A Systematic Approach to Assess Languages Proposed for Rules as Code

Ali, Sofia; Sileno, Giovanni; Van Engers, Tom

DOI
10.3233/FAIA251595

Publication date
2025

Document Version
Final published version

Published in
Legal Knowledge and Information Systems

Legal Knowledge and Information Systems
R. Markovich et al. (Eds.)

doi:10.3233/FAIA251595

A Systematic Approach to Assess Languages Proposed for Rules as Code

Sofia ALI [a,1], Giovanni SILENO [a] and Tom VAN ENGERS [a]
[a] *Informatics Institute, University of Amsterdam, the Netherlands*

Abstract. This paper presents a framework for evaluating approaches to formalizing legal sources. The framework integrates adoption-related criteria, extending the scope of assessment beyond purely technical dimensions. We build upon three contemporary comparison frameworks for normative specification languages, combining the Technology-Organization-Environment framework with grounded theory and adding criteria from other IT fields. As an illustration, we applied the resulting framework to RegelSpraak (RuleSpeak) from the Dutch Tax Administration.

Keywords. Rules as Code, Formal Language Comparison, Technology-Organization-Environment, Grounded theory, Technology adoption, RegelSpraak

1. Introduction

The rise of digitalization has led to a global increase in Rules as Code (RaC) initiatives aiming to encode regulations and directives into machine-readable formats [1,2], building on work in normative systems [3], on legal expert systems [4,5], and legal drafting requirements [6]. Various comparative frameworks have been proposed for formal languages and relative tools, focusing on logical languages [7], on tools for policy automation [8], or domain-specific languages (DSLs) [9,10], but a shared evaluation method is lacking so far. Studies often use limited and inconsistent/incomparable experiments [11], and focus primarily on formal and computational aspects while overlooking practical adoption. Against this background, we address the following questions: (RQ1) *which evaluation criteria recur in existing frameworks*; (RQ2) *which set of criteria covers adoption aspects relevant to practice*; and (RQ3) *what insights arise from applying the resulting framework.* We illustrate the approach with an evaluation of RegelSpraak, a controlled natural language (CNL) used by the Dutch Tax Administration (DTA).

Related works Formal legal languages are structured to translate legal texts into precise, machine-readable representations while preserving legal meaning and traceability to the source [11,12]. As the number of such legal representation languages has expanded, evaluation frameworks have become necessary. Early work [7] distinguished *epistemological*, *formal*, and *operational* criteria, while other studies [13] emphasized *organizational integration*, *maintainability*, and *strategic alignment*. Three contemporary comparative frameworks exist, each representing a unique evaluative viewpoint and a distinct practi-

[1]Corresponding author: `s.ali@uva.nl`. This work was partly funded by the Dutch Research Council (NWO) for the HUMAINER AI project (KIVI.2019.006).

cal focus. Chun et al. (2024) propose an eight-point comparative framework that extends prior work with usability, explainability, interoperability, and adoption-oriented criteria, and they apply it to legal DSLs [9]. Kaptijn and Klaver (2024) introduce an assessment framework for RaC with five scope categories and eight assessment aspects that integrate technical and user-oriented aspects [8]. Parvizimosaed et al. (2022) propose a ten-criterion framework for formal contract languages, organizing evaluation dimensions into ontology, language features, and analysis capabilities [10].

Research Gap Collectively, these three frameworks illustrate various evaluation perspectives, encompassing semantic and usability issues, compliance, contract and rule modeling, and user-centered design. However, none of these frameworks fully accounts for broader socio-technical factors, such as how well a language fits within an organization, how easily it can be adopted, or how it interacts with external conditions. Besides, various criteria are introduced without the support of more general literature, for instance, on the topic of technology adoption. This gap unveils the possibility of introducing a more integrated approach.

2. Method

Methodology To answer our research questions, we follow a three-fold approach. In an *analysis* phase, we collect the evaluation criteria proposed in recent comparison studies and organize them using the *Technology-Organization-Environment* (TOE) framework, initially developed in [14] and applied in technological innovation literature. In an *analogy/transfer learning* phase, we incorporate adoption-related criteria from other relevant IT fields, such as cloud computing and enterprise IT. In a *synthetic* phase, we apply *grounded theory* [15] to the identified criteria. The grounded theory process, via *open*, *axial*, and *selective* coding, allowed the identification of similarities, conceptual clusters, and broader evaluative dimensions. Through repeated analysis cycles, we resolved conceptual inconsistencies and semantic overlaps for a more coherent and empirically grounded evaluative structure.

As an illustrative example, we apply the resulting framework on RegelSpraak, a CNL based on RuleSpeak [16]. Its long-term use within the DTA for machine-readable tax rules provides a suitable setting to assess relevance.

Constructing the evaluative framework Applying this methodology, we first extracted evaluation criteria from three key contemporary studies on formal legal languages [8,9,10], presented in detail in the related works paragraph. We mapped all these criteria into the TOE framework, classifying each as technological, organizational, or related to the external environment. Table 1 presents this mapping. This clarifies that a formal legal language must be technically sound, adoptable by organizations, and suitable for the wider legal service provision environment. Where criteria could belong to multiple dimensions, e.g. *traceability* and *explainability*, we placed each criterion in the dimension it most directly applies.

As a second step, to ensure the framework is comprehensive, we added criteria from RaC-related studies and the broader literature on technology adoption. These additional factors capture organizations' challenges and drivers when adopting new technologies, complementing language-specific criteria. Table 2 shows the added factors within each TOE dimension.

TOE Dimension	Evaluation criteria	Sources
Technology	Analysis; Automated Verification; Capable of expressing a wide variety of contract genres; Capable of expressing a wide variety of legislative and regulative genres; Executable Analysis; Equipped with a formal semantics describing the language in terms of its underlying logics; Legal Concepts; Observables; Programming Paradigm; Time Support; Coding; Contract Reparations	[8,9,10]
Organization	Affordability; Explainability, Interpretation explanation; Transparency; Syntactically "low code" and user friendly with documentation and integrated development environment (IDE) support; Maturity; User-centered	[8,9]
Environment	Compliance; Compliance Monitoring; Easy to integrate; Oriented toward interoperability; Subcontracting; Traceability; Traceable Identifiability; Application-oriented; Open-source implementation available	[8,9,10]

Table 1. Mapping of evaluation criteria from key sources to TOE dimensions

Criteria per TOE dimension	Sources
Technology	
Relative advantage	[17,18,19,20]
Complexity	[17,21,19,20]
Compatibility	[20,17,21,18]
Security concerns	[17,18,20,19]
Expressiveness	[1,2]
Organization	
Top management support	[22,20,17,19,23,18]
Organization size	[20,18,17]
Cost savings	[20,17,19]
Readability by both legal and technical experts	[1]
Environment	
Political influence	[22,24]
Regulatory support	[25,26,18]
Public interpretability of encoded rules	[1,2]
Interoperability	[2,1]
Transparency of encoded rules	[2,1]

Table 2. Additional adoption criteria

As a third step, we merged and simplified all criteria into a final framework applying grounded theory. These categories, along with concise definitions, are shown in Table 3. The result is a comprehensive evaluation framework considering both technical and socio-organizational perspectives.

3. Case study: RegelSpraak

We applied the final TOE-grounded framework (Table 3) to RegelSpraak to validate the framework's utility. Table 4 presents the assessment per criterion: ✓ (criterion is supported), ✗ (criterion is not supported), and – (criterion is not applicable to the language's scope or design). In case of partial support, we used both symbols ✓✗.

Grounded Theory Category	Definition
Technology	
Formal semantics and foundations	Ensures the language has a rigorous, unambiguous formal basis
Expressiveness	Ability to capture a broad range of rule types and domain logic
Traceability	Links between code artifacts and their legal source references
Time management	Handling of time-dependent aspects and evolving state
Reasoning and logic	Advanced logic features for legal reasoning and inference
Security	Ensures data protection and prevents unauthorized access
Implementation fit	Refers to how easily a method can be understood and integrated within existing systems
Organization	
User-centric design	Accessible tooling, designed for usability by stakeholders
Efficiency and operational advantage	Practical benefits in development and execution
Explainability	Capacity to explain or justify outcomes of norm execution
Resource availability and support	Availability of sufficient resources, and leadership support
Maintainability and maturity	Stability, ease of updates, and maintenance processes
Environment	
Regulatory alignment	Alignment with external policy and regulatory requirements
Industry trends, external stakeholder demand	Alignment with broader trends and stakeholder requirements
International collaboration	Facilitation of international standards and cooperation

Table 3. Grounded Theory merged categories

The results demonstrate how RegelSpraak performs against the proposed criteria. It has a formal rule-based structure, suited for legal and regulatory domains, and works effectively for internal use by domain experts. However, it lacks formal and computational semantics, which disables automated reasoning and verification, while execution is only supported through *Agile Law Execution Factory (ALEF)* rather than directly in the language. From an organizational perspective, however, RegelSpraak benefits from resource availability and support, and its efficiency and operational advantage are evident in streamlined rule management at the DTA. The language also shows an implementation fit, integrating with *iKnow Rules* and *ALEF*. Given the language's non-executable scope, only security concerns were marked as not applicable. The environmental dimension, however, shows some challenges. Regulatory alignment is partial and oriented to internal goals. While external interest is growing, wider uptake depends on ALEF becoming open source, and there are no international collaborations or cross-national reuse cases. Overall, RegelSpraak shows strong socio-technical characteristics in its domestic context, but the absence of logic automation and direct execution semantics constrains broader applicability or transferability.

4. Conclusion

This paper presents a structured comparison framework for formal legal languages. The framework combines criteria from previous studies, the TOE framework, and related IT literature, and is constructed by applying grounded theory. For RQ1, we observed that contemporary frameworks consider adoption-focused elements, including usability, in-

Grounded Theory Category	RegelSpraak
Technology	
Formal semantics and foundations	✓✗ Provides a structured and unambiguous syntax, but lacks an explicit semantics as in logic-based languages.
Expressiveness	✓✗ Supports both fiscal and general rule modeling, though not as broadly expressive as formal logic languages.
Traceability	✗ Provides no inherent mechanism for linking code artifacts to legal sources, with traceability achieved through supporting methodology (Wetsanalyse) and tool (iKnow Cognitatie).
Time management	✓✗ Supports specification of time-dependent rules, with execution provided through ALEF.
Reasoning and logic	✗ Allows specification of repair and consistency rules that handle violations, but lacks a full reasoning engine; execution depends on ALEF.
Security	– Not applicable, as it does not handle directly execution-level access or data control
Implementation fit	✓ Integrates into IT systems like ALEF and iKnow Rules; low perceived complexity
Organization	
User-centric design	✓ Natural language structure with low-code logic improves usability
Efficiency and operational advantage	✓ Proven reductions in manual processing and improved decision- making
Explainability	✓✗ Produces structured, human-readable rule interpretations for both internal and external users; execution is however supported through ALEF rather than directly in the language.
Resource availability and support	✓ Backed by institutional support and technical expertise
Maintainability and maturity	✓ Long-term operational use; documented updates and institutional anchoring
Environment	
Regulatory alignment	✓✗ Primarily focused on national/internal goals; not aligned with external policy frameworks
Industry trends, external stakeholder demand	✓✗ Increasing interest from external stakeholders and public agencies, though wider adoption depends on ALEF becoming open source.
International collaboration	✗ No current collaborations or cross-national reuse cases

Table 4. Evaluation of RegelSpraak against the TOE framework

teroperability, explainability, and alignment with user needs. For RQ2, we observed that mapping these to TOE provided a balanced view but revealed gaps that necessitated additional criteria from adjacent IT adoption literature, reorganized using grounded theory. For RQ3, applying the refined framework to RegelSpraak revealed socio-technical factors and technical limits that earlier comparisons overlooked. The framework, therefore, offers a practical, wider basis for evaluation. In the future, we will consolidate the operationalization of criteria, applying our framework to multiple formal legal knowledge representation languages, and conduct additional iterations to test and refine its structure.[2]

[2]We acknowledge, for instance, that the specific scope/problem domain targeted by the tools under scrutiny is currently underdeveloped. This issue derives from the TOE framework, for which the scope remains implicit in the "relative advantage" category, always defined with respect to a certain problem domain.

References

[1] Rapson J, McLaughlin PA, Wong MW, Morris J, Pope R, Schneider M, et al. Rules as Code for a More Transparent and Efficient Global Economy. T7 Task Force 2: Digitalization of the Global Economy, Policy Brief. 2025 April:1-20.

[2] Mohun J, Roberts A. Cracking the code: Rulemaking for humans and machines. OECD Working Papers on Public Governance. 2020;42(1-109).

[3] Sergot MJ, Sadri F, Kowalski RA, Kriwaczek F, Hammond P, Cory HT. The British Nationality Act as a logic program. Communications of the ACM. 1986;29(5):370-86.

[4] Prakken H, Sartor G. Modelling reasoning with precedents in a formal dialogue game. In: Judicial applications of artificial intelligence. Springer; 1998. p. 127-83.

[5] Bench-Capon T, Sartor G. A model of legal reasoning with cases incorporating theories and values. Artificial Intelligence. 2003;150(1-2):97-143.

[6] Coupette C, Hartung D, Beckedorf J, Böther M, Katz DM. Law smells: Defining and detecting problematic patterns in legal drafting. arXiv preprint arXiv:211011984. 2021.

[7] Fensel D, Harmelen FV. A comparison of languages which operationalize and formalize KADS models of expertise. The Knowledge Engineering Review. 1994;9(2):105-46.

[8] Kaptijn B, Klaver J. Assessment of Rules as Code methods concerning the goal of assisting citizens personally with regulations. Digicampus voor GovTech4All; 2024.

[9] Chun A, Wong MW, Lauritsen M. Domain-Specific Languages and Legal Applications. The Journal of Robotics, Artificial Intelligence & Law. 2024;7(1):19-42.

[10] Parvizimosaed A, Sharifi S, Amyot D, Logrippo L, Roveri M, Rasti A, et al. Specification and analysis of legal contracts with Symboleo. Software and Systems Modeling. 2022;21(6):2395-427.

[11] Novotná T, Libal T. An Evaluation of Methodologies for Legal Formalization. In: Explainable and Transparent AI and Multi-Agent Systems. Cham: Springer International Publishing; 2022. p. 189-203.

[12] Bench-Capon TJM, Coenen FP. Isomorphism and legal knowledge based systems. Artificial Intelligence and Law. 1992 Mar;1(1):65-86.

[13] Schuwer RV. Het nut van kennissystemen; 1993.

[14] Tornatzky LG, Fleischer M. The Processes of Technological Innovation. Lexington Books; 1990.

[15] Tie YC, Birks M, Francis K. Grounded theory research: A design framework for novice researchers. SAGE open medicine. 2019;7(1):2050312118822927.

[16] Corsius M, Hoppenbrouwers S, Lokin M, Baars E, Cappellen GSV, Wilmont I. RegelSpraak: a CNL for executable tax rules specification. In: Proceedings of the seventh international workshop on controlled natural language (CNL 2020/21); 2021. .

[17] Al Hadwer A, Tavana M, Gillis D, Rezania D. A systematic review of organizational factors impacting cloud-based technology adoption using technology-organization-environment framework. Internet of Things. 2021;15:10040.

[18] Chittipaka V, Kumar S, Sivarajah U, Bowden JLH, Baral MM. Blockchain Technology for Supply Chains operating in emerging markets: an empirical examination of technology-organization-environment (TOE) framework. Annals of Operations Research. 2023;327(1):465-92.

[19] Park JH, Kim MK, Paik JH. The factors of technology, organization and environment influencing the adoption and usage of big data in Korean firms. 26th ITS Conference. 2015.

[20] Oliveira T, Thomas M, Espadanal M. Assessing the determinants of cloud computing adoption: An analysis of the manufacturing and services sectors. Information & management. 2014;51(5):497-510.

[21] Li L, Min X, Guo J, Wu F. The influence mechanism analysis on the farmers' intention to adopt Internet of Things based on UTAUT-TOE model. Scientific Reports. 2024;14(1):15016.

[22] Adade D, de Vries WT. An extended TOE framework for local government technology adoption for citizen participation: Insights for city digital twins for collaborative planning. Transforming Government: People, Process and Policy. 2024;19(1):53-73.

[23] Zhang H, Xiao J. Assimilation of social media in local government: an examination of key drivers. The Electronic Library. 2017;35(3):427-44.

[24] Adnan HR, Hidayanto AN, Kurnia S. Citizens' or government's will? Exploration of why indonesia's local governments adopt technologies for open government. Sustainability. 2021;13(20):11197.

[25] Gökalp E, Gökalp MO, Çoban S. Blockchain-based supply chain management: understanding the determinants of adoption in the context of organizations. Information systems management. 2022;39(2).

[26] Jans W, Denters B, Need A, Van Gerven M. Mandatory innovation in a decentralised system: The adoption of an e-government innovation in Dutch municipalities. Acta Politica. 2016;51:36-60.