FHIR guidelines or code libraries that validate FHIR resources and models to make sure they have the right structure, and that the implementation is secure – this step is based on existing Software.

Following these steps, someone can manage to design the structure needed to build a standard-based model (with FHIR resources in this case), that correspond to different use cases, even if a user has no knowledge of code.

V. CONCLUSIONS

The usage of standards emerges as the leader framework, that facilitates a seamless dialogue among divergent systems. In healthcare, FHIR essentially crafts a universal lexicon, a common tongue, allowing every elemental value to resonate clearly and coherently across varying systems, also including an API that is flexible and simple to understand. The adaptability of FHIR models stands out; they mesh effortlessly with countless entities, untangling complex relationships. Beyond its technical finesse, the FHIR model extends itself as a standardized medium, elegantly structuring a broad spectrum of healthcare information, from the simplicity of patient demographics to the nuanced realm of clinical observations. It defines the basic variables and structure that a health system (syntactic interoperability) must follow to be secure and communicate with other systems in common terminology.

In this paper, we analyzed an important piece of existing technology and literature. Then it was suggested a method that one can follow to build a data model towards achieving interoperability, in any Big Data domain. This method analyzes thoroughly the most important resources of the used standard (e.g., patient, condition, etc.), as well as their most important individual elements (e.g., resource type, element definition, code system, etc.). The above methodology is written comprehensively and is addressed to researchers, scientists, and developers without presupposing the knowledge of code. We consider that it provides a basis for designing a data model and can be applied to most use cases, in a plethora of sectors. The limited sources that exist for the creation of a standard-based model and the available resources are among the challenges of this paper. Current research focuses more on the tools of accomplishing this than on the early steps. It is proposed to construct a practical implementation guide based on the above-mentioned technique, and it will incorporate Artificial Intelligence (AI) capabilities that are relevant in a variety of use cases and related research projects [16][17]. The applicability of this data modeling process should be also considered in the domain of cultural heritage protection, since it can facilitate interoperability in data (e.g., image) exchanging processes, computer vision analysis, as well as three-dimensional (3D) modeling.

ACKNOWLEDGMENT

The research presented in this paper has received funding from the EU funded Project SmartCHANGE (GA no 101080965), and the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation under the call RESEARCH-CREATE-INNOVATE (project code: RESPECT - T2EDK-03741.

REFERENCES

[1] Data Interoperability, https://www.nnlm.gov/guides/data-glossary/data-interoperability

[2] T. Benson, G. Grieve. Principles of health interoperability. Cham: Springer International, 2021, pp. 21-40.

[3] S. Ajami, et al., Barriers for adopting electronic health records (EHRs) by physicians. Acta Informatica Medica, **21(2)**, 2013, p. 129.

[4] S. N. Duda, et al. HL7 FHIR-based tools and initiatives to support clinical research: a scoping review. Journal of the American Medical Informatics Association, **29(9)**, 2022, pp. 1642-1653.

[5] Health IT Buzz, https://www.healthit.gov/buzz-blog/health-it/the-heat-is-on-us-caught-fhir-in-2019

[6] G. Manias, A. Mavrogiorgou, A. Kiourtis, D. Kyriazis. SemAI: A novel approach for achieving enhanced semantic interoperability in public policies. In Artificial Intelligence Applications and Innovations: 17th IFIP WG 12.5 International Conference, AIAI 2021, 2021, pp. 687-699.

[7] A. Kiourtis, A. Mavrogiorgou, S. Nifakos, D. Kyriazis. A string similarity evaluation for healthcare ontologies alignment to HL7 FHIR resources. In Intelligent Computing: Proceedings of the 2019 Computing Conference, **1**, 2019, pp. 956-970.

[8] A. Kiourtis, et al. Identity management standards: A literature review. Computers and Informatics, **3(1)**, 2023, pp. 35-46.

[9] N. Pimenta, et al., Interoperability of Clinical Data through FHIR: A review. Procedia Computer Science, **220**, 2023, pp. 856-861.

[10] A. Chatterjee, N. Pahari, A. Prinz. HL7 FHIR with SNOMED-CT to achieve semantic and structural interoperability in personal health data: a proof-of-concept study. Sensors, **22(10)**, 2022, p. 3756.

[11] R. H. Dolin, et al., Introducing HL7 FHIR Genomics Operations: a developer-friendly approach to genomics-EHR integration. J. of the American Medical Informatics Association, **30(3)**, 2023, pp. 485-493.

[12] K. D. Mandl, et al., Push button population health: the SMART/HL7 FHIR bulk data access application programming interface. NPJ digital medicine, **3(1)**, 2020, p. 151.

[13] R. A. Bloomfield Jr, et al., Opening the Duke electronic health record to apps: implementing SMART on FHIR. International journal of medical informatics, **99**, 2017, pp. 1-10.

[14] A. Kiourtis, A. Mavrogiorgou, A. Menychtas, I. Maglogiannis, D. Kyriazis. Structurally mapping healthcare data to HL7 FHIR through ontology alignment. Journal of medical systems, **43**, 22019, pp. 1-13.

[15] A. Torab-Miandoab, et al. Interoperability of heterogeneous health information systems: a systematic literature review. BMC Medical Informatics and Decision Making, **23(1)**, 2023, p. 18.

[16] D. Kyriazis, et al., The CrowdHEALTH project and the hollistic health records: Collective wisdom driving public health policies. Acta Informatica Medica, **27(5)**, 2019, p. 369.

[17] A. Mavrogiorgou, et al., FAME: Federated Decentralized Trusted Data Marketplace for Embedded Finance. In 2023 International Conference on Smart Applications, Communications and Networking (SmartNets), 2023, pp. 1-6.