# Automating the Technical Validation of FHIR Implementation Guides: A Data-Oriented Framework Within the HfPT Project

Ivo A. Rodrigues[1,2], Leopoldo O. Silva[1,2(✉)], Cátia Pereira[4], Rute Almeida[3], Rui C. Ribeiro[1,2], and Francisco J. Duarte[1,2]

[1] CCG/ZGDV Institute, University of Minho, Azurém Campus, 4800-058 Guimarães, Portugal
leopoldo.silva@ccg.pt
[2] ALGORITMI Research Centre, University of Minho, 4710-057 Braga, Portugal
[3] RISE-Health, Faculty of Medicine, University of Porto, Alameda Prof. Hernâni Monteiro, 4200-319 Porto, Portugal
[4] Healthy Systems, Rua Engenheiro Ferreira Dias, 728 - 40, 4100-246 Porto, Portugal

**Abstract.** Several HL7 FHIR Implementation Guides were developed under the "Health From Portugal" initiative, particularly with Project A4.4.3 in mind, to address particular healthcare concerns in Portugal. Maintaining the accuracy and clarity of these Implementation Guides is crucial to guarantee compatibility at the national level. Still, conventional manual validation techniques usually take a long time, are prone to errors, and are difficult to scale. This paper discusses the creation and execution of an automated system meant for technical validation and certification assistance for FHIR Implementation Guides. We improved the execution of the official Java FHIR validator on large datasets by using a Python-based orchestration layer, which turned difficult *OperationOutcome* results into simple, well-organized formats. Effective gathering of error measurements and formulation of useful recommendations to improve IG are possible with this approach.

**Keywords:** HL7 FHIR · Software Engineering · Automated Quality Assurance · Workflow Automation · Data-Oriented Analysis · Interoperability Standards

## 1 Introduction

The digital change in healthcare systems depends greatly on the smooth sharing of clinical and administrative data across various platforms [1]. In this setting, the HL7 FHIR (Fast Healthcare Interoperability Resources) standard [2] has

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026, LNCS 16767, pp. 340–356, 2027.
https://doi.org/10.1007/978-3-032-30491-9_22

become the global common standard for healthcare compatibility, providing a flexible system based on current web technologies. However, the natural flexibility of FHIR requires the development of Implementation Guides (IGs), which are groups of profiles, terminologies, and business rules that tailor the basic standard to particular clinical, administrative, and national needs.

In Portugal, the national interoperability strategy, set by Portaria no 137/2017 [3], has seen significant progress due to the joint efforts of the Shared Services of the Ministry of Health (SPMS) [4] and the HL7 Portugal group [5]. A key achievement in this process is the "Health From Portugal" (HfPT) initiative [6], which has promoted the creation of a strong collection of IGs. These guides utilize the National FHIR Baseline (PT Core) [7] as their foundational semantics, broadening it to include particular situations such as monitoring chronic illnesses and managing hospital workflows. To promote the widespread adoption of these standards, it is essential to establish a validation system based on data that identifies any non-compliance at an early stage, ensuring that the Implementation Guides are technically prepared before undergoing the official HL7 certification process.

Making sure these Implementation Guidess and the data they generate are technically and mathematically correct can help us to better understand complex systems and improve decision-making. It is a tough work to be semantically right [8]. Historically, validation techniques have relied on unit-level testing or manual checking, which are sluggish, suspicious vulnerable to human errors and challenging to implement over the complicated architecture of systems based on FHIR [9]. Though the official HL7 validation tools frequently produce extremely technical and thorough results like the FHIR resource of OperationOutcome. This results can be difficult to be used and understood for non-expert stakeholders [10]. This causes a major delay in the technical certification process of national standards.

This paper outlines an automated system that links knowledge-based validation with a data-driven analysis layer, enabling systematic certification of FHIR standards in the HfPT project. By using a Python-driven orchestration layer on top of the HL7 Java Validator, the suggested solution allows for automated batch processing of resources across various FHIR versions and IG updates. The framework converts raw validation logs into organized, simplified formats, facilitating the systematic collection of error metrics and the creation of useful recommendations. Through the examination of two separate national use cases, the resp-ig [11] and the workflow-ep-ig [12], we show how this automated method improves the quality, consistency, and scalability of the FHIR certification process in Portugal.

## 2 Previous Work

The robust use of HL7 FHIR as an international standard for interoperability in healthcare is fundamentally based on its ability to maintain data uniformity through profiles and implementation guidelines [13]. However, the mere existence

of profiles and implementation guidelines does not guarantee the much-desired easy sharing of data. To achieve this, very rigorous validation methods are needed to reduce the gap between theoretical definitions and real-world applications [14]. Thus, this section presents a brief overview of the current technological environment regarding FHIR validation tools, existing techniques for ensuring compliance with implementation guidelines, and the specific problems faced by large-scale digital health projects in dealing with national standards with multiple versions. By reviewing the official HL7 validation system and its incorporation into third-party libraries, we establish the necessary basis for understanding the shortcomings of current manual methods and the need for automated coordination.

At the center of the FHIR conformance framework is the official HL7 FHIR Validator CLI [15], a Java-based standard tool made to conduct thorough technical evaluations of resources against basic specifications and specific profiles. This tool functions by importing Implementation Guide packages and terminology resources to evaluate four key aspects: structural integrity, cardinality constraints, value set bindings, and complex business rules represented via FHIRPath invariants. Although the HL7 FHIR Validator CLI is the best option for accuracy, the ecosystem is further improved by important reference libraries, particularly [16] and the Firely SDK for .NET systems [17]. These libraries feature integrated validation systems that are typically part of server-side configurations to ensure data compliance in real-time. Nonetheless, while they are quite effective on their own, these tools primarily serve as fundamental elements that require considerable manual configuration when used in extensive multi-IG settings. This often results in complex *OperationOutcome* reports that are difficult to merge without sophisticated coordination [18].

Even with the progress made in atomic validation methods, there remains a significant gap in the architecture, especially concerning the coordination of these tools for diverse and extensive datasets. Existing frameworks are mainly designed for checking individual units within transactional settings. However, there is insufficient direct assistance for managing structured batch processes that can accommodate various versions and different implementation guidelines at the same time. Additionally, the results generated by these systems are typically located within the common FHIR *OperationOutcome* resource [18], which is fundamentally complex and very technical. Although this aspect is crucial for communications between machines, it presents a major difficulty for stakeholders without specialized knowledge, like clinical implementation guides architects and certification organizations. [10]. At present, there is an evident deficiency in all-encompassing solutions that can unify these distinct outcomes into ongoing statistical reports or translate intricate error logs into simpler, actionable advice for programmers. Consequently, a more sophisticated framework is required to address this shortcoming, converting basic validation information into strategic insights that enhance a thorough quality assurance and certification procedure for nationwide interoperability standards.

Driven by the combined efforts of the Ministry of Health's Shared Services (SPMS) [4] and the local HL7 Portugal (HL7 PT) chapter [5], the evolution of Portugal's digital health landscape reveals a conscious desire to match international standards. From isolated pilot programs, the national interoperability plan has grown over the last ten years into a well-organized governance system directed at developing accepted Implementation Guides [4]. This development led to the establishment of the National FHIR Baseline (PT Core) [7], which acts as a common foundation for developing extensions and profiles customized to the particular administrative and clinical requirements of the Portuguese healthcare system. More sophisticated and field-specific Implementation Guides, such as those for hospital admission procedures [12] and chronic respiratory care [11], have been developed as these standards have improved to meet the complex needs of the national health programs. Still, the increasing number and complexity of these Portuguese profiles underline the need of a single certification organization [4]. Without a regular, automated validation system, the rapid expansion of the national IG directory might cause inconsistencies between application-level messages exchanged within public and private healthcare settings and clinical criteria.

## 3 Research Methodology

The approach used for this study is based on the Design Science Research Methodology for Information Systems (DSRM-IS) [19], which is a framework tailored for developing and assessing new IT artifacts aimed at addressing intricate issues within organizations. The first step began with identifying a significant operational challenge in the HfPT project: the existing manual FHIR validation processes have scalability and consistency issues. These problems hinder the rapid certification of national health standards. After this was identified, the main objective of the proposed solution was defined as developing an automated, data-driven system capable of performing complete and repeatable validations. Finally, it was also desired that this system would allow complex technical records to be converted into organized and useful information for various stakeholders.

The design and development phase concentrated on building a modular artifact that combines the canonical HL7 Java Validator CLI with a Python-based orchestration layer. This design was developed particularly to manage the ingestion of several Implementation Guide versions and different datasets, therefore guaranteeing the normalization of raw validation outputs into a single central data repository. The framework was used across two different national use cases, resp-ig and the workflow-ep-ig, to show how the artifact can be used in the real world. It looked at both internal reference examples and synthetic application-level messages.

Using automated pivot tables and statistical grouping, the artifact was assessed using a quantitative analysis of the created output to ascertain the framework's effectiveness in detecting structural, semantic, and syntax faults

throughout several FHIR versions. This methodical analysis validates that the artifact not only greatly lowers the time needed for IG certification but also raises the overall quality of the interoperability criteria by revealing repeated failure patterns often missed in conventional manual processes. Following the iterative cycles of the DSRM-IS, this study guarantees that the derived automation framework contributes significantly and robustly to the digital health governance infrastructure in Portugal.

## 4 The Proposed Automation Framework

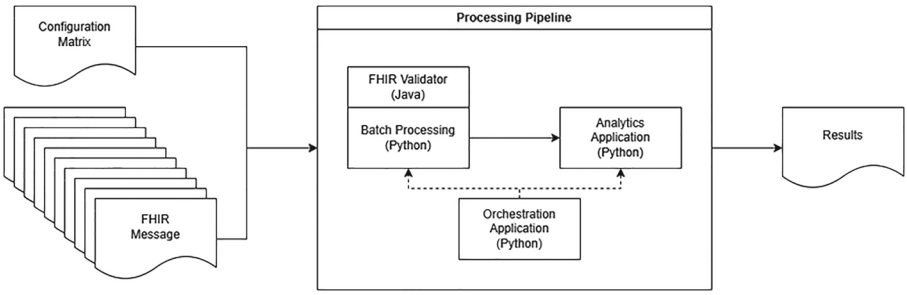


**Fig. 1.** Architectural workflow of the automated framework, illustrating the hybrid integration between the canonical Java-based validator and the Python-based components for orchestration, batch processing, and data analytics.

The core output of this research is the architecture of the proposed automation framework, created to deliver a scalable, repeatable, and data-driven quality assurance (QA) process for interoperability standards based on FHIR. This architecture acts as a high-level orchestration layer bridging the strategic needs of national digital health governance and low-level, atomic validation engines. The solution allows for the methodical intake of many datasets, from internal Implementation Guide (IG) reference examples to external application-level messages, across many FHIR versions and IG revisions by combining the canonical HL7 Java Validator CLI [15] with a flexible Python-based processing engine. The framework's modular architecture guarantees that the validation process is now a continuous and automated pipeline able to turn raw technical output, like the intricate *OperationOutcome* resource [18], into organized, multi-dimensional analytical reports instead of a fragmented, manual effort. The framework guarantees the technical readiness of interoperability profiles inside the HfPT project [6] by enabling the identification of repeated failure patterns and the creation of standardized maturity indicators via this orchestration.

This architecture acts as a high-level orchestration layer as it connects the strategic needs of national digital health governance and low-level atomic validation mechanisms. Thus, this solution allows for the methodical introduction

of many datasets, from internal reference examples from the Implementation Guide (IG) to external application-level messages, across various FHIR versions and IG revisions, combining the canonical HL7 Java Validator CLI [15] with a flexible Python-based processing mechanism. The modular architecture of this framework ensures that the validation process is now a continuous and automated pipeline capable of transforming raw technical output, such as the intricate *OperationOutcome* resource [18], into organized multidimensional analytical reports, rather than a manual and fragmented effort. The framework ensures the technical readiness of interoperability profiles within the HfPT project [6] as it allows for the identification of the most frequently recurring failure patterns and the creation of standardized maturity indicators through this orchestration.

The framework's system architecture is built on a hybrid integration model (Fig. 1), which combines the high-level flexibility of a Python-based orchestration layer with the specialized compliance of the Java-based validator. This architecture guarantees a clear division of responsibilities: the canonical HL7 Java engine manages the fundamental validation logic, while specialized Python components handle orchestration, batch management, and data analysis. A Python controller inside this design controls the execution environment by dynamically creating the command-line arguments needed to call the Java Archive (JAR) file of the validator. This programmatic invocation enables the exact definition of target FHIR versions, the dynamic loading of pertinent IG packages from both local and online repositories, and the mapping of several input data routes. The orchestrator allows for the methodical processing of varied message batches by abstracting the fundamental complexity of the Java execution environment. The framework therefore keeps perfect alignment with the official HL7 validation logic while acquiring the essential ability to automate sophisticated testing cases, such as multi-version comparisons, which are inherently challenging to handle via the native CLI alone.

The normalization process for data serves as an essential bridge that transforms unprocessed and diverse validation outcomes into a clear and organized analytical framework. Central to this operation is the intake of the FHIR *OperationOutcome* resource, which, while adhering to standards, is both lengthy and complex, thereby often hindering straightforward long-term or statistical evaluations. The structure employs a mapping technique to sort and organize essential diagnostic elements, including severity grades, error messages, and FHIRPath positions, into a streamlined and basic JSON schema (Table 1) to address this issue. Each issue is distinctly associated in this format with its corresponding input factors, which include the specific FHIR edition, the IG URL, and the identifier for the original message. In the final stage of the process, these standardized JSON objects are consolidated into table-based datasets to facilitate the creation of multi-dimensional pivot tables. These tables provide a high-level overview of compliance trends and allow for the methodical grouping and counting of issues by category and version, so enabling the automated identification of recurring abnormalities across huge datasets, a task that would be virtually unattainable by manual log inspection.

**Table 1.** Mapping of diagnostic fields.

| Original Field (FHIR) | Simplified Field | Analytical Objective |
|---|---|---|
| `issue.severity` | `severity` | Enables direct filtering by impact (e.g., Error vs. Warning). |
| `issue.code` | `issue_type` | Provides a standardized technical categorization of the failure. |
| `issue.details.text` | `description` | Translates low-level system logs into human-readable error descriptions. |
| `issue.expression` | `location` | Maps the error to the exact FHIRPath location within the resource. |
| *(Metadata)* | `fhir_version` | Allows for cross-version longitudinal analysis (e.g., R3 vs. R4). |
| *(Metadata)* | `ig_url` | Ensures traceability of the specific Implementation Guide version tested. |

**Table 2.** Example of the structured configuration matrix used to orchestrate simultaneous validation across multiple FHIR versions and implementation guide revisions.

| Test ID | FHIR Version | Implementation Guide URL | Input Data Path |
|---|---|---|---|
| `resp-v3` | 3.0.0 | https://hl7-pt.github.io/resp-ig/ | `./data/v3/resp-ig/` |
| `resp-v4` | 4.0.0 | https://hl7-pt.github.io/resp-ig/ | `./data/v4/resp-ig/` |
| `wf-ep-v4` | 4.0.0 | https://hl7-pt.github.io/workflow-ep-ig/ | `./data/v4/workflow/` |
| `pt-core` | 4.0.0 | https://hl7-pt.github.io/core/ | `./samples/core/` |

The batch processing approach is based on a matrix-driven methodology that lets one evaluate many interoperability situations at once in one execution cycle. The framework can organize validation across many test cases without needing hand adjustment by employing a structured configuration matrix (Table 2), which includes the particular mapping of FHIR versions, IG URLs, and target resource locations. Particularly important for controlling the shift between regular versions is this design; this is shown by the comparative analysis between versions 3.0.0 and 4.0.0 of the national IGs. Moreover, the approach is tuned to manage large datasets including outside application-level messages generated by healthcare software providers as well as the internal reference examples supplied in the guidelines. The framework ensures that every resource is evaluated within its particular environment while maintaining rigorous isolation between different test runs by automating the repetitive execution of the validation engine. This methodical approach essentially reduces the possibility of human error in parameter selection and offers a reliable, scalable basis for the longitudinal quality assurance of the growing portfolio of interoperability standards created under the HfPT project.

## 5 Case Studies: Domain Specifics

Within the HfPT initiative [6], different Implementation Guides (IGs) were developed [6] and two of these were chosen as the main examples to highlight the adaptability and resilience of our proposed structure. Within the Portuguese digital health ecosystem, these IGs represent various architectural functions, and thus one focuses on specialized clinical monitoring and the other on the structural management of patient workflows. Therefore, these two cases were chosen because they complement each other, as they allow an evaluation of the structure in relation to horizontal criteria, focused on the structure, and vertical needs, intensive in clinical terms, thus offering a complete testing ground for the diverse interoperability demands of the national health system.

The *resp-ig* [11] is a vertical clinical specification that aims to assist in the screening and treatment of chronic respiratory diseases. From a technical standpoint, this Implementation Guide has a high level of clinical complexity as it extensively uses specialized profiles of observation, condition, and medication administration resources. This guide also includes several customized extensions and links elements to certain medical vocabulary systems, such as LOINC and SNOMED-CT, along with unique coding systems created for Portuguese clinical research projects, in order to capture the subtleties of respiratory pathology. However, validating this Implementation Guide presents a significant challenge because it employs sophisticated questionnaire resources that require strict adherence to FHIRPath invariants and conditional logic to ensure data quality at the point of care. By using the automation framework of *resp-ig*, we hope to confirm the semantic integrity of clinical data and the correct execution of specialized extensions vital for long-term monitoring of patients in respiratory medicine.

The *workflow-ep-ig* [12] represents a standard for horizontal structure, while the previous case study was more focused on clinical aspects. Primarily concentrating on the *Encounter* resource and related administrative metadata, this IG defines the fundamental architecture for guiding the patient's path across the healthcare system. The letter "ep" stands for "episódio", which serves as the primary anchor for all clinical and administrative data produced during a hospital stay or outpatient visit. Strictly specifying state transitions via encounter status codes, this IG technically stresses the integrity of resource referencing guaranteeing that episodes are appropriately linked to patients, practitioners, and organizations. The validation of *workflow-ep-ig* is essential for institutional management and billing systems as it ensures the consistent use of institutional identifiers and admission types throughout the national health network. The automation framework evaluates the IG's capacity to maintain structural coherence throughout these high-level administrative processes, therefore guaranteeing that the structural backbone of the digital health record stays strong and compliant.

Table 3 summarizes the scope of the validation study and thus categorizes the IGs by function and focus. This classification shows the framework's versatility in handling the IGs' structure, as one is highly specialized in the clinical field

**Table 3.** Characteristics and scope of the selected FHIR Implementation Guides.

| IG Identifier | Focus Type | Primary FHIR Resources | Clinical/Administrative Domain |
|---|---|---|---|
| resp-ig | Vertical | *Observation, Condition, Questionnaire, MedicationAdministration* | Chronic Respiratory Disease Monitoring |
| workflow-ep-ig | Horizontal | *Patient, Organization, Location, Encounter* | Patient Admission and Episode Management |

while the other deals with broad administrative standards within the national data infrastructure.

## 6 Results

By applying this automated structure to the IGs selected by us, we obtained a rich collection of metrics. For the first case, the *resp-ig*, a total of 9 resources were analyzed, and we observed that from version 3.0.0 to version 4.0.0 there was a 70.4% reduction in serious errors, decreasing from 71 to 21. In addition, we detected a 285% increase in informative messages, which increased from 20 to 77. This indicates a successful shift towards improved semantic validation after overcoming structural obstacles. These results, outlined in Table 4, underscore the development of the respiratory care profiles within the HfPT initiative [6].

The performance indicators outlined in Table 4 show a change in the level of issues that marks the development of respiratory care profiles. This evolution from major structural obstacles to data confirmations indicates that as the Implementation Guide becomes more stable, the validation engine can perform more thorough semantic evaluations. This trend is further demonstrated in Table 4, where the notable decrease in errors is visually contrasted with the increase in informational messages, indicating a readiness for extensive interoperability testing.

**Table 4.** Quantitative validation metrics for the Respiratory Care IG.

| FHIR version | Total Resources* | Total Issues | Errors | Warnings | Information |
|---|---|---|---|---|---|
| 3.0.0 | 9 | 101 | 71 | 10 | 20 |
| 4.0.0 | 9 | 112 | 21 | 14 | 77 |
| *Change (%)* | 0% | +10.9% | -70.4% | +40% | +285% |

*Canonical reference resources defining the resp-ig specification profiles.

A corresponding and even more extreme development was noted in *workflow-ep-ig*, as shown in Table 5. In the six message groups evaluated, the framework

achieved complete resolution of major structural faults, decreasing from an average of 13 errors per message in version 3. 0. 0 to none in version 4. 0. 0. Nevertheless, while the severe errors disappeared, the overall number of warnings remained high, reaching a total of 177 issues in version 4. 0. 0. These numerical findings indicate that although administrative criteria reached structural accuracy more quickly than clinical guidelines, both eventually hit a point where the emphasis of validation changed from simple syntax adherence to more intricate relational and terminology standards.

**Table 5.** Quantitative validation metrics for the Workflow IG.

| FHIR version | Total Resources* | Total Issues | Errors | Warnings | Information |
|---|---|---|---|---|---|
| 3.0.0 | 6 | 237 | 78 | 159 | 0 |
| 4.0.0 | 6 | 177 | 0 | 177 | 0 |
| *Change (%)* | 0% | −25.3% | −100% | +11.32% | +0% |

*Canonical reference resources defining the workflow-ep-ig specification profiles.

The classification of the total 627 identified problems offers a detailed perspective on the technical difficulties present in various areas. Structural failures showed the most significant change in both data sets. In the *resp-ig*, these dropped from 41 to 1, while in the *workflow-ep-ig*, the 78 structural mistakes present in version 3. 0. 0, which included absent mandatory components in the `MessageHeader` and unrecognized attributes, were completely removed in version 4. 0. 0. This progress demonstrates the success of the iterative development process in aligning the national IGs with fundamental FHIR profiles. On the other hand, syntax and semantic errors displayed different trends. Syntax mistakes, particularly invalid elements, continued to pose a significant challenge for the *resp-ig* ($n = 13$ in version 4. 0. 0), mainly due to non-compliant resource identifiers that included underscores. In contrast, *workflow-ep-ig* did not present syntax problems ($n = 0$ in version 4.0.0), but encountered notable semantic and terminological issues. Our framework logged 54 invalid code warnings and 144 not-found warnings, related to unresolved participant types and the absence of national *CodeSystems*, such as the `location-physical-type` identifier.

Figure 2 illustrates the transition from high structural non-compliance in the earlier version to a more refined profile in version 4.0.0. In this later iteration, structural and invariant errors were nearly eliminated, leaving only residual syntax (invalid) and terminology (code-invalid) challenges. This shift signals a maturation process where basic resource integrity is achieved, allowing the validation engine to surface more specific data quality issues that were previously obscured by architectural misalignments.

Our automated system detected a significant technical trend, which is the effect of the validation scope, as we can observe in Fig. 3 With this, we understand why version 4.0.0 of *resp-ig* generates a higher number of logs compared to the

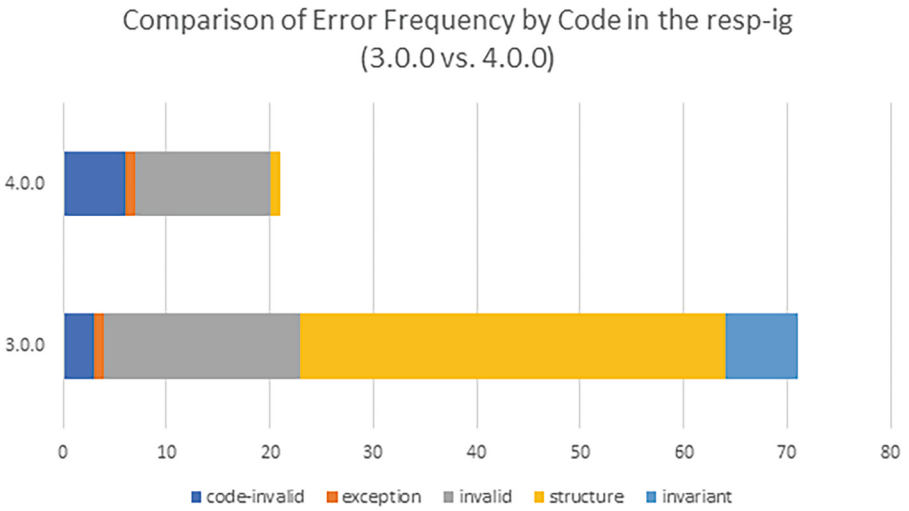


**Fig. 2.** Comparison of error frequency by issue category in the *resp-ig*.

previous version. The increase in logs is linked to the adoption of more detailed and complete business rules that identify inconsistencies that went unnoticed previously. In version 3.0.0, the frequent occurrence of structural errors ($n = 41$) is an obstacle and, therefore, often led the validation mechanism to prematurely interrupt processing a situation referred to as *short-circuit validation*. When the fundamental structure of an FHIR resource is faulty, the mechanism has difficulty in adequately evaluating more complex components, such as intricate slices, specific extensions, or internal business rules. Concerning the *short-circuit validation* issue, our system now suggests a step-by-step method that focuses on fixing the structure first. We found that requiring the validator to overlook structural components or simulate absent elements might result in incorrect positives in the semantic and business-rule aspects. Instead, our tool supports this sequential approach by offering transparent and well-organized feedback regarding structural obstacles. This enables developers to systematically address fundamental issues, thereby progressively unlocking more complex validation layers in a safe and dependable way. The gradual unlocking of validation layers is clearly illustrated in Fig. 3, where correcting structural mistakes led to an increase in the identification of informative and semantic messages.

In comparison, the upgrade to version 4.0.0 resulted in a significant decrease in structural problems, falling to almost zero ($n = 1$). By addressing these key architectural challenges, the framework allowed the validator to perform a deeper and more comprehensive examination of the entire feature tree. This improvement in clarity led to a 285% increase in informative messages, from 20 to 77, which means successful slice identifications and confirmations of informative rules. This relationship offers concrete proof that the advancement of the Implementation Guide is not only focused on error reduction but also on

improving analytical examination. Our research indicates that a standard can be genuinely verified for semantic correctness and accurate exchange of clinical data only after achieving structural reliability.

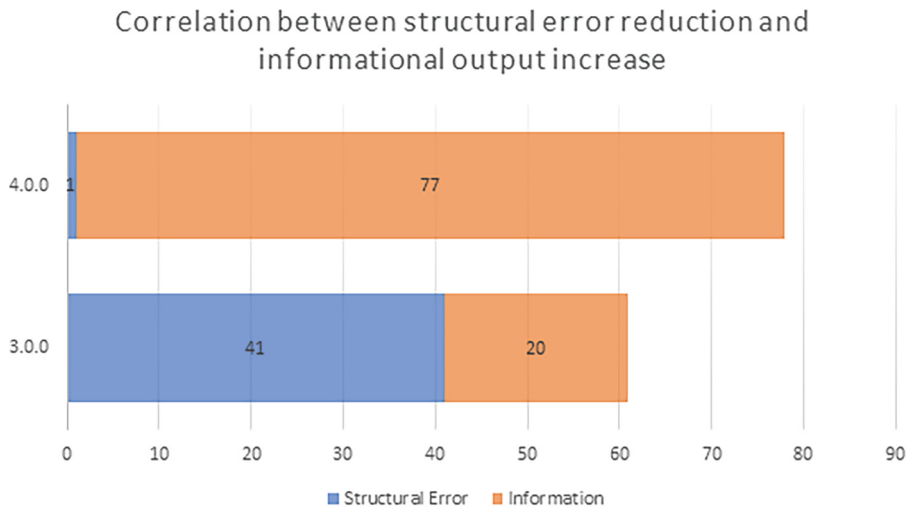


**Fig. 3.** Correlation between structural error reduction and informational output increase in the *resp-ig*.

Finally, a new aspect of referential integrity emerged as a significant obstacle for *workflow-ep-ig*. The system identified 138 information problems related to inaccessible resources, particularly *Encounter*, *Patient*, *Location*, and *Practitioner*, within the message packets. This observation reveals that, although each feature may conform in terms of structure, the logical connections between them in a clinical workflow still represent a vital area for technical checks. These findings imply that, as structural development progresses, the emphasis of the quality assurance process should shift to ensuring the consistency of the entire exchange of clinical messages, which is crucial for the practical application of national interoperability standards. These results are important as they show that FHIR resources, while they may be valid on their own, can encounter failures at the system level if the connections between them (such as between *Encounter*, *Patient* are disrupted. This underscores the significance of our orchestration layer, which extends beyond just validating individual resources to guarantee the overall functional integrity of the complete clinical message bundle.

The expanded analysis between the two interoperability guides offers a clear illustration of the development journey of national interoperability standards. Version 3.0.0 of both guides presented considerable structural interference, where fundamental architectural inconsistencies often masked more significant problems. However, when we examine the results of version 4.0.0, we detect a more refined and structurally sound environment. In *workflow-ep-ig*, we observe the

removal of all structural errors, which went from 78 to none. This indicates a state of technical readability for administrative headings. In *resp-ig*, by observing the almost complete elimination of persistent and structural errors, which went from 41 occurrences to just one, we realize that the internal clinical structure has become consistent.

Thus, this development highlights several aspects of non-compliance. *resp-ig* demonstrates stabilization in technical improvements, particularly with regard to naming conventions and internationalization codes, such as the use of `PTpt` tags. On the other hand, *workflow-ep-ig* indicates challenges in maintaining relational integrity, where resources may be synchronously correct but are not properly aligned in the message flow.

In short, our automation framework indicates that as an Implementation Guide evolves structurally, the main obstacle to technical certification changes. Thus, it is clear that both Implementation Guides have achieved a high level of structural quality, but the final phase of the quality assurance process requires specific contributions from the relevant area to ensure a smooth implementation within the Portuguese healthcare system.

## 7 Discussion and Actionable Recommendations

The main internal benefit of the framework is its capability to identify earlier failures in the reference models of the Implementation Guides. Our research indicates that both clinical and administrative guides necessitate different approaches for refinement to guarantee they are technically prepared. For the *resp-ig*, attention must be directed towards enhancing the technical aspects of the metadata. Ongoing syntax mistakes, like using underscores in identifiers such as `Communication.id`, reveal a flaw in manual quality assurance that could hinder the automated input into FHIR servers. In addition, the lack of clarity caused by the failure to address proprietary CodeSystems, like the *Inspirers* system, implies that IG authors need to verify that all terminology resources are either included in the IG package or registered with a national terminology server.

For the *workflow-ep-ig*, the emphasis on improvement turns to maintaining referential integrity. Even though version 4. 0. 0 reached a high standard of structure, the significant number of unattainable resource logs suggests that the logical arrangement of message groups is still not fully realized. IG creators ought to update the reference examples to confirm that all resources, such as *Patient*, *Practitioner*, and *Location*, are properly connected through the `MessageHeader.focus` or `Bundle.entry.request` elements, thus offering a consistent path for the validator [15].

From an external perspective, this system serves as a link for software developers, as it converts complex *OperationOutcome* records into simple compliance guides. By identifying unconventional language codes, such as the use of `PTpt` instead of the accepted `pt-PT`, along with problematic status transitions such as the `'not done'` condition in medication administration, gives developers clear directions for making necessary changes at the application level. In addition,

the framework emphasizes an essential need for grouping organization. Vendors adopting *workflow-ep-ig* are encouraged to extend their focus beyond the compliance of individual features to consider relational integrity. A validation report produced by this framework does not only indicate that a message has failed, but also identifies which specific link in the clinical workflow is faulty. In this way, we have an innovative approach that reduces the development burden on software developers. This helps ensure that messages at the application level are in strict compliance with national standards and thus promotes a more reliable digital health ecosystem.

The automation implemented through this system has transformed the assessment of the HfPT project's technical readiness from a time-consuming manual process to a structured governance model. The most notable effect is the speed of the feedback loop, where replacing manual checks with automated group verification allows national participants to analyze the effects of a single profile modification on various features in a matter of minutes. Thus, this system also offers an unbiased assessment of the maturity of national standards. The ability to visualize the transition from structural problems to semantic challenges allows project leaders to allocate specialized resources more strategically. A high level of structural inconsistencies in initial versions suggests the need for an architectural review, while continuous terminology alerts in subsequent versions indicate the need for deeper infrastructure changes, such as updates to terminology servers. In conclusion, this data-driven strategy ensures that the transition from a preliminary specification to a fully developed national standard is fast and technically sound. The main value of the proposed framework lies in its ability to convert granular technical records into strategic and actionable insights for diverse stakeholders.

**Table 6.** Actionable Recommendations for Implementation Guide resp-ig and Application Refinement

| Finding | Impact | Target | Recommendation |
|---|---|---|---|
| Underscores in Ids | Prevents ingestion in FHIR Servers | IG Authors | Replace _ with - in all resource identifiers. |
| Invalid Language Code | Breaks localization/search | Software Vendors | Standardize on BCP-47 tags (use pt-PT instead of PTpt). |
| Unresolved CodeSystems | Semantic validation failure | Project Governance | Bundle proprietary CodeSystems within the IG package. |
| Deprecated Extensions | Future interoperability risk | IG Authors | Replace `ordinalValue` extension with `itemWeight`. |
| Broken References | Data integrity loss | Software Vendors | Ensure Encounter resources always link to a valid Patient. |

As summarized in Table 6, the findings are categorized not only by their technical impact but also by the specific stakeholder responsible for remediation, thus bridging the gap between low-level validation and national digital health governance.

By differentiating between Implementation Guide authors, software vendors, and project oversight, the framework ensures that validation comments are sent to the correct organizational level. For example, issues such as outdated extensions or unusual syntax in identifiers are recognized as flaws that occur early in the process. When these issues are addressed at the Implementation Guide author level, it is possible to ensure that official reference examples can act as a reliable source of truth for the entire ecosystem. On the other hand, issues related to localization, such as BCP-47 codes, and relational integrity are directed to software vendors, offering a straightforward plan for necessary application-level modifications for technical readiness. Finally, when we talk about broad issues such as the challenge of dealing with proprietary terms like the *Inspirers* CodeSystem, these are identified as governance tasks. All these issues require comprehensive solutions, such as registering code systems in centralized national terminological databases. This specific method transforms FHIR validation from a reactive pass/fail system into a proactive quality improvement process that safeguards the long-term robustness of the Portuguese interoperability framework.

## 8 Conclusion

The main achievement of this research is the creation of a robust tool that transforms FHIR validation from a disorganized manual process into an automated, data-driven process aligned with the national certification schedule. By integrating the official HL7 validation mechanism through a Python automation layer, it's possible to achieve significant efficiency gains, allowing for the simultaneous processing of hundreds of resources across multiple versions in a much shorter time than traditional manual assessments. Additionally, the structure provided an outstanding degree of technical precision for the HfPT initiative [6]. As shown in the chosen examples for *resp-ig* and the *workflow-ep-ig*, the structure effectively pinpointed minor yet critical compliance problems, such as irregular resource IDs, wrong localization codes, and faulty reference links, which frequently go unnoticed in manual evaluations. Therefore, the dual approach emphasized the various growth challenges encountered by both clinical and administrative protocols, providing national parties with a straightforward way to verify a variety of interoperability models.

Although the system is very efficient, it does have drawbacks due to the existing global FHIR landscape. Its effectiveness and diagnostic precision are heavily linked to the standard Java HL7 validator CLI, which means it can be influenced by any lags in execution or variations in interpretation within the main framework. Additionally, the structure encounters a meaning gap concerning the resolution of terminology, as it has difficulty in addressing proprietary or specific CodeSystems, like the *Inspirers* system, without depending on outside terminology servers or local collections, resulting in a major hindrance. These relationships emphasize that although automation enhances the process, the overall success of technical validation is deeply connected to the presence and dependability of national terminology systems.

The subsequent stage of this study will concentrate on incorporating the automation framework into Continuous Integration and Continuous Deployment (CI/CD) processes. By integrating verification within the ongoing development workflow, IG creators and software providers can achieve built-in compliance, obtaining immediate feedback as they alter profiles or messages. Additionally, the framework will be expanded to include the complete range of the HfPT project, including the Implementation Guides for Adverse Drug Reactions (*adr-ig*), Neonatology (*obs-neonat-ig*), and Community Health (*community-health-ig*). This development strategy aims to establish the framework as a foundational element of the Portuguese health information governance approach, which supports the overarching objective of maintaining national standards that are strong, dependable, and genuinely interoperable.

**Acknowledgments.** Authors would like to acknowledge the project HfPT - Health from Portugal, co-financed by the RRP (Recovery and Resilience Plan) and the NextGenerationEU Funds: Notice 2022-C05i0101-02 Mobilizing Agendas/Alliances for reindustrialization, Project no C630926586-00465198.

R.A. acknowledges support from Fundação para a Ciência e a Tecnologia, I.P., under the Scientific Employment Stimulus - FCT Tenure (Grant 2023.13694.TENURE.056).

## References

1. Sinaci, A.A., et al.: A data transformation methodology to create findable, accessible, interoperable, and reusable health data: Software design, development, and evaluation study. J. Med. Internet Res. **25**, e42822 (2023). https://doi.org/10.2196/42822
2. HL7 International: HL7 FHIR - Fast Healthcare Interoperability Resources Overview (2026). https://www.hl7.org/fhir/overview.html. Accessed 18 Mar 2026
3. Ministério da Saúde: Portaria no. 137/2017: Define os princípios e as normas de interoperabilidade de dados e sistemas de informação de saúde." Diário da República no. 75/2017, Série I, 2017. https://dre.pt/dre/detalhe/portaria/137-2017-106927891
4. SPMS - Serviços Partilhados do Ministério da Saúde: SPMS - Serviços Partilhados do Ministério da Saúde - Home (2024). https://www.spms.min-saude.pt/. Accessed 18 Mar 2026
5. HL7 Portugal: HL7 Portugal - Health Level Seven International (2024). https://www.hl7.pt/. Accessed 18 Mar 2026
6. Health from Portugal: Health from Portugal (HfPT) - Agenda Mobilizadora para o Setor da Saúde (2024). https://www.healthfromportugal.pt/pt/. Accessed 18 Mar 2026
7. HL7 Portugal: HL7 Portugal FHIR National Baseline (PT Core) (2024). https://hl7-pt.github.io/core/. Accessed 18 Mar 2026
8. Leroux, H., Metke-Jimenez, A., Lawley, M.: Towards achieving semantic interoperability of clinical study data with fhir. J. Biomed. Semantics **8**, 41 (2017). https://doi.org/10.1186/s13326-017-0148-7

9. Walonoski, J., Scanlon, R., Dowling, C., Hyland, M., Ettema, R., Posnack, S.: Validation and testing of fast healthcare interoperability resources standards compliance: Data analysis. JMIR Med. Inform. **6**, e10870 (2018). https://doi.org/10.2196/10870
10. T. Benson and G. Grieve, *Principles of health interoperability: NOMED CT, HL7 and FHIR*.07 2016.https://doi.org/10.1007/978-3-319-30370-3
11. HL7 Portugal: Respiratory Care FHIR Implementation Guide (resp-ig) (2024). https://hl7-pt.github.io/resp-ig/. Accessed 18 Mar 2026
12. HL7 Portugal: Workflow Episode of Care FHIR Implementation Guide (workflow-ep-ig) (2026). https://hl7-pt.github.io/workflow-ep-ig/. Accessed 18 Mar 2026
13. Ayaz, M., Pasha, M.F., Alzahrani, M.Y., Budiarto, R., Stiawan, D.: The fast health interoperability resources (fhir) standard: systematic literature review of implementations, applications, challenges and opportunities. JMIR Med. Inform. **9**, e21929 (2021). https://doi.org/10.2196/21929
14. Nan, J., Xu, L.-Q.: Designing interoperable health care services based on fast healthcare interoperability resources: Literature review. JMIR Med. Inform. **11**, e44842 (2023). https://doi.org/10.2196/44842
15. HL7 International: Using the FHIR Validator (2024). https://confluence.hl7.org/spaces/FHIR/pages/35718580/Using+the+FHIR+Validator. Accessed 18 Mar 2026
16. Smile Digital Health: HAPI FHIR: The Open Source FHIR API for Java (2026). https://hapifhir.io/. Accessed 19 Mar 2026
17. Firely: Firely SDK for .NET architectures (2026). https://firely-team.github.io/firely-net-sdk/. Accessed 19 Mar 2026
18. HL7 International: FHIR OperationOutcome Resource Definition. Accessed 18 Mar 2026. https://www.hl7.org/fhir/operationoutcome.html (2026)
19. Peffers, K., Tuunanen, T., Rothenberger, M., Chatterjee, S.: A design science research methodology for information systems research. J. Manage. Inf. Syst. **24**, 45–77 (2007)