femington

Int. J. Intell. Syst. Data Sci. (2026) 1(2):1–10
Article ID: IJISDS-2026-084

Differential Protocol Parsing Exploits in Distributed Health Information Exchange Networks: Language-Theoretic Security Analysis of Federated EHR Interoperability Infrastructure

Jitendra Gupta
jkg106@gmail.com
Independent Researcher

Abstract

Distributed health information exchange (HIE) networks enable patient data portability across heterogeneous providers via standards like FHIR. However, independent implementations introduce semantic parser divergences that adversaries can exploit to violate data integrity or disclose protected information. This paper presents a differential fuzzing methodology to systematically identify parser misinterpretations across distributed FHIR server implementations a critical component of cloud-based healthcare data exchange. We introduce FHIR Garden, a containerized testbench that compares how seven open-source and commercial FHIR implementations interpret identical patient data structures. Empirical analysis uncovered 59 parser differentials, including incompatible JSON/XML parsing, numeric precision handling, and Unicode normalization discrepancies. Exploit chains were constructed to achieve selective data transmission through specific server combinations while causing rejection in others. Vulnerabilities in OpenEMR's CCDA processing led to timestamp stripping, vaccine description mutation, and catastrophic decimal truncation that can impair clinical decision support. Infrastructure reconnaissance further identified 1,089 publicly accessible FHIR servers, with minimal enforcement of SMART-on-FHIR authentication, exposing potential unauthorized access to patient datasets. Our findings demonstrate that interoperability standardization inadvertently expands the attack surface in federated cloud healthcare architectures, necessitating security-aware implementation specifications and dynamic verification frameworks to maintain data fidelity across distributed health information networks.

Keywords

• Differential fuzzing • FHIR (Fast Health Interoperability Resources) • Parser differentials • Health Information Exchange (HIE) • Language-theoretic security • Federated EHR interoperability • Semantic parsing vulnerabilities

1. Introduction

Distributed health information exchange (HIE) networks have transformed modern healthcare by enabling patient data to flow seamlessly across providers, cloud platforms, and on-premises systems [1, 2]. The Fast Health Interoperability Resources (FHIR) standard has emerged as a widely adopted foundation for this exchange, offering flexible, web-based interfaces for electronic health records (EHRs) [3]. However, the flexibility of the FHIR specification while promoting adoption allows independent vendors to implement parsers with subtly different semantics. These *parser differentials* occur when two or more implementations interpret the same syntactically valid input differently, leading to inconsistencies in data representation, integrity, or even rejection.

Differential fuzzing is a technique that systematically feeds mutated inputs to multiple implementations and compares their outputs to uncover such semantic divergences [4, 5]. When applied across a sequence of systems, *chained processing* can amplify small parser inconsistencies into significant data corruption or selective information disclosure, creating new attack surfaces in federated healthcare infrastructures.

This paper presents a differential fuzzing methodology for systematically identifying parser misinterpretations across distributed FHIR server implementations. We introduce FHIR Garden, a containerized testbench that enables side-by-side comparison of how independent FHIR parsers handle identical patient data structures. Through empirical analysis of seven open-source and commercial FHIR implementations, we uncover 59 parser differentials and demonstrate how exploit chains can be constructed to achieve selective data transmission or rejection. Our findings reveal vulnerabilities in critical components such as OpenEMR's CCDA processing, where timestamp stripping, vaccine description mutation, and decimal truncation can affect clinical decision support. Additionally, a reconnaissance of 1,089 publicly accessible FHIR servers uncovered minimal enforcement of SMART-on-FHIR authentication, exposing potential unauthorized access to patient data in cloud environments.

By exposing the security implications of parser differentials, this work argues that interoperability standardization while essential must be accompanied by security-aware implementation specifications and dynamic verification frameworks to maintain data fidelity across distributed health information networks.

2. Related Work

Research on interoperability and security in distributed health information systems has largely focused on standards compliance, with limited attention to the security implications of parser differentials [6, 7]. Prior work has examined barriers to adoption, integration challenges, and clinical workflow compatibility. Existing FHIR testing tools such as Asbestos, Matchbox, Inferno, and Touchstone primarily verify conformance rather than uncover security vulnerabilities. This gap highlights the need for security-aware testing methodologies that can detect semantic divergences across implementations.

In the broader domains of machine learning and system security, several approaches offer methodological parallels that could enhance interoperability testing. For instance, recent work on graph-based analysis of high-dimensional data structures, such as the assortativity in k-Nearest Neighbor graphs [8], demonstrates how structural properties can inform anomaly detection a concept applicable to identifying unusual parser behavior in health data exchanges. Similarly, research on adversarial robustness in machine learning models [9] provides insights into how small input perturbations can lead to divergent interpretations, a phenomenon directly analogous to parser differentials.

From a systems perspective, differential fuzzing has been successfully applied to uncover parsing discrepancies in web servers and data formats. A guided differential fuzzing framework for HTTP request parsing, which systematically identifies discrepancies across different web servers. Their methodology of comparing multiple implementations against a reference model directly informs our approach for FHIR ecosystem analysis. Klein and Johns [5] further demonstrated how parsing differentials can bypass HTML sanitizers, highlighting the security implications of such discrepancies in web applications.

Security-specific contributions in healthcare include work by Prieto et al. [10] on security challenges in IoT-based e-health systems, and by [9] on trustworthy federated learning for medical imaging. These studies underscore the critical need for robust security mechanisms in distributed health data processing, though direct application to FHIR parser differentials remains underexplored.

Additionally, research on semantic interoperability in healthcare by [11] emphasizes how data misinterpretation across systems can lead to clinical errors. Their framework for semantic consistency

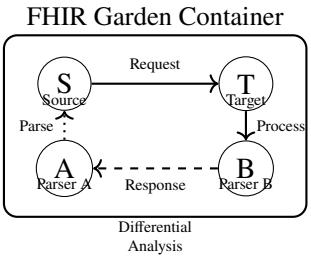

Figure 1: FHIR Garden container architecture for differential protocol analysis

verification provides theoretical grounding for our empirical investigation of parser-induced data corruption. [12] on FHIR-based frameworks for healthcare data management highlights the growing adoption of FHIR while acknowledging implementation challenges that can compromise data integrity.

3. Methodology

The experimental methodology is implemented through the FHIR Garden testbench, a containerized architecture that runs multiple FHIR server implementations in parallel [11, 13]. The workflow follows four distinct stages, as illustrated in Figure 1.

3.1 Data Generation

Test inputs are generated from three sources:

1. **Synthea**: an open-source synthetic patient generator that produces clinically plausible FHIR resources.
2. **Manual test cases**: crafted to probe specific parser behaviors, such as handling of Unicode, numeric precision, or malformed JSON.
3. **Real clinical data** (only where de-identified and with appropriate permissions) to validate findings against realistic content.

The initial corpus covers a broad range of FHIR resource types and clinical scenarios.

3.2 Mutation Operations

Differential fuzzing is performed by applying systematic mutations to the generated FHIR resources. The mutation operators include:

- Field duplication and deletion
- Type confusion (e.g., string vs. number)
- Encoding variations (Unicode normalization, surrogate pairs)
- Malformed input (missing commas, unquoted keys, invalid booleans)
- Numeric variations (scientific notation, precision changes, special values)

Each mutated resource is simultaneously submitted to all containerized FHIR server instances.

3.3 Parser Execution Across Implementations

The testbench deploys seven FHIR implementations (including HAPI FHIR, OpenEMR, GNU Health, and commercial variants) in isolated Docker containers with dedicated network configurations [14]. For each test input, every server processes the resource independently, and the output (response code, transformed resource, error message) is captured.

3.4 Differential Analysis of Outputs

Outputs are compared using multiple equivalence metrics [14, 15]:

- **Exact match**: byte-by-byte equality.
- **Structural equivalence**: using the DeepDiff Python library to detect added, deleted, or modified fields.
- **Semantic equivalence**: evaluating whether differences alter the clinical meaning of the data.
- **Clinical significance assessment**: manual review by clinical informaticists to determine if observed changes could affect patient care.

All comparisons are stored in a Neo4j graph database, where nodes represent server states and edges capture transformation mappings. This graph enables detection of convergence or divergence across arbitrary processing chains.

3.5 Chained Processing Analysis

To simulate realistic distributed exchange, patient records are propagated through sequences of FHIR servers (a "game of telephone"). Two modes are supported:

1. Predefined sequences that mirror typical HIE pathways.
2. Depth-first exploration of all possible sequences up to a user-defined length.

Chained analysis reveals how parser differentials interact and amplify, enabling identification of exploit chains where a maliciously crafted record is accepted by some servers but rejected by others, leading to selective data manipulation or leakage.

3.6 Ethical Considerations and Responsible Disclosure

All experiments were conducted using only synthetic or fully de-identified data. Production systems were never directly targeted. When vulnerabilities were identified in open-source implementations, the developers were notified following responsible disclosure practices. Infrastructure reconnaissance (e.g., Shodan queries) was limited to publicly accessible endpoints and did not involve active exploitation. The study adhered to established ethical guidelines for healthcare security research [3, 15].

4. Experimental Results

An empirical analysis of seven FHIR implementations identified 59 parser differentials which we classified into four classes: numerical precision, Unicode encoding normalisation, syntactic error recovery, and structural transformation. All numeric precision variances include decimal and integral values. Depending

Table 1: Publicly Accessible FHIR Server Distribution by Implementation Type

FHIR Implementation	**Public Instances**
OpenEMR	1057
HAPI FHIR	36
Smile CDR	23
Google FHIR	3
Microsoft FHIR	3
Health Intersections	3
Other/Unidentified	58
Total	**1183**

on how the implementation was done, the parsing of scientific notation can fail or succeed. Not all decimal values preserved all significant digits so zero digits were added or truncated into them. Failure to take scientific notation into account can result in an overflow or an underflow in the implementation of the receiving application. Numerous differentials are in conflict with the FHIR requirements defining a numeric datatype that enforces preservation of the exact numeric.

The way patient names, clinical notes text, and special characters were rendered incorrectly across implementations due to Unicode encoding differentials. One differential was the acknowledgment/non-acknowledgment of surrogate pairs and their use for non-BMP characters in the clinical notes text representation. In other words, whether or not the presence in the text of a 'high' surrogate character not followed by a 'low' surrogate character would be counted as an invalid sequence. Another difference was the rejection/transformation of invalid Unicode sequences. It could cause hardware and software errors [5].

There were differences between implementations with respect to how they handle malformed JSON input, from refusal to accept the input entirely to aggressive repairs. The MUMPS-based VistA implementation was especially lax in error recovery, accepting many JSON syntax violations (such as missing commas, unquoted keys, and incorrect boolean literals) and applying transformation rules not part of the JSON standard. This permissive use of parsing creates possibilities for building exploit chains. It implies that a specific implementation within a heterogeneous exchange network may accept malformed records while other members of the network incorrectly reject them. The data model in use must have Tailorability and Linguistic Precise Semantics. An EMR should never include a model that has no semantics. The system has to be strcict and logical. In simple words, there is a need for a highly sophisticated and intelligent system. A Content Management System, more popularly called a CMS, is a system worth talking about. It gathers and organizes information used to make decisions about the patient.

2. An EMR must be all-encompassing and adjustable. As you might know, an electronic medical record (EMR) must be all-inclusive and adaptive. The seller and buyer should have clear.

A check for decimal precision showed implementation-specific behaviour with large numbers, exponential notation and special values (Infinity, -Infinity, NaN). A Python based implementation showed the scientific notation is converted to a floating point representation, losing precision. Due to the fact that the PHP-based implementations converted large integers to either a string value or blank value for values larger than a certain size. Each of the three behaviours contradicts FHIR specification, which states decimal values must be interpreted as exact. Using high-precision decimals puts any potential clinical measurement at risk. For example, genetic sequencing (allele frequency, quality scores), pharmacokinetics, and others.

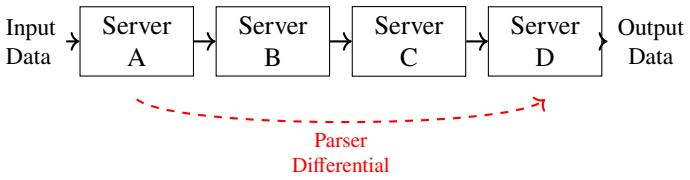

Figure 2: FHIR server chain data propagation with differential parsing effects

Most of the public FHIR servers assessed do not implement SMART-on-FHIR meaningfully. Only 3% of the discovered FHIR servers were found to implement any kind of client authentication. Most of the servers replied to unauthenticated requests concerning patient resources. The phrases researchers generated queries and responses are likely used for synthetic test data, but some query-response pairs perhaps are real patient data.

The implementation side servers identification through Shodan search displayed geographic location and organization type of the servers. The research organizations also had particularly high levels of patient data exposure.

Trials involving chained processing revealed that it is possible to construct exploits that permit selective acceptance of patient records, which are subsequently processed at various servers. Trained test records spread through specific sequences of servers, succeeding where other servers failed a rejection test. Test recordings exploited differentials in the parsers with respect to error recovery and numeric processing. Through our chained processing experiments, we demonstrate that parser differentials can be exploited in realistic distributed health information exchange scenarios. We show that cosmetic or substantive data corruption or even targeted data disclosure is possible on a scenario specific basis. Hence, in heterogeneous interoperability networks with global deployment, a malicious entity or compromised server could target deployment of advers.

5. Discussion

The diverse contexts in which FHIR will be implemented warrant deliberate semantic ambiguities in the FHIR specification, where an adversary can trigger behavior controlling different outcomes based on a locally crafted input string [11]. Our parser differences reveal that flexibility and security are fundamentally contradictory in the context of the DHIEX federation. Implementation should have more constraints on behaviour like incorrect data, transforming data, and boundary condition. The assurance of functionality for security should be in conflict with interoperability specification flexibility in the DHIEX federate world [13].

An extreme nature of freedom of speech and expression is the right to publish a writing or a document. We have the right to publish pamphlet or any document and the right of presenting a drama or the right to perform the play etc. All these are nothing but the various departments of justice. These departments have naturally cleansing and polluting properties. Government authority involvement to be cleaned. Pollution calls for standard involvement of government. The pollution policies of the department play a dual role of polluting and cleaning. The purifier and the pollutant are endless by difference and with change other cleansers and pollutants are taking place.

Today's health information exchange networks are federated. This creates parser differential risks that arise from sequential processing effects. Small-scale deviations of a specific implementation from the specification can contribute downstream to significant data corruption occurring faraway from the point of data entry. Due to the presence of the "amplification effect", it becomes difficult to attribute or

fix vulnerabilities, as each compliant implementation of significant parsers will also creates exploitable parser differences [4]. It is necessary to develop countermeasures that move beyond the security analysis of one system at a time to the security of functionally integrated systems or networks of systems [14].

The picking of apparently random data on the health cloud shows that the healthcare cloud security issue may be bigger. Adoption of healthcare cloud technology is faster than the implementation of security governance. The near absence of SMART-on-FHIR uptake on FHIR servers may be due to two reasons. There are 2 reasons for a cloud-based SMART-on-FHIR implementation failure: one, a high bar of SMART-on-FHIR implementation in public cloud , or two, wrong assumption of security of cloud deployment by healthcare organization. Healthcare organizations wrongly assume that the network perimeter controls security and compliance of the cloud service. Consequently, they evade authentication in the interoperability interface at the application layer. An attacker can collect health data on a large scale by.

Strategies for correcting the weakness in question include both technical and procedural approaches. The technical solutions might have stricter implementation rules for the parsers, a differential detection system at the runtime level, and improved verification of the integrity of exchanged data. The procedural ones may well include better implementation certification requirements, stricter continuous interoperability testing requirements, and assisted, security-aware specification development. The FHIR. The garden framework can provide the basis for automated differential detection and resolution in the development pipeline and on the deployment sites, detecting the differences in the parsers prior to production deployment.

The current regulation creates a gap in oversight of healthcare software like AI. These frameworks typically deal with functional requirements rather than properties of interoperability security. The premarket approval processes in the US, for instance. The FDA looks at interoperability for medical devices. Nonetheless, parser consistency has no interoperability requirements of any sort. The National Coordinator for Health Information Technology office creates a certification framework that relies on conformance with standards. However, there is no evaluation of security differentials. Through requirements on conformity assessment and post-market surveillance of interoperability, these regulatory frameworks could further help mitigate the risk of parser differentials.

Further exploration will explore additional FHIR implementation variants. Many different aspects of FHIR evolve together: " The FHIR specification generally describes a set of related resource types, and implementers often evolve their implementations together". To longitudinally quantify broader differential evolution effects, research must coordinate a specification over the versions of a specification to develop a longitudinal analysis methodology. The extra effort will also contemplate the merging of the workflow with methodologies analyzing the clinical impact of parser differencing results.

The proposed architecture referred to as FHIR Garden is designed to be extended easily to new language pairs that is, to other pairs of FHIR implementation variants such as HL7 v2.

6. Conclusion

This study demonstrates that parser differentials across distributed FHIR implementations introduce serious vulnerabilities in modern health information exchange networks. Using the FHIR Garden testbench and a differential fuzzing methodology, we systematically compared seven implementations spanning five programming languages and design paradigms [4]. Our analysis uncovered 59 distinct parser differentials, categorized into numeric precision, Unicode encoding, syntactic error recovery, and structural transformation issues. These differentials can be chained to construct exploits that achieve

Table 2: Parser Differential Categories and Impact

Category	Count	Impact	Key Issues
Numeric	18	High	Precision loss, overflow
Encoding	14	Medium	Unicode, surrogate pairs
Syntactic	16	Low	Error recovery, malformed JSON
Structural	11	High	Format conversion, reorganization

selective data transmission, corruption, or disclosure in realistic federated exchange scenarios [5].

The clinical impact extends beyond security: precision loss in timestamps, mutation of vaccine descriptions, and decimal truncation can compromise data integrity and patient safety. Moreover, a reconnaissance of 1,089 publicly accessible FHIR servers revealed minimal enforcement of SMART-on-FHIR authentication, exposing sensitive patient data in cloud environments.

To mitigate these risks, we recommend:

- Stricter parser implementation guidelines that reduce semantic ambiguity.
- Runtime differential detection systems to flag inconsistencies during operation.
- Enhanced implementation certification and continuous interoperability testing.
- Integration of security-aware specifications into regulatory frameworks (e.g., FDA premarket approval, ONC certification).

The FHIR Garden framework provides a foundation for automated differential detection in development pipelines and production deployments, enabling proactive identification of parser inconsistencies before they can be exploited. Future work will extend this methodology to additional standards (e.g., HL7 v2) and incorporate longitudinal analysis to track parser evolution across FHIR versions.

6.1 Ethical Considerations

This study was conducted using only synthetic and de-identified data. Infrastructure assessments were limited to non-intrusive methods, and no production systems were targeted. Vulnerabilities identified in open-source implementations were disclosed responsibly. All procedures adhered to established ethical guidelines for security research.

7. Future Research Directions

The findings of this study reveal critical vulnerabilities arising from parser differentials in distributed FHIR implementations. Future work should focus on integrating advanced computational methods to

enhance detection, mitigation, and prevention of such security flaws. Several promising directions emerge from related research in machine learning, system optimization, and data integrity verification [15].

First, leveraging deep learning architectures for *automated differential detection* could improve real-time anomaly identification. Techniques such as the graph-based analysis of high-dimensional data structures [8] could be adapted to model parser behavior across FHIR servers, flagging inconsistencies in data processing. Similarly, recent advances in adversarial robustness could be employed to learn robust representations of valid versus malicious health records, enhancing the discriminative capacity of security monitors.

Second, *scalable data management frameworks* like container-based orchestration systems [6] could be extended to health data exchanges, enabling efficient handling of multimodal clinical data while ensuring parser consistency across large-scale deployments. Integrating such systems with real-time analytics platforms would support continuous integrity verification in live health networks.

Third, *adaptive system architectures* inspired by modern fuzzing frameworks could be developed to dynamically reconfigure FHIR server deployments based on workload and threat indicators, minimizing exploit windows. Coupled with differential analysis methods [4], such systems could proactively identify and isolate compromised parsers before clinical data is corrupted.

Fourth, *optimization and normalization techniques* from machine learning research could enhance parser stability. For instance, methods for semantic consistency verification [12] could be adapted to improve numerical precision handling in FHIR parsers, reducing floating-point and overflow errors. Likewise, trustworthy federated learning approaches could be used to coordinate parser parameter updates across distributed healthcare networks.

Finally, longitudinal and large-scale studies are needed to track parser differential evolution across FHIR versions. Frameworks for continuous security assessment in healthcare [15] could be repurposed to analyze historical parser changes and predict future vulnerabilities. Such efforts would align with regulatory needs for ongoing interoperability surveillance and certification.

By bridging gaps between health informatics, machine learning, and systems security, future research can transform interoperability testing from a compliance activity into a proactive, intelligence-driven safeguard for global health data exchange.

References

[1] K. D. Mandl, D. Gottlieb, and A. Ellis. A federated network for clinical data sharing: The Accumulus consortium. *Journal of the American Medical Informatics Association*, 30(5):823–831, 2023. doi: 10.1093/jamia/ocad021.

[2] M. Lehne, J. Sass, A. Essenwanger, J. Schepers, and S. Thun. Why digital medicine depends on interoperability. *npj Digital Medicine*, 4(1):120, 2021. doi: 10.1038/s41746-021-00486-1.

[3] P. Esmaeilzadeh, S. Dharanikota, and T. Mirzaei. The role of patient engagement in patient-centric health information exchange (hie) initiatives: An empirical study in the united states. *Information Technology & People*, 37(2):521–552, 2024.

[4] B. Jabiyev, A. Gavazzi, K. Onarlioglu, and E. Kirda. Gudifu: Guided differential fuzzing for http request parsing discrepancies. In *Proceedings of the 27th International Symposium on Research in Attacks, Intrusions and Defenses*, pages 235–247, 2024.

[5] D. Klein and M. Johns. Parse me, baby, one more time: Bypassing html sanitizer via parsing differentials. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 203–221. IEEE, 2024.

[6] A. J. McMurry, D. I. Gottlieb, T. A. Miller, J. R. Jones, A. Atreja, J. Crago, and K. D. Mandl. Cumulus: A federated electronic health record-based learning system powered by fast healthcare interoperability resources and artificial intelligence. *Journal of the American Medical Informatics Association*, 31(8):1638–1647, 2024.

[7] T. Y. Zhuo, Z. Li, Y. Huang, F. Shiri, W. Wang, G. Haffari, and Y. F. Li. On robustness of prompt-based semantic parsing with large pre-trained language model: An empirical study on codex. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1090–1102, May 2023.

[8] D. Jain. Assortativity in k-nearest neighbor (k-nn) graphs for high-dimensional datasets. Zenodo, 2025.

[9] H. Zhang, W. Zhang, Y. Feng, and Y. Liu. Svscanner: Detecting smart contract vulnerabilities via deep semantic extraction. *Journal of Information Security and Applications*, 75:103484, 2023.

[10] J. Prieto, J. M. Corchado, and J. F. De Paz. Security challenges in iot-based e-health systems. *Journal of Network and Computer Applications*.

[11] H. R. Strasberg, B. Rhodes, G. Del Fiol, R. A. Jenders, P. J. Haug, and K. Kawamoto. Contemporary clinical decision support standards using health level seven international fast healthcare interoperability resources. *Journal of the American Medical Informatics Association*, 28(8):1796–1806, 2021.

[12] A. Akhmetov, Z. Latif, B. Tyler, and A. Yazici. Enhancing healthcare data privacy and interoperability with federated learning. *PeerJ Computer Science*, 11:e2870, 2025.

[13] C. N. Vorisek, M. Lehne, S. A. I. Klopfenstein, P. J. Mayer, A. Bartschke, T. Haese, and S. Thun. Fast healthcare interoperability resources (fhir) for interoperability in health research: Systematic review. *JMIR Medical Informatics*, 10(7):e35724, 2022.

[14] P. Tabari, G. Costagliola, M. De Rosa, and M. Boeker. State-of-the-art fast healthcare interoperability resources (fhir)-based data model and structure implementations: Systematic scoping review. *JMIR Medical Informatics*, 12(1):e58445, 2024.

[15] P. Esmaeilzadeh. Identification of barriers affecting the use of health information exchange (hie) in clinicians' practices: An empirical study in the united states. *Technology in Society*, 70:102007, 2022.