

## RESEARCH

# On the Use of Metaprogramming and Domain Specific Languages: An Experience Report in the Logistics Domain

Pedro HT Costa\*, Edna D Canedo and Rodrigo Bonifácio

## Abstract

In this paper we present the main design and architectural decisions to build an enterprise system that deals with the distribution of equipment to the Brazilian Army. The requirements of the system are far from trivial, and we have to conciliate the need to extract business rules from the source code (increasing software flexibility) with the requirements of not exposing the use of declarative languages to use a rule-based engine and simplifying the tests of the application. Considering these goals, we present in this paper a seamless integration of meta-programming and domain-specific languages. The use of meta-programming allowed us to isolate and abstract all definitions necessary to externalize the business rules that ensure the correct distribution of the equipment through the Brazilian Army unities. The use of a domain specific language simplified the process of writing automatic test cases related to our domain. Although our architecture targets the specific needs of the Brazilian Army, we believe that logistic systems from other institutions might also benefit from our technical decisions.

**Keywords:** Software Architecture; Generative Programming; Domain Specific Languages; Military Logistics

## Introduction

One of the logistics' phases to distribute equipment to an army force contemplates the *distribution planning* of materials throughout military unities. This phase is known as *material endowment*, and involves several steps, including the specification of relevant materials, often independent of vendor; the definition of rules that link military materials and equipment's to organizational unities (e.g., departments, regiments, and military posts); the execution of the rules for deriving the materials originally intended for each *abstract military unit*; and the designation of the materials that must be actually distributed to the *concrete military organizations*, which may suffer variations due to the inclusion and suppression of positions for a particular organizational unit of an army force. In the context of this paper, the reader might consider the scenario of the Brazilian Army.

Decision support systems are essential in this domain, given the importance and complexity of the processes related to the allocation of military employment materials. In particular, the rules used to predict

military materials and equipment are not trivial, and implementing these rules directly in the source code makes the systems difficult to understand and maintain. This problem mostly occurs because the number of conditions that need to be expressed is proportional to the complexity of the organizational structure and the number of materials. In the case of an entire army force, this number is noteworthy. In addition, a "*hard-coding*" solution does not allow domain experts to trivially specify and test new derivation rules. For these reasons, it is preferably to implement these rules declaratively, using languages, tools, or libraries that support some sort of inference mechanism.

Nevertheless, introducing either a logic language (e.g., Prolog or Datalog) or a specific library (such as Drools [1]) in a complex enterprise system may lead to some type of *architectural conflict*, particularly in environments whose whole ecosystem system is well established and based on a set of constraints related to both language and libraries usage. Furthermore, these solutions require a reasonable learning curve, which can hamper software maintenance activities.

This article describes the main design and architectural decisions adopted to implement the mechanisms that assist decision makers to plan the distribution of military materials and equipment across the organiza-

\*Correspondence: pedro.costa@aluno.unb.br  
Computer Science Department, University of Brasília - (UnB), P.O. Box 4466, 70910-900 Brasília, BR  
Full list of author information is available at the end of the article

tional structure of the Brazilian Army. These decisions aim to both (a) abstract the adoption of a business rule engine using **meta-programming** and (b) allow domain experts to simulate the definition and testing of rules using a **domain specific language**. Accordingly, this paper presents the following contributions:

- A description of the use of generative programming techniques to raise the abstraction level related to the adoption of a logic programming language in the specification of rules for the distribution of materials and equipment through organizational unities.
- A report of an empirical evaluation of our design and architectural decisions, reinforcing that they fulfill not only the needs of the domain specialists, but also existing technical constraints.

Although we discuss the distribution of materials and equipment for the military domain (which is far from trivial), we believe that the architectural decisions discussed here can be exploited to other logistics and endorsements scenarios in which it is also desirable to transparently introduce the support for rule based engines in enterprise systems.

We organized this article as follows. Section presents the concepts related to the domain of material logistics. Section presents the concepts related to business rules and rule based engines and Section presents some background information related to *generative programming*—techniques that we use to support the design and implementation of our approach. Section presents the main design and architectural decisions we take to develop the mechanism responsible for the distribution of materials. In Section we present the results of an empirical study that validate our approach. Finally, Section and Section relate our work with the existing literature and present some final remarks.

## Technical Background

### Military Logistics Systems

Military Logistics is the method used to implement military strategies and tactics as a method to gain competitive advantage [2]. Logistics is defined as the art of moving armies, as well as accommodating and supplying the military [3]. Strategy and tactics provide the framework for conducting military operations, while logistics provides the means. In the context of this paper, we are more interested in the supplying aspects of logistics, which refers to the set of activities that deals with the provisioning of the material of different classes (including vehicles, armaments, munition, uniform, and fuel) necessary to the military unities (including a rich organizational structure and a huge number of soldiers).

In order to facilitate the administration and control of materials and equipment, the Brazilian Army uses a catalog system that is similar to one defined by the North Atlantic Treaty Organization (OTAN) [4], which classifies items into groups and classes. The Brazilian Army catalog uses a classification that considers the *type*, *class*, and *family* of the materials, ignoring some details (such as specific brands and suppliers). More detailed information is considered by other areas of the logistic system. Several rules govern the identification of materials and equipment, to avoid duplication and simplify the process of decision making. A logistic system is the integrated set of organizations, personnel, principles, and technical standards intended to provide the adequate flow for supplying materials [5]. This is a complex system, and thus must be organized in different modules.

In this paper we present the main design and architectural decisions related to one of the modules (SISDOT) of the *Integrated Logistics System* (SIGELOG) of the Brazilian Army. SISDOT deals with the identification of the necessary materials and equipment (hereafter MEMs) for the organizational unities of the Brazilian Army. Each organizational unity is a combination of a set of *generic organizations of specific types*, which describes the expected number of military unities of an organization (including armies, divisions, brigades, and so on). The goal of SISDOT is to estimate the set of MEMs that should be assigned to a generic organization, producing one of its main products (named QDM). Based on a set of QDMs, SISDOT then derives another estimate of MEMs that should be assigned to a concrete organizational unity (a QDMP).

Before generating a QDM, it is necessary to catalog all relevant MEMs to a given class of materials and then specify *high-level rules* (HLRs) that relate MEMs to the generic organizational unities, considering different elements of the structure—from the type of the unity and sub-unities to the qualifications of a soldier. That is, it is possible to elaborate a HLR that states that a military ambulance would be assigned to all *drivers within medical unities*. Differently, it is also possible to state that an automatic rifle will be assigned to all *sergeant that has been qualified as a shooter*. There is some degree of flexibility for writing these rules, that might be from two different kinds: rules that consider the entire structure of the Brazilian Army and rules that refer to specific elements of the organizational structure. The former details rules that are more specific for a set of military qualifications. The second allows the generation of rules that might be reused through different unities.

Deriving QDMs and QDMPs is not trivial and its complexity is due to several factors, including the rich

organizational structure of the Brazilian Army, the huge number of existing functions and qualifications that must be considered, and the reasonable number of MEMs. The process might also vary depending on the class of the materials. Moreover, it is possible to specify the *high-level rules* in different ways, which in the end might interact to each other leading to undesired results (e.g., the same material being assigned several times to a given soldier).

Due to number of logical decisions, using an imperative language to implement the process of QDM generation is not an interesting choice. However, the SIGELOG architecture constraints the development to use the Java Enterprise Edition platform—so we had little space to introduce a more suitable language (such as Prolog) to generate QDMs. Other constraints guided the architectural decisions of SISDOT, including:

- Maintenance: the set of languages and libraries are limited to those already used in the SIGELOG development.
- Performance: considering a total of 1000 *high-level rules*, the system must be able to generate a QDM in less than 10 seconds.
- Testing: the technical decisions must not compromise the testability of the system, being expected a high degree of automated tests.

To deal with the first constraint, we decided to use a meta-programming approach that derives *low-level rules* for Drools (a rule based engine for Java) *on the fly*, from a set of *high-level rules* represented as domain objects that must exist in a database. In this way, it is not necessary to understand the language used to specify rules in Drools. Surely, although this approach increases flexibility and allows us to externalize the *low-level rules*, it might introduce some undesired effect on the performance of the system: we generate all low-level rules every time a QDM is built. We are using a cache mechanism and performance tests (as we discuss in Section ) shows that our architecture fulfills the second constraint. For the third constraint, we implemented a DSL that simplifies the specification and execution of test cases.

In the next sections (Section and Section ) we introduce some technical background that might help the reader to understand the design and architectural decisions we take while implementing SISDOT.

### Rule Based Systems

The process for deriving QDMs consists of applying a set of business rules that relates MEMs to the organizational structure of the Brazilian Army. In this paper, we named this set of business rules as *high-level rules*. A business rule is a compact, atomic, and well-formed

statement about an aspect of the business that can be expressed in terms that may be directly related to the business and its employees using a simple and unambiguous language that might be accessible to all stakeholders [6]. This type of structure is very important for organizations, as they need to deal more and more with complex scenarios. These scenarios consist of a large number of individual decisions, working together to provide a complex assessment of the organization as a whole [7].

Rule Based Systems (RBS), also known as production systems or expert systems, are the simplest form of artificial intelligence that help to design solutions that reason about and take actions from business rules (a feature that is interesting in the context of SISDOT), using rules such as knowledge representation [8]. Instead of representing knowledge in a declarative and static way as a set of things that are true, RBS represents knowledge in terms of a set of rules that tells what to do or what to accomplish in different situations [8]. RBS consist of two basic elements: a *knowledge base* to store knowledge and an *inference engine* that implements algorithms to manipulate the knowledge represented in the knowledge base [8, 9, 10]. The knowledge base stores facts and rules [11]. a fact is a usually static statement about properties, relations, or propositions [11] and should be something relevant to the initial state of the system [8]. A rule is a conditional statement that links certain conditions to actions or results [12], being able to express policies, preferences, and constraints. Rules can be used to express deductive knowledge, such as logical relationships, and thus support inference, verification, or evaluation tasks [11].

A rule “if-then” takes a form like “*if x is A then take a set of actions*”. The conditional part is known as antecedent, premise, condition, or left-hand-side (LHS). The other part is known as consequent, conclusion, action, or right-hand-side (RHS) [8, 12]. Rule-based systems works as follows [8]: it starts with a knowledge base, which contains all appropriate knowledge coded using “if-then” rules, and a working memory, which contain all data, assertions, and information initially known.

The system examines all rule conditions (LHS) and determines a subset of rules whose conditions are satisfied, based on the data available at the working memory. From this information set, some of the rules are triggered. The choice of rules is based on a conflict resolution strategy. When the rule is triggered, all actions specified in its then clause (RHS) are executed. These actions can modify the working memory, the rule base itself, or take whatever action that the system programmer chooses to include. This *trigger rule loop* con-

tinues until a termination criterion is met. This end criterion can be given by the fact that there are no more rules whose conditions are satisfied or a rule is triggered whose action specifies that the program should be finalized.

Drools is an example of a business logic integration platform (BLiP) written in the Java programming language and that fits some requirements of SISDOT (in particular, its integration with Java and the Wildfly application server). We use Drools as the *rule-based system* in the QDM generation. However, to abstract the use of a RBS, we generate the Drools rules using a meta-programming approach, a specific technique for *generative programming*.

### Generative Programming

Generative Programming (GP) is related to the design and implementation of software components that can be combined to generate specialized and highly optimized systems, fulfilling specific requirements [13]. With this approach, instead of developing solutions from scratch, domain-specific solutions are generated from reusable software components [14]. GP is applied on the manufacturing of software products from components in a system in an automated way (and within economic constraints), i.e., using an approach that is similar to that found in industries that have been producing mechanical, electronic, and other goods for decades [15]. The main idea in GP is similar to the one that led people to switch from Assembly languages to high-level languages.

In our context, we use GP to deal with two different aspects of SISDOT. First, we raise the level of abstraction by automatically generating low-level Drools rules from an existing set of *high-level rules* (previously defined by the domain experts). This is a type of *meta-programming* based on template engines—we use a template to generate a set of facts and rules (which are the fundamental mechanisms of logic programming) from domain objects. We also raise the level of abstraction to assist developers in the specification and execution of test cases, using a domain specific language and a set of customized libraries and tools.

### Metaprogramming and Template Engines

Metaprogramming concerns to the implementation of programs that generate other programs. In the context of SISDOT, we use a metaprogramming technique to generate programs that are interpreted by Drools. The main goal is to abstract the use of a specific language for implementing the *low-level rules* used to generate QDMs. There are many different approaches for metaprogramming, including preprocessing, source to

source transformations, and template engines. A template engine is a generic tool for generating textual output from templates and data files. They can be used in the development of software that requires the automatic generation of code according to specific purposes [16].

The need for dynamically generated web pages has led to the development of numerous template engines in an effort to make web application development easier, improve flexibility, reduce maintenance costs, and enable parallel encoding and development using HTML like languages. In addition to being widely used in dynamic web page generation, templates are used in other tasks, being one of the main techniques to generate source code from high level specifications. More recently, languages such as Clojure, Ruby, Scala, and Haskell adopted the use of template and *quasiquote* mechanisms as a strategy for code reuse and program generation (some of them considering type systems). A template is a text document that combines placeholders and linguistic formulas used to describe something in a specific domain [17].

In this work we use the FreeMarker template engine, which is an open-source software designed to generate text from templates [18]. FreeMarker was chosen for several reasons, such as: it follows a general purpose model; it is faster than other template engines; it supports shared variables and model loaders; and its templates can be loaded or reloaded at runtime without the need to (re)compile the application [16, 19].

### Domain Specific Languages

A language is a set of valid sentences, serving as a mechanism for expressing intentions [20]. Creating a new language is a time-consuming task, requires experience, and is usually performed by engineers specialized in languages construction [21]. To implement a language, one must construct an application that reads sentences and reacts appropriately to the phrases and input symbols it discovers. The need for new languages for many growing domains is increasing, as well as the emergence of more sophisticated tools that allow software engineers to define a new language with reasonable effort. As a result, a growing number of DSLs are being developed to increase developer productivity within specific domains [21].

Domain Specific Languages (DSL) are specification languages or programming languages with high level of abstraction, simple and concise [22], focused on specific domains, and are designed to facilitate the construction of applications, usually declaratively, with limited expressiveness lines of code, which solve these specific problems [23]. A DSL is a computer language of limited expressiveness, through appropriate notations and

abstractions, focused and generally restricted to a specific problem domain [24, 25]. DSLs have the potential to reduce the complexity of software development by increasing the level of abstraction for a domain. According to the application domain, different notations (textual, graphical, tabular) are used [26].

In this work we used Xtext [27] (a language workbench) to implement a DSL that helps developers to specify and execute test cases related to the process of QDM generation. The main rationale to building this DSL was the significant costs related to specifying *high-level rules* (particularly during acceptance tests). In the next section we present more details about our architectural decisions and implementation of our metaprogramming and DSL approach, which deals with the generation and test of QDMs.

### Software Testing

Controle e garantia de qualidade são atividades essenciais para qualquer empresa que produza produtos para serem usados por outros. Nos primórdios da computação a qualidade era responsabilidade exclusiva do programador. Padrões para garantia da qualidade de software foram introduzidos no desenvolvimento de *software* de contrato militar durante a década de 1970 e se espalharam rapidamente no desenvolvimento de *software* no mundo comercial [28].

O Software Engineering Body of Knowledge (SWE-BOK) [29] constitui o primeiro consenso internacional desenvolvido sobre o conhecimento fundamental exigido por todos os engenheiros de *software*, tornando-se um guia de uso e aplicação das melhores práticas de Engenharia de *Software*, de maneira sensata e razoável. O SWEBOK é organizado em 10 áreas de conhecimento (Knowledge Areas - KA) e um tópico referente às disciplinas relacionadas com a Engenharia de *Software*. A KA de Qualidade de *Software* cobre técnicas estáticas, aquelas que não requerem a execução do *software* sendo avaliado, enquanto técnicas dinâmicas são cobertas na KA Teste de *Software*.

A KA Teste de *Software* consiste na verificação dinâmica do comportamento de um programa com um conjunto finito de casos de testes, selecionados de um domínio geralmente infinito de execuções, para confirmar o comportamento especificado esperado. Os processos de verificação e validação determinam se os produtos de desenvolvimento de uma determinada atividade estão em conformidade com os requisitos dessa atividade e se o produto satisfaz seu uso pretendido e as necessidades do usuário. A verificação é uma tentativa de garantir que o produto seja construído corretamente, no sentido de que os produtos de saída de uma atividade atendem às especificações impostas a eles em atividades anteriores. A validação é uma tentativa de

garantir que o produto certo seja construído, ou seja, o produto atenda ao seu propósito específico [29]. Teste de *software* é uma das atividades de verificação e validação. Existem outros métodos para garantir a qualidade, como verificação baseada em modelo ou prova de teorema, mas a maioria dos desenvolvedores prefere o teste porque é mais simples em conceito e mais fácil de implantar [30].

A motivação do teste de *software* é afirmar a qualidade do *software* com métodos que possam ser aplicados de maneira econômica e eficaz tanto em larga escala como em pequena escala [28]. O teste de *software* é uma área muito ampla, que envolve muitas outras áreas técnicas e não técnicas, como especificação, design e implementação, manutenção, processos e gerenciamento de problemas em engenharia de *software* [31].

O teste destina-se a mostrar que um programa faz o que se pretende fazer e descobrir defeitos do programa antes de ser colocado em uso [32]. Porém, é importante citar uma frase de Edsger W. Dijkstra [33]: "testes de programas podem ser usados para mostrar a presença de *bugs*, mas nunca para mostrar sua ausência!". Embora essa frase indique um limite da utilidade dos testes, também indica uma das propriedades mais úteis dos testes: encontrar *bugs* [34].

O estudo realizado por Elberzhager et al [35], identifica abordagens existentes para reduzir o esforço de teste sem que o *software* final perca qualidade. A maioria dos artigos analisados, cerca de 50%, discute a automação de testes. Este resultado não é surpreendente, porque ao aplicar abordagens de automação de teste, muito tempo de teste pode ser salvo, reduzindo o custo do teste de *software* [36]. A automação permite mais ciclos de teste devido a testes repetíveis e execuções de teste mais frequentes. Também facilita a verificação rápida e eficiente de mudanças de requisitos e correções de erros e minimiza os erros humanos [37], [36].

Na literatura são descritas diversas técnicas para geração de casos de testes e dados para testes, como [38], [36], [39], [40], [41], [42], [43]: técnicas de geração aleatória de dados de teste, que escolhem entradas arbitrariamente até que entradas úteis sejam encontradas. Técnicas de geração de dados de teste simbólico, que criam expressões algébricas para várias restrições no programa, atribuindo valores simbólicos à variável. A execução simbólica para geração automática de testes possui dois tipos de grupos: geração de teste estático, feita sem executar o programa; e técnica de geração dinâmica, que determina quais casos de teste atendem ao requisito durante a execução do programa. Existe também a técnica de geração automática de testes baseada em busca, onde o domínio de entrada do sistema sendo testado é usado como espaço de pesquisa e são selecionados os dados que atendem ao objetivo da geração. O processo para encontrar

a melhor solução possível entre todas as disponíveis é conhecido como otimização. Esta técnica de geração de testes pode utilizar diversos algoritmos, como: subida de encosta, algoritmos genéticos (GA), Ant Colony Optimization (ACO), Bee Colony Algorithm (BCA), Harmony Search (HS), Particle Swarm Optimization (PSO), programação dinâmica, entre outros algoritmos utilizados para solucionar problemas de busca, que geralmente utilizam heurísticas.

Para a geração de testes desenvolvida neste trabalho foi usada a técnica Property-based Testing (PBT), que tornou-se uma técnica padrão para melhorar a qualidade do *software* em uma ampla variedade de linguagens de programação [44]. PBT é uma técnica de teste de caixa-preta, semi-automática, de alto nível na qual, ao invés de escrever muitos testes de unidade manualmente, o testador simplesmente especifica propriedades gerais que o sistema sob teste (System under test - SUT) deve satisfazer, e geradores que produzem entradas aleatórias bem distribuídas para as partes do sistema que são testadas. Como em todas as técnicas de testes aleatórios, a chance de encontrar um erro e a confiança no SUT aumentam com o número de testes gerados [45].

PBT é uma técnica que está recebendo muita atenção recentemente, é fortemente relacionada a testes baseados em modelos, mas onde o foco está na geração de testes individuais automaticamente a partir de especificações na forma de testes de propriedades de alto nível [46]. PBT é uma estratégia que leva a uma família de poderosas ferramentas de teste automáticas. Geralmente, são ferramentas para geração e execução de casos de teste com base nas especificações. Geralmente trabalhando como uma biblioteca, elas permitem que o desenvolvedor/testador anote as especificações do programa, usando uma linguagem de programação de sua escolha, na forma de propriedades que devem ser satisfeitas. A partir dessas especificações, as ferramentas geram, executam e verificam automaticamente os resultados de um grande número de casos de teste aleatórios para ver se as propriedades são mantidas [47]. Existe uma variedade de ferramentas PBT que surgiram ao longo dos anos inspiradas no QuickCheck original para Haskell [48], como: JUnit-QuickCheck para Java, theft para C, ScalaTest and ScalaCheck para Scala [49].

Ao contrário dos métodos de teste convencionais, em que o comportamento de um programa é testado apenas em alguns casos pré-determinados, o PBT enfatiza a definição de propriedades e, em seguida, testa sua validade em relação a pontos de dados amostrados aleatoriamente. Como o teste baseado em propriedade (PBT) usa um pequeno número de pontos de dados amostrados aleatoriamente, ele ainda fornece apenas

uma resposta aproximada à questão de saber se uma propriedade é satisfeita em todos os pontos de dados. No entanto, pode fornecer mais confiança do que o teste unitário convencional, porque os pontos de dados amostrados aleatoriamente podem cobrir casos problemáticos que não foram previstos pelos desenvolvedores [49].

PBT apresenta três vantagens sobre o teste manual [30]. Primeiro, escrevendo apenas propriedades e omitindo dados de teste, os desenvolvedores podem reduzir significativamente o tamanho e complexidade do código de teste. Segundo, ao executar testes baseados em propriedade contra entradas geradas aleatoriamente, os desenvolvedores podem encontrar mais *bugs* do que testes manuais, pois os testes baseados em propriedade podem verificar propriedades com muitas entradas, então a chance de encontrar *bugs* é aumentada. Finalmente, as propriedades podem ser vistas como uma especificação formal leve do programa. Escrever especificações formais promove o entendimento da especificação e da implementação e é eficaz para compartilhar as especificações entre a equipe. É importante que essa especificação seja executável. Essa especificação executável pode evitar incompatibilidade entre especificação e implementação, o que geralmente é o caso se a especificação for gravada em um documento separado.

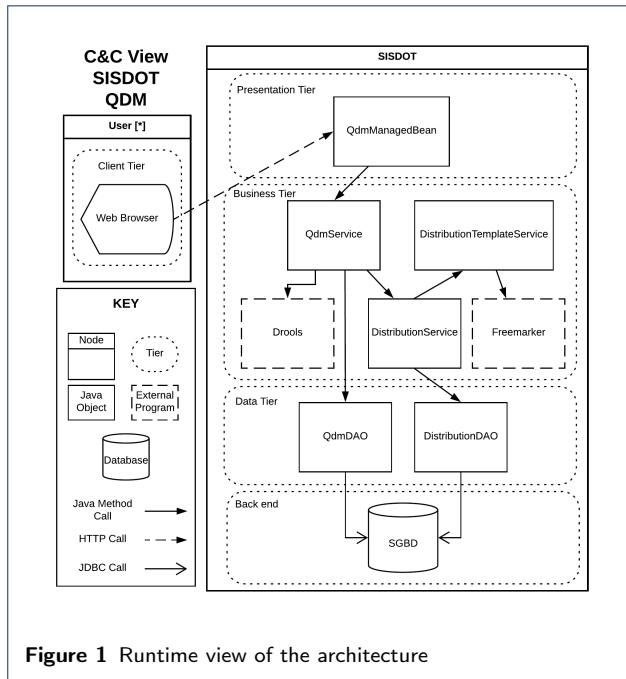
## Architectural Decisions

SISDOT is based on the client-server architectural style (according to a typical Java enterprise application), where the components interact to each other by requesting services from other components [50]. SISDOT is also organized into logical groups of components (known as tiers)—a strategy often used in the client-server architectural style. The multi-tier architectural pattern can be classified as a *component and connector* pattern, depending on the criterion used to define the tiers [51]. Indeed, there are several criteria, such as component type, sharing of the execution environment, and so on [50].

The multi-tier pattern also induces some topological constraints that govern how the components might interact with other components. Specifically, connectors might only exist between components of the same tier or that reside in an adjacent tier [50]. In addition, according to Bass et al., the organization of a system using tiers supports software maintenance, facilitates the definition of security policies, and might increase the performance, availability, and modifiability of the systems [51].

**Runtime Architectural Description for QDM Generation**  
In order to further detail and illustrate the SISDOT architecture, and the main architectural decisions to

address the constraints introduced in Section , here we present a scenario that refers to the generation of QDMs using a runtime architectural description, whose components, relationships, and interactions are illustrated in Figure 1).



**Figure 1** Runtime view of the architecture

arrumar exemplo mais completo de DRL  
falar sobre as regioes da DRL? (qc, fracao, cargo)

Considering the user interaction, a previously authenticated specialist on the distribution of equipment throughout the Brazilian Army, using her web browser, requests the web page that allows the generation of QDMs (one of the main products of SISDOT). After selecting the *generic organizational unity* for which a QDM would be generated, the system sends a request to the server. The client tier is responsible for these actions. The request is then received by a component of the presentation tier, more specifically by a *managed bean* that works as a *front controller* (in this case, a JSF controller).

This controller validates the request data, before sending a new request to the business tier. In the business tier, the service that implements the business rules related to the QDM domain object is invoked to start the QDM generation process, which includes actions to (a) transform *high-level rules* into the *low-level rules* specified using the Drools Rule Language (DRL), (b) execute these rules to instantiate a business object that represents a QDM, and (c) save this object in the persistence layer (a relational database).

This process for generating a QDM starts by retrieving a list of previously registered *high-level rules*.

An HLR, regardless of its type (either based on the full structure of the organization or based on its components), is related to one or more *military materials / equipment* (MEMs) and defines the respective amount that should be assigned to a military unit (which ranges from an organization, a center, a department, a brigade, or even a military function or qualification of a soldier).

A high-level rule has one or more distribution rules. Each rule has a type, a value, and an associated description. For example, a rule for a department type has the value of the department identifier and the description of the department name. For a better use of the database, avoiding the definition of numerous columns that would inevitably be null for several rule types, we decided to persist the distribution rules of an HLR using JSON (JavaScript Object Notation), which is converted back to an object when it is retrieved. The set of these rules defines exactly who should receive the MEMs specified by the HLR.

One of the factors that motivated the use of a rule based engine was the similarity between *high-level rules* and the “if-then” rules of these mechanisms. An HLR has a set of conditions that fits the conditional part of a rule (“if”) and has a set of actions (“then”) that, in this case, trigger the distribution of MEMs throughout the expected military unities.

The list of retrieved *high-level rules* contains only Java objects. In order to be able to use the rule engine in a transparent way for the end users (and also for developers in future maintenance scenarios), we decided to use a meta-programming approach, translating these objects into a set of rules (a program in logic programming) that might be used by a rule-based engine. As mentioned before, we decided to use the *Drools rule-based engine* in the context of SISDOT (Figure 1), mostly because of its integration capabilities with both JEE systems and the Wildfly application server. Accordingly, the *high-level rules* are translated into DRL rules (Listing 1 presents an example of a *low-level Drools rule*). To this end, we use a template engine (FreeMaker) to implement this particular transformation (Figure 1), using a template that contains the required *markups* for transforming HLR into *low-level* Drools rules. That is, for each HLR, the template engine populates the template, resulting in a DRL rule consistent with the converted HLR.

The RHS of the Drools rule of Listing 1 iterates over all selected military roles (that satisfy the constraints of the LHS). We use a helper class to assign the equipment id (in this case a *rifle* whose id is 1051000007) to the individual roles. Besides providing means to populate a QDM, this helper class also checks some constraints related to the QDM domain model.

```

1 rule "HLR 1"
2   when
3     $qc: QuadroDeCargosV0(
4       (operational == OPERATIONAL || operational == NOT_OPERATIONAL)
5       , natureId not in (50, 51)
6       , valueId in (23, 41, 90, 50, 42, 31, 21, 32, 20, 30, 40)
7     )
8     $roles: List( size > 0 ) from accumulate (
9       $dep: DepartmentV0(
10        depId not in (14086, 1130)
11      ) from $qc.deps
12      and
13      $role: RoleV0(
14        function == "ROLE"
15        , roleId not in (364, 539, 121, 2205, 528, 2204, 539)
16        , rankId in (24, 23, 44, 22, 42)
17        , armsId not in (1055, 10, 833, 5308, 5393, 729, 5110, 15, 5112, 5390, 12)
18      ) from $dep.roles
19      and
20      QualificationV0(
21        code not in ("782", "756", "793", "768", "765", "751", "921", "749", "767", "748")
22      ) from $role.qualifications
23      and
24      ObservationV0(
25        code in ("40D", "40K")
26      ) from $role.observations;
27
28     collectList( $role )
29   )
30   then
31     for(int i=0; i < $roles.size(); i++){
32       RoleV0 c = (RoleV0) $roles.get(i);
33       helper.add(1051000007, c.getFractionId(), c.getRoleId());
34     }
35 end

```

**Listing 1** Example of a low-level Drools rule

A QDM is generated for a generic organizational unity (hereafter QC, from *Quadro de Cargos* in Portuguese), selected by the user on the client tier. Therefore, the user-specified QC must be retrieved from the Brazilian Army's enterprise database through a specific Data Access Object (DAO) in the data tier [52]. From the list of DRL rules, created based on existing *high-level rules*, the rule engine is instantiated and these rules are compiled, so that they can be triggered by Drools. The QC domain object is then inserted into the working memory as a fact, an information that is always considered true. The rules are only executed when their conditions are satisfied, based on the specified QC organizational structure. When a rule is valid, that is, it has been activated and the *Left Hand Side* (LHS) conditions are valid, the QDM is populated with the MEMs specified in the *Right Hand Side* (RHS) of

the rules. Therefore, we generate a complete QDM object after verifying all valid rules (previously registered in the database) for the selected QC.

After generating a QDM domain object, it is saved on the database and we let the domain expert know of the success of the operation using a simple user interface message. The domain expert can then perform the necessary operations on the generated QDM, including a workflow involving its edition, homologation, and validation.

#### A Domain Specific Language for testing QDM Generation

As explained before, to generate a QDM, a set of *high-level rules* must be previously declared. This is a time-consuming task, particularly when using the interface



of the system. In order to facilitate the definition of the *high-level rules* used in the automated test scripts, we decided to implement HLRDSL—a DSL (Domain Specific Language) that enables us to specify *high-level rules* in a clear, objective, and declarative way (and most important, without using the interface of the SISDOT system).

We implemented HLRDSL using Xtext, which generates plugins that allow code editing in both Eclipse and IntelliJ, plus an editor that can be embedded in a web application. This way, a developer writing test scripts might use HLRDSL to take advantages of the functionality provided by these plugins. To implement a DSL using Xtext, it is first necessary to declare a grammar expressed using a syntax similar to ANTLR [53]. Listing 2 presents the grammar of our DSL. From this grammar, the Xtext tool suite generates a lexer, a parser, a set of classes representing the AST (Abstract Syntax Tree), and an editor with several features typically available in IDEs.

```
1 grammar br.unb.cic.sisdor.chamador.HLRDSL
2 with org.eclipse.xtext.common.Terminals
3 generate hlrdsl "http://www.unb.br/sisdor/hlrdsl"

5 Model:
6 rules += Rule*;

8 Rule:
9 'rule' code = INT '{'
10 ('type' '=' type=OmType ',')?
11 'materials' '=' '[' items+=Item (',' items+=Item)*
12 'cnds' '=' '[' cnds+=Condition (',' cnds+=Condition)
13 '}' ;

15 Item: '(' codot=STRING (':' name=STRING)? ':'
16 amount=INT ')';

17 Condition:
18 '(' ('cond' '=' cond=ConditionType ',')?
19 'type' '=' type = RuleType ', '
20 'values' '=' '[' values+=STRING (',' values+=STRING)
21 ']' ;

23 enum OmType: BOTH | OPERATIONAL | NOT_OPERATIONAL;
24 enum ConditionType: AFF | NEG;
25 enum RuleType: NATURE | SUB_NATURE | VALUE |
26 DEPARTMENT | ARMS | ROLE_QC | SUB_CIRCLE |
27 RANK | ROLE | QUALIFICATION | OBSERVATION ;
```

Listing 2 Xtext grammar defining the DSL structure

Listing 3 presents a simple example of a HLR declaration using our DSL. The goal of this rule is to

distribute the materials (defined in the list of materials construct) for *operational* OMs (see the type construct) and *militaries not working in the specified list of departments* with a set of qualifications and roles. While the definition of this rule using our DSL requires 12 lines of code, the corresponding definition of the same rule using a Java test case requires more than 50 lines of imperative code.

```
1 rule 1 {
2     type = OPERATIONAL,
3     materials = [
4         ("1051000007" : "Rifle" : 1),
5         ("1020100011" : "Rifle Carrying Case" : 2)
6     ],
7     cnds = [
8         (cond=NEG, type=DEPARTMENT, values=[14086 ,
9             ↪ 1130]),
10        (type=QUALIFICATION, values=[782, 756]),
11        (type=ROLE, values=[2, 12, 15, 21])
12    ]
13 }
```

Listing 3 Example of a HLR declaration using our DSL

Our first test approach consists first in the specification of the required *high-level rules* in a HLRDSL file. Next, the *behavior under test* is detailed as features using the Cucumber [54] framework (we present an example of a feature in Listing 4). After that, the developer details the implementation steps of the features by (a) specifying which *high-level rules* (previously declared using our DSL) should be considered in the test execution, (b) specifying which QC the QDM should be generated, and (c) specifying the expected results in terms of the properties of the expected QDM. To this typical scenario, we use a testing infrastructure that supports several facilities, that aims to increase productivity, such as: database connectivity, auxiliary methods for executing queries in the database, and predeclared rules in the rule engine.

That is, HLRDSL simplifies the process of specifying *high-level rules* and its primary goal was to assist in test activities. However, after presenting to domain experts some examples of *high-level rules* specified using our DSL, we realized that HLRDSL might also be useful to simulate the specification of rules, allowing a better understanding of the effect of each HLR for building QDMs.

```
1 Feature: QDM 757311
2 As a user
3 I want to generate a QDM from HLRs

5 Scenario: Test 1
```

```

6 Given QC with code 757311
7 And HRL: 1, 2
8 When I generate a QDM
9 And the results are consistent
10 Then positions must be populated
11 | idDep | idPos | codot | amount |
12 | 6765748 | 10403 | 1051000025 | 1 |
13 | 6765748 | 10403 | 1020100017 | 1 |
14 | 6765748 | 10403 | 1020100013 | 1 |
15 | 6765748 | 10403 | 1051000006 | 1 |
16 | 6765748 | 21328 | 1051000025 | 1 |
17 | 6765750 | 10122 | 1051000025 | 1 |
18 | 6765793 | 13276 | 1051000025 | 6 |
19 | 6765796 | 24655 | 1051000025 | 1 |
20 | 6765796 | 24429 | 1051000025 | 2 |
21 | 6765798 | 24429 | 1051000025 | 3 |
22 | 6765798 | 11433 | 1051000025 | 9 |
23 And the final result must be:
24 | codot | amount |
25 | 1020100013 | 74 |
26 | 1020100017 | 74 |
27 | 1051000006 | 74 |
28 | 1051000025 | 74 |

```

**Listing 4** *Cucumber feature*

Our second test approach automatically generates test cases from our DSL and from a random sample of simple combinations that characterize artificial organizational unities. This *property based approach* involves a custom Maven plugin that currently generates simple test cases, freeing the tester to create only complex tests related to the QDM generation.

These test cases are generated from a combination of conditions, for example: one test case only for operational OMs and another for non-operational OMs; a test case for the different kinds of OMs (artillery, infantry, cavalry, aviation and helicopters, general services, and so on), combining with the operational type, such as: operational infantry OMs and non-operational infantry OMs. These combinations are generated for simple cases so that they are easy to verify without the need to manually implement Java test scripts for generating QDMs.

O *plugin* define diversos cenários que serão gerados para cada QC declarado. Cada cenário é responsável pela geração de um chamador no formato DSL e um arquivo de definição do cenário no formato do cucumber. Para cada cenário é gerado um QDM para o QC, usando o chamador gerado e o resultado deste QDM é conferido de acordo com os dados previstos na definição do cenário do cucumber. Foram criadas diversas combinações de cenários e para cada uma delas alguns dos valores das propriedades são gerados aleatoriamente, seguindo a técnica PBT.

Os cenários de teste foram divididos em três grupos, de acordo com as seções da DRL (QC, fração e cargo). Isso foi feito pois cada região é testada de forma quase que independente, possuindo dependência apenas da validade da região anterior. Por exemplo: a região de fração só é avaliada se a região de QC tiver sido avaliada como verdadeira, então essas regiões podem ser testadas de forma independente ao assumir que a região anterior é verdadeira e que foram gerados casos de testes que ajudaram a validar o funcionamento das condições da região anterior. Ao assumir a independência dos grupos, para fins de teste, é possível focar nas combinações das propriedades de cada grupo, facilitando a geração de cenários de teste.

Cada grupo de cenários é responsável por testar várias das combinações das propriedades da entidade relativa ao grupo (QC, fração ou cargo). Em cada cenário é definido um chamador e uma lista com os resultados esperados, que são gerados de acordo com a combinação específica de propriedades deste cenário. Como o MEM definido no chamador não interfere na execução das regras, foi definido um MEM padrão para todos os chamadores gerados, assim como a quantidade igual a um (unitário). Isso foi feito para diminuir a complexidade do algoritmo de geração ao não precisar acessar o banco de dados várias vezes para recuperar MEMs de forma aleatória, tendo em vista que não interferem na avaliação das regras envolvidas na geração do QDM.

No grupo de QC existem as seguintes propriedades: operacionalidade, natureza, subnatureza e valor. Foram criados cenários específicos para cada propriedade e para combinações entre elas. Para a propriedade natureza do QC, por exemplo, foram definidos os seguintes cenários:

- natureza válida e tipoOM válido: este cenário gera um QDM válido, com resultados válidos, pois neste cenário é criado um chamador com os mesmos valores de natureza e operacionalidade (tipoOM) do QC para o qual será gerado o QDM, ou seja, natureza e tipoOM válidos;
- natureza válida e tipoOM válido (AMBOS): este cenário funciona de forma parecida com o anterior, mas o tipoOM é definido como "AMBOS" no chamador. Este tipoOM indica que o QC pode ser operacional ou não-operacional;
- natureza válida e tipoOM inválido: o valor da natureza é o mesmo do QC, pois é válido, e o valor do tipoOM é alterado. Por exemplo: se o QC for operacional o valor é alterado para não-operacional e vice-versa;
- natureza inválida e tipoOM válido: o valor da natureza deve ser inválido no chamador, ou seja, um valor diferente da natureza do QC. A geração deste valor inválido de natureza é realizado

de forma randômica, usando a seed configurada no plugin maven, em uma lista de códigos de naturezas reais previamente definida;

- natureza inválida e tipoOM inválido: os valores de natureza e tipoOM são inválidos, e são gerados conforme citado anteriormente: invertendo o valor de operacionalidade e recuperando randomicamente um código de natureza da lista declarada previamente;
- negar natureza, inválido: as combinações citadas anteriormente nesta lista possuem chamadores com regras do tipo "AFIRMACAO", o tipo padrão de uma regra. Mas existe o tipo "NEGACAO", onde a regra é negada, ou seja, a condição (na DRL) é válida quando o valor testado não pertence à lista de valores especificados na regra. Neste cenário, por exemplo, há uma regra de negação da natureza de forma que o QDM seja inválido. Então foi criada uma regra para "NATUREZA" do tipo "NEGACAO" definindo como valor o código da natureza do QC. Com isso, esta regra não é válida para este QC, sendo válida para todas as outras naturezas exceto a deste QC;
- negar natureza, válido: para este cenário é criado um chamador com uma regra para "NATUREZA" do tipo "NEGACAO", mas gerando um código de natureza diferente da natureza do QC. Com isso, esta regra é válida para este QC.

Seguindo a mesma linha de combinações utilizadas para cenários envolvendo natureza do QC, foram criados cenários para subnatureza e valor. Não foram criados cenários para tipoOM pois esta é uma propriedade de preenchimento obrigatório no chamador e a checagem de sua validade já é exercitada em todos os cenários do grupo. Na Table 1 são apresentados os cenários definidos para o Valor do QC e suas combinações. Na primeira coluna é apresentado o número do cenário, criado apenas para facilitar uma possível referência a uma linha específica da tabela. As outras colunas contém os valores das propriedades do QC sendo combinadas. As células com valor 0 indicam que o valor desta propriedade, indicada pela coluna, é inválido, ou seja, deverá ser selecionado um valor randomicamente de uma lista previamente definida e este valor deve ser diferente do valor que essa propriedade apresenta no QC. As células com o valor 1 indicam que o valor desta propriedade é válido, ou seja, é o mesmo valor definido no QC para esta propriedade. As células sem valor indicam que essa propriedade não faz parte da combinação.

No grupo Cargo existem as seguintes propriedades: código do cargo, subcírculo ao qual o cargo pertence, o posto (patente) necessário para ocupar o cargo, e a

**Table 1** Cenários definidos para Valor de QC

Cenário	Valor	Subnatureza	Natureza	TipoOM
1	0			0
2	0			1
3	1			0
4	1			1
5	0		0	0
6	0		0	1
7	0		1	0
8	0		1	1
9	1		0	0
10	1		0	1
11	1		1	0
12	1		1	1
13	0	0	0	0
14	0	0	0	1
15	0	0	1	0
16	0	0	1	1
17	0	1	0	0
18	0	1	0	1
19	0	1	1	0
20	0	1	1	1
21	1	0	0	0
22	1	0	0	1
23	1	0	1	0
24	1	0	1	1
25	1	1	0	0
26	1	1	0	1
27	1	1	1	0
28	1	1	1	0

arma (Arma, Quadro ou Serviço) necessária para ocupar o cargo. O subcírculo está relacionado às graduações (posto/patente) existentes, servindo como uma forma de agrupamento. Por exemplo: o subcírculo superior agrupa as patentes de coronel, tenente-coronel e major; o subcírculo subalterno agrupa os tenentes.

Para cada cenário do grupo de cargos foi selecionado um cargo aleatoriamente da lista de cargos do QC, utilizando a seed configurada no plugin maven. De acordo com o cargo selecionado são geradas as regras conforme a combinação de propriedades especificada para o cenário. Para cada cenário, além das regras, é gerada também uma lista contendo os resultados esperados, ou seja, os dados que se espera encontrar no QDM gerado para o QC.

Para este trabalho foram especificados 100 cenários. Todos os cenários gerados, independente do grupo, contém um chamador, com as regras referentes à combinação de propriedades definida para o cenário, e uma lista de resultados esperados, que serão comparados com os dados do QDM gerado. Para cada QC declarado no plugin é gerada uma lista de cenários. Para cada QC é gerado um arquivo cucumber contendo os cenários definido para este QC. A geração do arquivo é realizada com o uso do Freemarker, usando um template previamente definido.

O processo de geração automática de testes pode ser resumido da seguinte maneira: o plugin maven deve ser

declarado e configurado, indicando a lista com os códigos dos QCs para os quais devem ser gerados testes. Então, para cada QC da lista são gerados cenários, com variações em suas propriedades de acordo com o QC sendo testado. Os cenários gerados são transformados em arquivos do cucumber e de DSL, que representam os cenários gerados. Houve então uma conversão dos cenários gerados com base na técnica Property-based Testing (PBT) em casos de teste por exemplo, que serão executados automaticamente durante a fase adequada do ciclo de vida de *build* do maven. Existe a possibilidade de geração de relatório da execução dos testes em alguns formatos, como: html e json.

In this second approach, for each of the possible combinations, we generate a HLR that meets the conditions and a cucumber feature that exercises the HLR and compares the expected results declared in the feature. Each HLR is considered during the generation of the QDM to the QCs specified in the maven plugin. That is, to use this architecture characteristic (here we consider testability as an architecture concern), we “feed” the maven plugin with a list of QC codes for which the QDMs should be generated, besides other optional information, which have default values, such as: database connection string, seed to be used in the random selection of values within a combination (e.g., considering 20 different kinds of organizational unities, randomly take 5 for generating the set of combinations), and the output directory where the code should be exported.

After running the plugin, with the generated code, the tests are executed in a similar way to the test definitions detailed before. The tests run along with the other unit tests declared, whether DSL-based or not.

## Empirical Study

For the validation of the proposed architecture we conducted an empirical study, which consists of the evaluation of the main architectural decisions of SISDOT, which relate to the generation of QDMs. The case study aims to answer the following research questions:

- **RQ.1:** Does our proposed approach, based on meta-programming, generate correct QDMs?
- **RQ.2:** Does the use of a DSL reduce the effort necessary to implement the test cases that target QDM generation?
- **RQ.3:** Does our proposed approach, based on meta-programming, generate QDMs within a satisfactory time-frame?

**RQ.1** deals with the most important concern: correctness regarding the QDM generation. It is intended to demonstrate that the solution not only automatically generates the rules, eliminating the work that the developer would have to implement the solution

in Java, but also that the proposed approach leads to the expected QDMs when considering a well defined set of *high-level rules*. **RQ.2** is related to the productivity of the tester and demonstrates how much code the developer have to write when using our DSL, compared to the effort to specify *high-level rules* using the Java language in unit test cases. During the activities of requirements elicitation, we identified a soft-goal stating the acceptable time for generating a QDM, which should not be above 10 seconds, when using 1000 *high-level rules*. Thus, **RQ.3** serves to ensure that the QDMs are generated in a satisfactory time.

## Data Collection and Analysis Procedures

We answer our first question qualitatively. Although this might look like disappointing (under the perspective of empirical methods), we have strong evidences collected from the domain experts that our approach has produced correct QDMs. We believe that this correctness is due to several factors, including the involvement of the domain experts that helped us to full understand the requirements and the expertise of our development team. Nevertheless, the decision of not writing the low-level rules at the source code level, but instead using a program generation approach that has reduced significantly the effort needed to *hand write* those rules, has also contributed to achieve this confidence about correctness.

In a meeting with the domain experts, one of the specialists said: “I got surprised with the correctness of the solution. In the previous version of the system, we had to have several interactions until we got confident about the correct operation of the *high-level rules*”. The first time that we generate QDMs, the process was validated. It is important to note that “life is not a bed of roses”, and we broke the initial implementation after the introduction of a new, at first not so related feature that, in the end, cut across several classes of the system. We fixed the related issues recently.

To answer the other two research questions, we collected the following metrics: lines of code (LOC) and execution time. The choice of these two metrics is related to the size of the code, which may indicate higher productivity; and to the fulfillment of the functional requirement that determines the maximum execution time for the generation of QDMs.

**arrumar: nao sao 66**

We use as a benchmark 66 *high-level rules* declared in its database. These *high-level rules* correspond to real rules used by the Brazilian Army to perform some tests related to the generation of QDMs. We converted each HLR written using JUnit to their correspondent one specified using our DSL. Table 2 presents a sample of the collected data. The first column contains the HLR

code, the second column contains the number of lines of code to represent the HLR using our DSL, and the third column contains the number of lines of code to represent the same rule in the Java language.

**Table 2** Comparing the number of lines of code need to specify test cases using our DSL and Java

HLR Code	DSL	Java	Difference	%
1	10	26	16	62
2	10	25	15	60
3	10	32	22	69
4	11	52	41	79
5	15	41	26	64
6	9	16	7	44
7	18	21	3	15
8	12	25	13	52
9	9	11	2	19
10	11	23	12	53
122	10	48	38	80
123	8	44	36	82
124	10	54	44	82
125	8	39	31	80
126	12	47	35	75
127	9	43	34	80
128	9	39	30	77
129	9	27	18	67
130	12	25	13	52
131	12	45	33	74

Regarding the runtime metrics, we collected the data using an Intel(R) Core™ i7-4790 processor, with 16GB of RAM, running Ubuntu 16.04.4 LTS 64-bit operating system, with 4.15.0-24-generic kernel. We used 16 QCs, chosen to represent the various OM types, natures, and sub-natures. For each QC, we generated QDMs using random sets of *high-level rules* with size: 20, 40, 50, 60, 100, 200, 400, 700, 1000, 1500, 2000. Since there were only 66 real *high-level rules* already declared, larger quantities contain repeated *high-level rules* (which in the end will duplicate the data within a given QDM). The selection of *high-level rules* was performed at random, but with the same seed, so that the same sets were used in the generation of QDMs for different QCs. Table 3 shows a sample of the results obtained with the execution. The first column contains the QC code, and the second column contains the time in milliseconds for the generation of the QDMs using the indicated amount of *high-level rules* on the respective column header.

pegar tempos da dissertacao, incluir AMAN

We exported the collected data to the CSV format, so that we could perform a data analysis using the R environment [55]. In the data analysis for **RQ.2**, which compares the number of lines of code between the two distinct approaches for representing *high-level rules*, we created two additional columns: *difference*, which measures the difference in the number of lines of code necessary to specify *high-level rules* using Java and our DSL; and *percent of reduction*, which indicates

**Table 3** Time (ms) for generation of QDMs

QC	60	100	400	700	1000	2000
206410	184	275	944	1829	2279	4707
231303	197	282	975	1731	2541	4987
500311	201	291	948	1804	2289	4950
618322	175	264	922	1640	2412	4990
727310	173	302	937	1652	2466	4992
1450311	182	269	890	1649	2527	4785
2203310	197	263	909	1619	2344	4976
5406194	176	253	884	1598	2393	4610
7020003	212	241	922	1610	2361	4470
9131000	171	252	888	1609	2345	4471

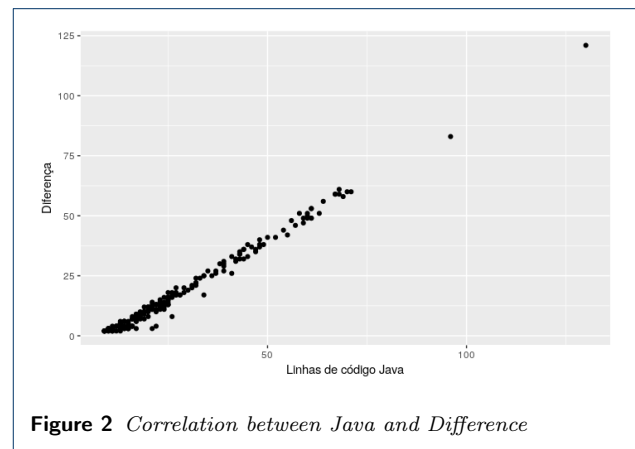
how much less code is necessary to specify a HLR using our DSL, when compared to the direct specification in Java code.

**Table 4** Analysis of DSL and Java comparison data

	Min.	Median	Mean	SD	Max.
DSL	7.00	9.00	9.67	2.01	18.00
Java	9.00	23.00	28.66	18.47	130.00
Difference	2.00	13.00	18.98	18.25	121.00
%	14.29	56.26	54.66	21.68	93.08

Table 4 presents some descriptive statistics of the collected data, containing the following columns: minimum value, median, mean, standard deviation, and maximum value. There is a strong correlation (0.982) between the lines of code in Java and the *difference* measurements. This correlation can be seen in Figure 2. It is possible to identify that as the size of the Java code grows, the difference for the representation in DSL also increases.

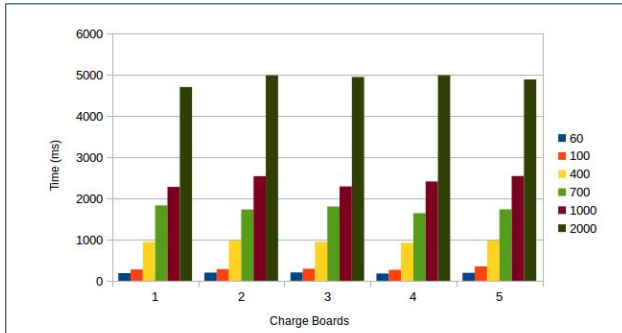
regerar figuras em ingles



**Figure 2** Correlation between Java and Difference

Figure 3 shows the time necessary to generate QDMs for 5 different QCs. For each of them, 6 QMS were generated, with different amounts of *high-level rules*, 60, 100, 400, 700, 1000 and 2000.





**Figure 3** QDM Generation Time (ms)

### Discussion, Lessons Learned, and Threats to Validity

When analyzing the data related to RQ.2, it was possible to realize that the representation of a HLR using our DSL is on average 45% smaller than the representation of the same rule in Java. This indicates a possible productivity gain for the developer when writing test cases, since she will have to write a smaller amount of code. In addition, using our DSL, we specify *high-level rules* using a declarative approach, which is close to the vocabulary of the problem domain. With respect to the third research question, the limit imposed by the functional requirement of 5 seconds was not exceeded, even when considering 2000 *high-level rules*, which is twice the number of *high-level rules* expected for the system in production. It is important to notice that these number (5 seconds as time limit and 1000 *high-level rules* were collected together with the stakeholders of SISDOT).

The general approach discussed in this paper involves the use of metaprogramming to generate low-level Drools rules from *high-level rules* and the use of a DSL to specify *high-level rules* during test activities. We have some previous experience using Drools, and, at the beginning of this research collaboration, we realized that the use of a rule based engine (such as Drools), could simplify the computation of QDMs (one of the main products of SISDOT). Nevertheless, we should try to abstract the use of Drools, mostly due to architectural conformance—we should not introduce new languages during the development of SISDOT. For this reason, we applied a transformational approach, translating Java domain objects into Drools rules at runtime. Perhaps due to the structure of the *high-level rules*, the task of implementing this transformational approach was not difficult. We believe that it was the right decision considering the original set of requirements. Surely, a number of uncertainties arise, including: “will our proposed approach support the expectations of the end-users w.r.t. correctness and time

constraints?”. We address this question in our empirical evaluation.

However, we found that testing the process of QDM generation was time consuming, and then we designed a DSL to address the main bottleneck: the specification of *high-level rules* either using Java test scripts or the SISDOT user interface (the common way for specifying *high-level rules*). Our choice for using Xtext to this end had simplified the entire process for implementing a DSL. Even considering that HLRDSL is a simple and small DSL, we argue that the set of Xtext tools is really simple and productive, reducing the overall effort to implement the HLRDSL grammar, parser, and program generator. We could have used other tools and techniques as well, for instance a metaprogramming approach to support our test activities. However, such approach would not have solved the main problem that motivated us to implement HLRDSL. In this case, our goal was to design a simple approach for specifying *high-level rules*. In the end, HLRDSL started to be considered an interesting approach for use not only in test activities, but also for simulating QDMs without the need to persist *high-level rules* in the production environment.

Our empirical study presents several limitations. First, we discuss *correctness* based on the results of both integration and acceptance tests, though also considering the opinion of the domain experts that are currently confident that our approach for QDM generation is working properly. Surely, we could have used a more formal approach for discussing correctness, and we postpone this investigation to a future work. Nevertheless, considering the scope of this paper, the feedback from the domain experts and testing outcomes reduce some of our uncertainties about correctness.

Second, we only considered 66 real *high-level rules* in our study, which served as a basis for measuring the time to generate QDMs. SISDOT is expected to have around 1000 *high-level rules* in the production environment. To mitigate the threat related to the small number of available *high-level rules* for testing, we considered their repetition when carrying out a performance test of SISDOT. In this way, we expect that the system would present a behavior similar to the production environment (with respect to performance). The same situation occurs with the comparison of lines of code between the declarations of *high-level rules* in DSL and Java. As the number of *high-level rules* is relatively small, it may not be possible to generalize the results we found.

### Related Work

We addressed the main architectural concerns of SISDOT using well-known patterns for enterprise systems [56] and software architecture patterns [57].

Moreover, the ideas of using rule based engines to support inference within software systems are also not new [58, 59], and have been particularly explored in the health domain [60, 59, 61]. For instance, Van Hille et al. present a comparison between Drools and OWL + SWRL for implementing the reasoning procedures of the alert module of an implantable cardioverter defibrillator. The authors conclude that Drools provides greater expressiveness when compared to the other approach. Our decision to use Drools to implement low-level rules was mostly based on its integration to the JEE platform, instead of its expressiveness.

Dingcheng Li and colleagues report on the use of the Drools engine to model and reason about *electronic health records* [59]. To this end, the authors of this mentioned work implemented a *model-to-model translational approach* that converts the specification of phenotyping algorithms expressed using the Quality Data Model (from the National Quality Forum<sup>[1]</sup>) into Drools rules. Similarly, Jung et al. present an approach for executing medical logic modules expressed in ArdenML using Drools [61]. They also used a model-to-model approach implemented using XSLT. Our approach could also be characterized as being based on *model-to-model transformations*, since we translate a business entity model (represented as the *high-level rules* Java domain objects at runtime) into Drools rules, though using FreeMarker as template engine.

Several DSLs have been proposed to support test activities. For instance, Mugridge and Cunningham present the *Fit Framework for Integrated Tests*, which is based on a DSL [62]. The Cucumber framework also uses a DSL for expressing the expected behavior of software features [63]. Here, we designed a small DSL to reduce the effort for setting up a suite of acceptance tests (based on Cucumber) for QDM generation, particularly helping the test team to specify *high-level rules*.

## Final Remarks

In this paper we detailed the main design and architectural decisions we use to implement the mechanisms necessary to support the distribution of materials through the Brazilian Army unities. These decisions included the use of a metaprogramming approach to derive high level business rules in rules specific to a rule based system (Drools) and the definition of a DSL that facilitates the declaration of *high-level rules* to build test scripts.

We carried out an empirical study to validate the proposed architecture, based on the evaluation of one of the main scenarios of our logistic system (SISDOT),

related to the generation of QDMs. From the results obtained, it was possible to observe a reduction in the average number of lines of code required for the declaration of a *high-level rules* using our DSL, in comparison with the representation of the same HLR in Java. Another result of the case study shows that our approach fulfills a non functional requirement that constraints the expected time to generate QDMs. This limit was not exceeded even when using twice the number of *high-level rules* expected for the system in production environment.

Although the architecture discussed here considers the specific needs of the Brazilian Army, we believe that logistic systems from other institutions might benefit from our technical decisions as well.

### Competing interests

The authors declare that they have no competing interests.

### Author's contributions

Text for this section . . .

### Acknowledgements

We would like to thank the anonymous referees for the valuable comments and suggestions, which have led to several improvements of this paper. This work is partially supported by PROMISE-EB, a software modernization project involving the University of Brasília and the Brazilian Army.

### References

- Browne, P.: JBoss Drools Business Rules. Packt Publishing Ltd, ??? (2009)
- Rutner, S.M., Aviles, M., Cox, S.: Logistics evolution: a comparison of military and commercial logistics thought. *The International Journal of Logistics Management* **23**(1), 96–118 (2012). doi:10.1108/09574091211226948
- Prebilić, V.: Theoretical aspects of military logistics. *Defense & Security Analysis* **22**(2), 159–177 (2006). doi:10.1080/14751790600764037
- OTAN, L.C.: NATO Logistics Handbook. 2nd Edition, 2012 (2012). [https://www.nato.int/docu/logi-en/logistics\\_hndbk\\_2012-en.pdf](https://www.nato.int/docu/logi-en/logistics_hndbk_2012-en.pdf)
- Brasil, M.D.D.: Manual de Campanha C 100-10 - Logística Militar Terrestre (2ª edição, 2003). [http://bibliotecamilitar.com.br/wp-content/uploads/2016/02/02-manual\\_c\\_100-10-logistica\\_militar\\_terrestre.pdf](http://bibliotecamilitar.com.br/wp-content/uploads/2016/02/02-manual_c_100-10-logistica_militar_terrestre.pdf)
- Graham, I.: Business Rules Management and Service Oriented Architecture: a Pattern Language. John Wiley & Sons, ??? (2007)
- Salatino, M., De Maio, M., Aliverti, E.: Mastering JBoss Drools 6. Packt Publishing Ltd, ??? (2016)
- Grosan, C., Abraham, A.: Rule-Based Expert Systems, pp. 149–185. Springer, ??? (2011). doi:10.1007/978-3-642-21004-4\_7. [https://doi.org/10.1007/978-3-642-21004-4\\_7](https://doi.org/10.1007/978-3-642-21004-4_7)
- Lucas, P.J.F., van der Gaag, L.C.: Principles of Expert Systems. International computer science series. Addison-Wesley, ??? (1991)
- Gallacher, J.: Practical introduction to expert systems. *Microprocessors and Microsystems* **13**(1), 47–53 (1989). doi:10.1016/0141-9331(89)90034-3
- Hayes-Roth, F.: Rule-based systems. *Commun. ACM* **28**(9), 921–932 (1985). doi:10.1145/4284.4286
- Abraham, A.: Rule-Based Expert Systems. American Cancer Society, ??? (2005). Chap. 130. doi:10.1002/0471497398.mm422. <https://onlinelibrary.wiley.com/doi/abs/10.1002/0471497398.mm422>
- Czarnecki, K.: Generative programming - principles and techniques of software engineering based on automated configuration and fragment-based component models. PhD thesis, Technische Universität Illmenau, Germany (1999). <http://d-nb.info/958706700>

<sup>[1]</sup><http://www.qualityforum.org/Home.aspx>

14. Arora, R., Bangalore, P., Mernik, M.: Developing scientific applications using generative programming. In: 2009 ICSE Workshop on Software Engineering for Computational Science and Engineering, pp. 51–58 (2009). doi:10.1109/SECSE.2009.5069162. <https://ieeexplore.ieee.org/document/5069162/>
15. Barth, B., Butler, G., Czarnecki, K., Eisenecker, U.: Generative programming. In: Frohner, A. (ed.) Proceedings of the Workshops on Object-Oriented Technology. ECOOP '01, pp. 135–149. Springer, Berlin, Heidelberg (2002). <http://dl.acm.org/citation.cfm?id=646781.705926>
16. Benato, G.S., Affonso, F.J., Nakagawa, E.Y.: Infrastructure based on template engines for automatic generation of source code for self-adaptive software domain. In: Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE, pp. 30–35 (2017). doi:10.18293/SEKE2017-147. [http://ksiresearchorg.ipage.com/seke/seke17paper/seke17paper\\_147.pdf](http://ksiresearchorg.ipage.com/seke/seke17paper/seke17paper_147.pdf)
17. Segura, S., Durán, A., Troya, J., Cortés, A.R.: A template-based approach to describing metamorphic relations. In: 2017 IEEE/ACM 2nd International Workshop on Metamorphic Testing (MET), pp. 3–9 (2017). doi:10.1109/MET.2017.3
18. Radjenovic, J., Milosavljevic, B., Surla, D.: Modelling and implementation of catalogue cards using freemarker. Program **43**(1), 62–76 (2009). doi:10.1108/00330330910934110
19. Parr, T.J.: Web application internationalization and localization in action. In: Proceedings of the 6th International Conference on Web Engineering. ICWE '06, pp. 64–70 (2006). doi:10.1145/1145581.1145650. <http://doi.acm.org/10.1145/1145581.1145650>
20. Parr, T.: Language Implementation Patterns: Create Your Own Domain-specific and General Programming Languages. Pragmatic Bookshelf, ??? (2010)
21. Karsai, G., Krah, H., Pinkernell, C., Rumpe, B., Schindler, M., Völkel, S.: Design guidelines for domain specific languages. CoRR **abs/1409.2378**(October), 7 (2014)
22. Raja, A., Lakshmanan, D.: Domain specific languages. International Journal of Computer Applications **1**(21), 105–111 (2010). doi:10.5120/37-640. arXiv:1011.1669v3
23. Neeraj, K.R., Janardhanan, P.S., Francis, A.B., Murali, R.: A domain specific language for business transaction processing. In: 2017 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), pp. 1–7 (2017). doi:10.1109/SPICES.2017.8091270. <https://ieeexplore.ieee.org/document/8091270/>
24. Fowler, M., Parsons, R.: DSL: Linguagens Específicas de Domínio. Bookman, ??? (2013)
25. van Deursen, A., Klint, P., Visser, J.: Domain-specific languages: An annotated bibliography. SIGPLAN Not. **35**(6), 26–36 (2000). doi:10.1145/352029.352035
26. Pfeiffer, M., Pichler, J.: A comparison of tool support for textual domain-specific languages. In: Proceedings of the 8th OOPSLA Workshop on Domain-Specific Modeling. DSM'08, pp. 1–7 (2008). <http://www.dsmforum.org/events/dsm08/papers.html>
27. Eysholdt, M., Behrens, H.: Xtext: Implement your language faster than the quick and dirty way. In: Proceedings of the ACM International Conference Companion on Object Oriented Programming Systems Languages and Applications Companion. OOPSLA '10, pp. 307–309. ACM, New York, NY, USA (2010). doi:10.1145/1869542.1869625. <http://doi.acm.org/10.1145/1869542.1869625>
28. Pressman, R.S.: Software Engineering: a Practitioner's Approach. McGraw-Hill, ??? (2005)
29. Bourque, P., Fairley, R.: SWEBOK - Guide to the Software Engineering Body of Knowledge, Version 3.0. 2014
30. Yatoh, K., Sakamoto, K., Ishikawa, F., Honiden, S.: ArbitCheck: A Highly Automated Property-Based Testing Tool for Java. In: 2014 IEEE Seventh International Conference on Software Testing, Verification and Validation Workshops, pp. 405–412. IEEE, ??? (2014). doi:10.1109/ICSTW.2014.68. <http://ieeexplore.ieee.org/document/6825695/>
31. Luo, L.: Software testing techniques. Technical report, Software Engineering Institute, Carnegie Mellon University (2001). <https://www.cs.cmu.edu/texttildelowluluo/Courses/17939Report.pdf>
32. Sommerville, I.: Software Engineering. Addison-Wesley, ??? (2011)
33. Dijkstra, E.W.: Structured programming, pp. 1–82. Academic Press Ltd., London, UK, UK (1972). Chap. Chapter I: Notes on Structured Programming. <http://dl.acm.org/citation.cfm?id=1243380.1243381>
34. Yenigün, H., Yilmaz, C., Ulrich, A.: Advances in test generation for testing software and systems. International Journal on Software Tools for Technology Transfer **18**(3), 245–249 (2016). doi:10.1007/s10009-015-0404-z
35. Elberzhager, F., Rosbach, A., Münch, J., Eschbach, R.: Reducing test effort: A systematic mapping study on existing approaches. Information and Software Technology **54**(10), 1092–1106 (2012). doi:10.1016/j.infsof.2012.04.007
36. Mahadik, P., Bhattacharyya, D., Kim, H.-j.: Techniques for automated test cases generation: A review. International Journal of Software Engineering and Its Applications **10**, 13–20 (2016). doi:10.14257/ijseia.2016.10.12.02
37. Xu, D., Xu, W., Kent, M., Thomas, L., Wang, L.: An automated test generation technique for software quality assurance. IEEE Transactions on Reliability **64**(1), 247–268 (2015). doi:10.1109/TR.2014.2354172
38. Hooda, I., Chhillar, R.: A review: Study of test case generation techniques. International Journal of Computer Applications **107**(16), 33–37 (2014). doi:10.5120/18839-0375
39. Elghondaky, R., Moussa, S., Badr, N.: A comprehensive study for software testing and test cases generation paradigms. In: Proceedings of the International Conference on Internet of Things and Cloud Computing. ICC '16, pp. 50–1507. ACM, New York, NY, USA (2016). doi:10.1145/2896387.2896435. <http://doi.acm.org/10.1145/2896387.2896435>
40. Ali, S., Briand, L.C., Hemmati, H., Panesar-Walawege, R.K.: A Systematic Review of the Application and Empirical Investigation of Search-Based Test Case Generation. IEEE Transactions on Software Engineering **36**(6), 742–762 (2010). doi:10.1109/TSE.2009.52
41. Harman, M., McMinn, P.: A Theoretical and Empirical Study of Search-Based Testing: Local, Global, and Hybrid Search. IEEE Transactions on Software Engineering **36**(2), 226–247 (2010). doi:10.1109/TSE.2009.71
42. Sahoo, R.K., Ojha, D., Mohapatra, D.P., Patra, M.R.: Automated Test Case Generation and Optimization : A Comparative Review. International Journal of Computer Science and Information Technology **8**(5), 19–32 (2016). doi:10.5121/ijcsit.2016.8502
43. Khurana, N., Chhillar, R.S.: A comparison of evolutionary techniques for test case generation and optimization. Journal of Theoretical and Applied Information Technology **95**(19), 5285–5295 (2017)
44. Lampropoulos, L., Gallois-Wong, D., Hrițcu, C., Hughes, J., Pierce, B.C., Xia, L.-y.: Beginner's luck: a language for property-based generators. In: Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages - POPL 2017, pp. 114–129. ACM Press, New York, New York, USA (2017). doi:10.1145/3009837.3009868. 1607.05443. <http://dl.acm.org/citation.cfm?doid=3009837.3009868>
45. Loscher, A., Sagonas, K.: Automating Targeted Property-Based Testing. In: 2018 IEEE 11th International Conference on Software Testing, Verification and Validation (ICST), pp. 70–80. IEEE, ??? (2018). doi:10.1109/ICST.2018.00017. <https://ieeexplore.ieee.org/document/8367037/>
46. Earle, C.B., Fredlund, L.-A., Marino, J., Arts, T.: Teaching Students Property-Based Testing. In: 2014 40th EUROMICRO Conference on Software Engineering and Advanced Applications, pp. 437–442. IEEE, ??? (2014). doi:10.1109/SEAA.2014.74. <http://ieeexplore.ieee.org/document/6928850/>
47. Castro, L.M.: Advanced management of data integrity: property-based testing for business rules. Journal of Intelligent Information Systems **44**(3), 355–380 (2015). doi:10.1007/s10844-014-0335-2
48. Claessen, K., Hughes, J.: QuickCheck: A Lightweight Tool for



- Random Testing of Haskell Programs. In: Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming - ICFP '00, vol. 35, pp. 268–279. ACM Press, New York, New York, USA (2000). doi:10.1145/351240.351266. <http://portal.acm.org/citation.cfm?doid=351240.351266>
49. Chepur, A., Rathee, M.: Checking laws of the blockchain with property-based testing. In: 2018 International Workshop on Blockchain Oriented Software Engineering (IWBOSE), vol. 2018-Janua, pp. 40–47. IEEE, ??? (2018). doi:10.1109/IWBOSE.2018.8327570. <http://ieeexplore.ieee.org/document/8327570/>
  50. Clements, P., Bachmann, F., Bass, L., Garlan, D., Ivers, J., Little, R., Merson, P., Nord, R., Stafford, J.: Documenting Software Architectures: Views and Beyond. Addison-Wesley Professional, ??? (2011)
  51. Len, B., Paul, C., Rick, K.: Software Architecture in Practice. Addison-Wesley, ??? (2013)
  52. Alur, D., Malks, D., Crupi, J.a.: Core J2EE Patterns: Best Practices and Design Strategies, 2nd edn. Prentice Hall, ??? (2003)
  53. Parr, T.: The Definitive ANTLR 4 Reference. Pragmatic Bookshelf, ??? (2013)
  54. Wynne, M., Hellesoy, A., Tooke, S.: The Cucumber Book: Behaviour-driven Development for Testers and Developers. Pragmatic Bookshelf, ??? (2017)
  55. Crawley, M.J.: The R Book. John Wiley & Sons, ??? (2013)
  56. Fowler, M.: Patterns of Enterprise Application Architecture. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (2002)
  57. Buschmann, F., Meunier, R., Rohnert, H., Sommerlad, P., Stal, M.: Pattern-Oriented Software Architecture, Volume 1: A System of Patterns. Wiley, ??? (1996)
  58. Ordóñez, A., Eraso, L., Ordóñez, H., Merchan, L.: Comparing drools and ontology reasoning approaches for automated monitoring in telecommunication processes. *Procedia Computer Science* **95**, 353–360 (2016). doi:10.1016/j.procs.2016.09.345. Complex Adaptive Systems Los Angeles, CA November 2–4, 2016
  59. Li, D., Endle, C.M., Murthy, S., Stancl, C., Suesse, D., Sottara, D., Huff, S.M., Chute, C.G., Pathak, J.: Modeling and executing electronic health records driven phenotyping algorithms using the nqf quality data model and jboss® drools engine. In: AMIA Annual Symposium Proceedings, vol. 2012, p. 532 (2012). American Medical Informatics Association
  60. Mantas, J., *et al.*: Comparing drools and ontology reasoning approaches for telecardiology decision support. *Quality of Life Through Quality of Information: Proceedings of MIE2012* **180**, 300 (2012)
  61. Jung, C.Y., Sward, K.A., Haug, P.J.: Executing medical logic modules expressed in ardenml using drools. *Journal of the American Medical Informatics Association* **19**(4), 533–536 (2011)
  62. Mugridge, R., Cunningham, W.: Fit for Developing Software: Framework for Integrated Tests (Robert C. Martin). Prentice Hall PTR, Upper Saddle River, NJ, USA (2005)
  63. Wynne, M., Hellesoy, A.: The Cucumber Book: Behaviour-Driven Development for Testers and Developers. Pragmatic Bookshelf, ??? (2012)