Bridging the Gap: Integrating Heterogeneous Clinical Data into HL7 FHIR

Teresa Conte†, Nadia Brancati†, Martina Russo†, Mario Sicuranza†,
†Institute for High Performance Computing and Networking (ICAR)
National Research Council of Italy (CNR), Naples, Italy
Email: {teresa.conte, nadia.brancati, martina.russo, mario.sicuranza}@icar.cnr.it

***Abstract*—The fragmentation of patient data across different and disconnected systems, ranging from electronic health records to Artificial Intelligence (AI)-based diagnostic tools, poses a major challenge to the delivery of an efficient and accurate healthcare system. This paper proposes a modular and interoperable architecture designed to integrate heterogeneous clinical data from different sources, including structured clinical records, socio-health information, patient-generated data, and outputs from AI-based diagnostic systems such as imaging analysis.**

The proposed architecture facilitates seamless data harmonization and supports clinical decision-making by structuring integrated information through the HL7 Fast Healthcare Interoperability Resources (FHIR) standard. This enables standardized data exchange and full interoperability with existing Health Information Systems, including Electronic Health Records and Telemedicine Platforms. An Implementation Guide is proposed as a reference framework for validating the FHIR resources produced by the architecture. In addition, a key feature of the architecture is its embedded Clinical Decision Support System, which dynamically identifies and presents only the clinically relevant information required for diagnostic reasoning and risk assessment.

***Index Terms*—HL7 FHIR, Interoperability, Machine Learning, Decision Support System.**

I. Introduction

The digital transformation of healthcare has led to an exponential increase in the generation of clinical data from heterogeneous sources such as Health Information Systems (HISs), wearable devices, imaging platforms, laboratories, and mobile applications. The increasing complexity and heterogeneity of this clinical data pose significant challenges to data interoperability, clinical reasoning, and diagnostic efficiency. Modern healthcare systems need to support not only the aggregation of this multifaceted information but also its semantic harmonization and contextualization to provide meaningful, actionable insights for healthcare professionals.

Despite this abundance, data often remain isolated, unstructured, and non-interoperable, limiting their effective use in clinical practice and research. This fragmentation significantly hampers the implementation of advanced tools such as Artificial Intelligence-based Clinical Decision Support Systems (AI-CDSS), which require standardized and semantically rich datasets to ensure reproducibility, reliability, and personalization of care [1]. Missing or incomplete modalities, inconsistent data formats, and variability in acquisition protocols are frequent, often limiting model generalizability. Addressing these issues requires harmonization strategies, natural language processing for unstructured text, and robust imputation or domain adaptation methods.

To overcome these challenges, the HL7 Fast Healthcare Interoperability Resources (FHIR) standard has emerged as a modern, web-based framework for structuring and exchanging health data. FHIR enables granular representation of clinical concepts through modular resources (e.g., Patient, Condition, Observation), facilitating flexible integration across systems via RESTful APIs [2]. The adoption of FHIR not only promotes interoperability and data quality but also supports traceability, explainability, and regulatory compliance, critical aspects in the context of AI integration. A recent review revealed that over 98% of CDSS developed between 2018 and 2021 leveraged FHIR as the primary interoperability standard, underscoring its growing relevance in digital health initiatives [3].

Several initiatives and studies have demonstrated the potential of the standard FHIR in enhancing clinical workflows and enabling AI-based applications. The platform ”SMART” based on FHIR framework, presented in [4] for instance, allows the development of secure and reusable applications within Electronic Health Records (EHRs) through the use of FHIR profiles and authentication protocols such as OAuth2 [1] and OpenID Connect [2]. Research by Major et al. [5] showed how integration of a FHIR back-end with one of the most widely adopted EHRs, ”Epic” (5 leading EHR providers and is the most adopted EHR system across the globe) [6], enables real-time analysis of clinical notes, medications, and vitals data, thus improving AI-based decision-making.

Similarly, Lamprinakos et al. demonstrated how FHIR-based mobile applications can facilitate personalized care by connecting clinicians, patients, and pharmacists [7].

FHIR is increasingly being used to define structured data models that support precision medicine. To support the FHIR standard in defining precise instructions, Implementation Guides (IGs) are used. They help identify and summarize typical requirements for the development of specific architectures.

[1] https://oauth.net/2/
[2] https://openid.net/developers/how-connect-works/

For instance, in oncology, the mCODE (here is the IG [3]) initiative provides a core set of FHIR profiles for cancer data, supporting reuse in both clinical and research contexts [8]. Other projects, such as OSIRIS in oncology (here is the IG [4]) , extend this approach to genomic data and multi-center studies, enabling standardized data exchange across institutions [9].

Moreover, numerous FHIR IGs have been developed at both international and national levels to ensure semantic coherence and clinical applicability in real-world settings. The International Patient Summary (IPS) [5], conforming to ISO 27269, defines a minimal dataset of essential clinical information to support unscheduled and cross-border care, enabling semantic alignment across jurisdictions. The European Hospital Discharge Report (EHDR) [6] provides a structured model for hospital discharge summaries throughout Europe, with a strong focus on the modularity of clinical content and continuity of care. At the national level, the Italian Personal Health Diary (in italian known as ”Taccuino Personale dell'Assistito”) [7] allows patients to autonomously input anamnesis-related data into the EHRs, promoting structured, patient-centered documentation and encouraging greater engagement in the healthcare process.

In this research, the aim is to improve the integration of heterogeneous data from different sources, such as patient family history, and medical image reports, enabling AI-based CDSS use in clinical practice by adopting standardized and interoperable frameworks accordance with HL7 FHIR. Inspired by successful models and IGs discussed in the literature, we present our architecture, designed to assist clinicians by providing AI-supported diagnostic insights based on standardized input and to empower patients by promoting consistent and transparent communication of health data. Our architecture uses for the definition and validation of clinical data the ”Anamnesi in Remoto” FHIR IG [8] which we appropriately defined for the purpose of the architecture. Moreover, this approach offers a replicable framework for many clinical domains, especially in contexts where access to expert interpretation is limited.

II. Architecture

In this study, we propose a modular architecture designed to collect and integrate data from different and heterogeneous sources, including HISs, EHRs, and proprietary systems such as mobile applications. The architecture allows patients to contribute their own clinical information in the form of anamnesis and integrates outputs from AI algorithms (e.g., image-based risk assessments). The integrated data are then standardized using the HL7 FHIR, enabling full interoperability with existing health IT systems.

[3]https://build.fhir.org/ig/HL7/fhir-mCODE-ig
[4]https://ig-osiris.cancer.fr/ig/osiris/
[5]https://build.fhir.org/ig/HL7/fhir-ips
[6]https://build.fhir.org/ig/hl7-eu/hdr
[7]https://www.hl7.it/fhir/taccuino
[8]https://anamnesi.na.icar.cnr.it/

A key design principle of the proposed architecture is its ability to transform raw and fragmented clinical data into structured, diagnosis-oriented information, supporting risk assessment and decision-making. The architecture ensures that only the relevant and necessary clinical information is presented to clinicians, contextualized, and enriched with AI-based assessments, thus minimizing information overload and facilitating precise diagnostics. The architecture has five core objectives:

Aggregate and Data Collection: Consolidating data from multiple sources, such as: i) medical and family history information via patient interviews or digital forms; ii) diagnostic reports and data obtained through AI-based diagnostic tools for interpretation of imaging and iii) structured data from HIS platforms.

Data Standardization: Harmonizing all collected and processed data by transforming them into HL7 FHIR-compliant resources, ensuring semantic interoperability and system integration.

Focused and Rapid Clinical Data Visualization: Reporting of only the most relevant, diagnosis-specific information through an intuitive user interface, streamlining clinical workflows, and enhancing decision-making efficiency.

Decision Support: Generating a clinical summary and AI risk assessment based on the integrated dataset, to assist healthcare professionals in diagnostic reasoning.

Integration with HISs: Exchanging standardized data with existing systems (EHSs and telemedicine platforms) via HL7 FHIR.

Figure 1 presents a comprehensive view of the proposed modular architecture, which is centered around a data flow that enables both healthcare professionals and patients to contribute and interact with relevant health information. The architecture includes the following key components:

Data Sources and Input

- *HIS/EHR Systems* Clinical data originates from traditional HISs and EHRs, accessible to healthcare professionals.
- *Patient Front-End:* Patients can input their own clinical data (e.g., anamnesis) through a dedicated application.
- *Diagnostic Images and Reports:* Imaging data (e.g., ultrasound, radiographs) and associated diagnostic reports are introduced into the pipeline.

Data Aggregation and Processing

- *Aggregated and Collected Data:* This module consolidates all incoming data from heterogeneous systems, patient inputs, and diagnostic sources.
- *Diagnostic Image Processing:* A specialized module processes medical images using AI-based techniques. It includes an Information Features Extractor, which identifies clinically relevant features to support risk analysis and reporting.

Standardization and Interoperability

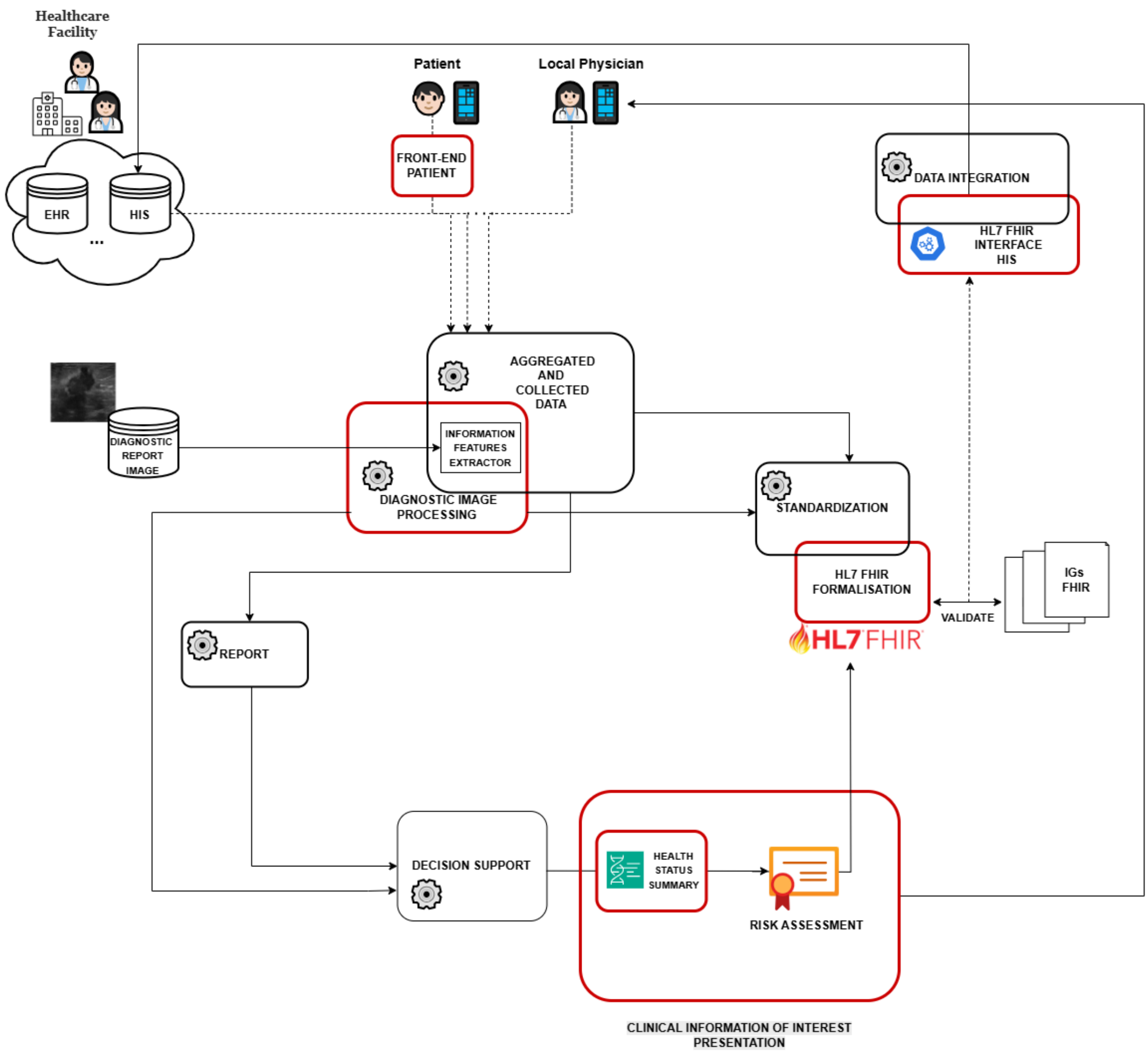

Fig. 1. Architecture System.

- *HL7 FHIR Formalisation:* The module maps all aggregated data into the HL7 FHIR format.
- *Data Integration:* A dedicated FHIR interface for heterogeneous systems ensures interoperability and the ability to communicate with external healthcare infrastructures.
- *Validation:* The formatted data is validated against defined FHIR IGs to ensure compliance and consistency with our solution.

Clinical Information and Risk Assessment

- *Report Generation:* A report module compiles the outputs into a structured, clinician-friendly document.
- *Decision Support:* A decision support system leverages the processed and standardized data to assist clinicians in diagnosis and treatment planning.
- *Clinical Information of Interest Presentation:* Finally, only the most relevant and diagnosis-specific information is presented to clinicians via an intuitive and responsive interface.
- *Risk Assessment:* This module evaluates patient risk using both structured clinical data and AI outputs, producing a Health Status Summary.

In the following, we discuss the architecture in greater depth, examining each constituent module, its internal functionality, and how it contributes to the overall system workflow and objectives. Particular attention is given to the interactions between components and the role each plays in ensuring data interoperability, clinical relevance, and usability. In the Figure 1, the modules that implement the front-end functionality of the architecture are contoured in red, while in black boxes, the back-end components are reported.

- *Front-end Patient*
  It allows the patient to independently collect a range of anamnestic information (possibly specific to the clinical condition). The module allows the management of data obtained by a form presented to the patient for the collection of medical history and other relevant information (allergies, medications, family history, symptoms, etc.). This module also allows the patient to retrieve any information present in the system through interaction with the HL7 interface module.
- *Diagnostic Image Processing Module (DIPM)*
  It performs analysis and processing of diagnostic images (e.g., detection, segmentation, classification, feature extraction) using Machine Learning algorithms or specific analysis tools. The aim of this module is to return an AI-based risk assessment.
- *HL7 FHIR Formalisation Module*
  This module allows formalising, through international standard HL7 FHIR profiles, the data provided by the patient's anamnesis form and the DIPM's image processing results. This module uses the FHIR profiles and resources (e.g., Patient, Observation, Diagnostic Report, Allergy Intolerance, Medication Statement), which are appropriately defined to ensure compliance with FHIR standards. It manages the creation of FHIR resources and carries out validation for the aim of proposed architecture.
- *Clinical Information of Interest Presentation Module*
  This module allows identifying and collecting all and only the clinical information of interest for a specific diagnosis (patient's clinical condition and medical history). It thus enables the presentation of an integrated and intuitive view of the information: personal data, structured medical history, and image processing results.
- *Health Status Summary Module*
  It applies logical rules and algorithms to extract and analyze integrated data (history, image results, other available data), and generates a concise summary of the patient's health status, highlighting key information, potential risks (AI-based risk assessment), and recommendations.
- *HL7 FHIR Interface with HIS*
  It enables the Clinical Information of Interest Presentation Module to communicate with the existing platform (EHR) using the HL7 FHIR standard, and supports FHIR operations such as Create, Read, Update of Patient resources, Observation, Diagnostic Report, and other relevant ones.

III. Implementation Guide

The proposed solution requires the definition and exchange of information to validate the data structures (FHIR resources) both during construction and communication. For this aim, an Implementation Guide has been developed, which enables: i) Standardized formalization of the data structures, and ii) Validation of FHIR resources against the functional and regulatory requirements of the solution, both during creation and in the exchange/interoperability between systems. This approach ensures consistency, integrity, and interoperability of data throughout the entire lifecycle of the FHIR resources. The FHIR IG, associated with the proposed solution, is "Anamnesi in Remoto" [9]. The IG was defined with the support of HL7 Italia's members. The objective was to enable structured, interoperable, and reusable representation of clinical data, particularly for remote anamnesis. This process involved a detailed analysis of existing standards and implemented solutions (IPS FHIR IG, EHDR FHIR IG, and Taccuino FHIR IG in italian), followed by the design and formalization of a context-specific FHIR IG aligned with the architectural and interoperability requirements of the system.

The following methodology included: (i) comparative analysis of existing IGs and standards (CDA/FHIR) ; (ii) definition of a conceptual model focused on the most relevant elements for anamnesis; (iii) selection and mapping of appropriate FHIR resources; (iv) creation of customized profiles using FHIR Shorthand (FSH) [10] [11]; and (v) generation and validation of FHIR structures using SUSHI [11] and IG Publisher [12] tools. More details on the technical methodology used in the definition of IG through the official FHIR tools are described in the work presented in [13].

The resulting IG specifies fourteen customized FHIR profiles, complemented by customized ValueSets and CodeSystems to support the classification of clinical concepts. The underlying information model encompasses essential clinical data domains, including symptoms, diagnoses, clinical procedures, pharmacological treatments, allergies, implanted medical devices, informed patient consent, and diagnostic reports.

Figure 2 illustrates the class diagram that delineates the core concepts identified and profiled within the defined IG. This diagram provides a structured representation of the key entities, their attributes, and the relationships necessary to model the domain according to the HL7 FHIR specifications and project requirements. The primary classes represented in the diagram include, but are not limited to:

- *AllergyIntolerance_Patient*: Used to represent the patient's known allergies and intolerances.
- *CarePlan_Patient*: Models healthcare professionals and organizations involved in patient care delivery.
- *Condition_Patient*: Records the patient's medical conditions, including diagnoses.
- *Consent_Patient*: Represents the patient's expressed authorization regarding the collection, access, use, and sharing of their health data.
- *Device_Patient*: Represents the implanted or otherwise utilized medical devices associated with the patient, including monitoring equipment, therapeutic devices, and

[9]https://anamnesi.na.icar.cnr.it/

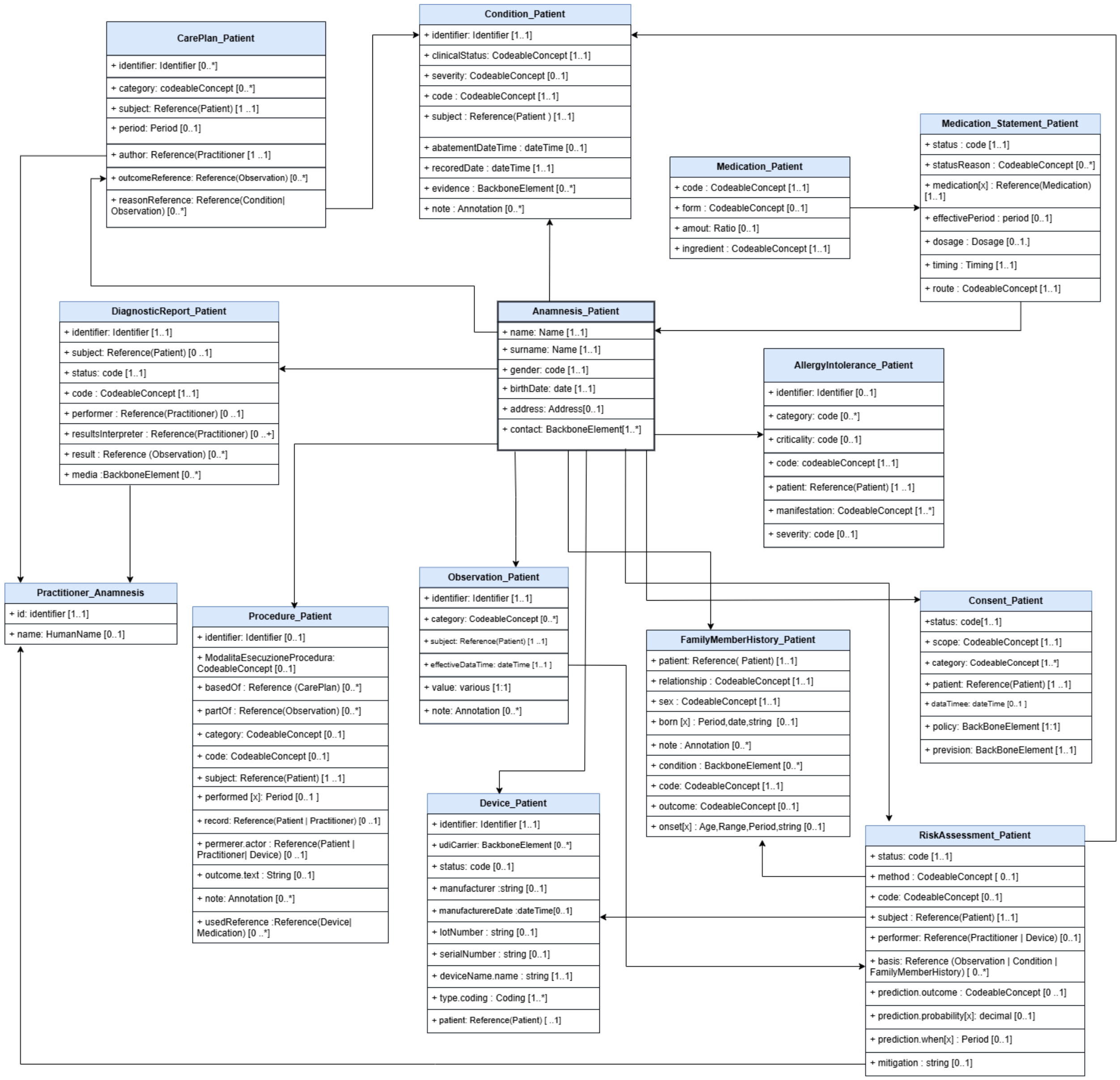

Fig. 2. Overview of the Information Model and profiled FHIR resources.

prosthetic devices such as implants (e.g., hip or knee prostheses).

- *DiagnosticReport_Patient*: Contains the structured clinical findings and diagnostic interpretations generated from laboratory or imaging data.
- *FamilyMemberHistory_Patient*: Represents the clinical history of genetic conditions that could affect the patient.
- *Medication_Statement_Patient*: Used to represent the set of medications taken by the patient.
- *Medication_Patient*: Used to describe the characteristics of drugs such as active ingredient, pharmaceutical form, and dosage.
- *Observation_Patient*: Represents measurements or assertions made about a patient's health status.
- *RiskAssessment_Patient*: Encapsulates the results and parameters of clinical risk evaluations derived from patient data and AI algorithms.
- *Anamnesis_Patient*: Represents the list of patient's information, including demographic and identification attributes.
- *Practitioner_Anamnesis*: Represent the health professional involved in patient care, e.g., the physician in charge of the examination.
- *Procedure_Patient*: Details clinical interventions or treat-

ments.

In the information model, the patient is the central entity that connects all the data. The patient's clinical status and interventions are described by Condition, Observation, and Procedure, while the patient's consent for data processing is described by Consent. Family Member History provides the hereditary context of the patient's diseases, while Medication/MedicationStatement records medication use. The Diagnostic Report aggregation the patient's laboratory and imaging results. RiskAssessment integrates clinical and AI-derived risks. Finally, Practitioner identifies the responsible healthcare professional.

Following the formal definition of the IG, it was systematically integrated within the established architectural framework to enable the structured representation of all information acquired from heterogeneous external systems. This integration was specifically designed to ensure full compliance with the HL7 FHIR standard, thereby providing a consistent and standardized methodology for managing and exchanging healthcare data across diverse platforms. The architecture leverages the IG as a foundational element to harmonize disparate datasets, transforming heterogeneous inputs into interoperable FHIR-compliant resources. This approach facilitates semantic consistency and syntactic interoperability, which are critical for maintaining data integrity and enabling efficient communication among various healthcare information systems. In addition, the previously exposed modules in the section II, responsible for clinical information, utilize the IG to validate clinical data (as FHIR resources) they produce. This validation process ensures that all generated resources strictly adhere to the defined profiles, constraints, and value sets specified within the IG, thereby guaranteeing conformance to recognized interoperability standards. Consequently, the validated resources are made available in a format that supports seamless integration and accessibility by external clinical systems (compliance with HL7 FHIR) and authorized healthcare professionals, fostering an environment of reliable, interoperable data exchange that enhances clinical decision-making and supports coordinated patient care.

The defined IG offers a scalable and reusable solution that can be extended to several clinical domains (e.g., oncology follow-up, chronic disease monitoring), representing a valuable contribution to the evolution of the Italian EHR initiative as well as to modern, patient-centered digital health paradigms founded on semantic interoperability principles.

IV. Conclusions

In this study, we adopted the HL7 FHIR standard to design an architecture capable of integrating ICT Clinical Solutions, and HISs, including EHRs, socio-health datasets, patient-reported outcomes, and AI-based risk assessment outputs. We propose a modular and fully interoperable architecture designed to facilitate seamless data harmonization and exchange. It is based on the definition of a specific IG. At the core of this architecture lies the HL7 FHIR standard, which serves to normalize disparate data formats and guarantee comprehensive interoperability with extant HISs, such as EHRs and Telemedicine platforms. A pivotal element within the system is an embedded CDSS, engineered to dynamically filter and prioritize clinically relevant data, thereby augmenting diagnostic precision and supporting robust risk-stratification processes. Through the implementation of standardized data exchange protocols and an intelligent decision support mechanism, this architecture effectively addresses the inherent complexity associated with integrating diverse healthcare datasets. Moreover, it achieves this with scalability and operational efficiency, positioning the solution as a viable approach to meet the demands of contemporary digital health ecosystems.

Acknowledgment

References

[1] S. Ramgopal, L. N. Sanchez-Pinto, C. M. Horvat, M. S. Carroll, Y. Luo, and T. A. Florin, "Artificial intelligence-based clinical decision support in pediatrics," *Pediatric research*, vol. 93, no. 2, pp. 334–341, 2023.

[2] R. Sreejith and S. Senthil, "Smart contract authentication assisted graphmap-based HL7 FHIR architecture for interoperable e-healthcare system," *Heliyon*, vol. 9, no. 4, 2023.

[3] P. Taber, C. Radloff, G. Del Fiol, C. Staes, and K. Kawamoto, "New standards for clinical decision support: a survey of the state of implementation," *Yearbook of medical informatics*, vol. 30, no. 01, pp. 159–171, 2021.

[4] J. C. Mandel, D. A. Kreda, K. D. Mandl, I. S. Kohane, and R. B. Ramoni, "SMART on FHIR: a standards-based, interoperable apps platform for electronic health records," *Journal of the American Medical Informatics Association*, vol. 23, no. 5, pp. 899–908, 2016.

[5] V. J. Major, W. Wang, and Y. Aphinyanaphongs, "Enabling AI-augmented clinical workflows by accessing patient data in real-time with FHIR," in *Proceedings of the 2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*. IEEE, 2023, pp. 531–533.

[6] J. Chishtie, N. Sapiro, N. Wiebe, L. Rabatach, D. Lorenzetti, A. A. Leung, D. Rabi, H. Quan, and C. A. Eastwood, "Use of Epic electronic health record system for health care research: scoping review," *Journal of medical Internet research*, vol. 25, p. e51003, 2023.

[7] G. C. Lamprinakos *et al.*, "Using FHIR to develop a healthcare mobile application," in *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare (MOBIHEALTH)*. IEEE, 2014, pp. 132–135.

[8] T. J. Osterman, M. Terry, and R. S. Miller, "Improving cancer data interoperability: The promise of the minimal common oncology data elements (mCODE) initiative," *JCO Clinical Cancer Informatics*, vol. 4, pp. 993–1001, Oct 2020.

[9] J. Guérin, Y. Laizet, V. Le Texier, L. Chanas, B. Rance, F. Koeppel, F. Lion, S. Gourgou, A. L. Martin, M. Tejeda, M. Toulmonde, S. Cox, E. Hess, M. Rousseau-Tsangaris, V. Jouhet, and P. Saintigny, "OSIRIS: A minimum data set for data sharing and interoperability in oncology," *JCO Clinical Cancer Informatics*, vol. 5, pp. 256–265, Mar 2021.

[10] FHIR Community, "Fhir shorthand (fsh) - hl7 fhir ig definition language," https://build.fhir.org/ig/HL7/fhir-shorthand/.

[11] ——, "Sushi - sushi unshortens short hand inputs," https://github.com/FHIR/sushi.

[12] HL7 International, "FHIR IG Publisher," https://github.com/HL7/fhir-ig-publisher.

[13] T. Conte and M. Sicuranza, "Sistema FHIR-based per la raccolta strutturata dell'anamnesi in remoto: Approccio, standard e implementazione," CNR-ICAR, Rapporto Tecnico RT-ICAR-NA-2025-05, 2025. [Online]. Available: https://intranet.icar.cnr.it/wp-content/uploads/2025/06/RT-ICAR-NA-2025-05.pdf