

A DIGITAL LEAN FRAMEWORK FOR EARLY-STAGE URBAN REGULATORY DECISION-MAKING

Caroll Francesconi¹, Eric Forcael², & Reinaldo Valdebenito³

ABSTRACT

Early-stage regulatory decisions play a critical role in the delivery of construction projects. However, regulatory processes are frequently characterized by fragmented information, limited transparency, and iterative rework, generating significant waste that contradicts Lean Construction principles. From a Lean perspective, regulatory decision-making can be understood as a production system whose performance directly affects downstream planning, cost, and schedule reliability. This paper proposes a digital, Lean-oriented framework that supports early-stage regulatory decision-making by classifying administrative records using data-driven methods. Using secondary, publicly available data from urban land-use certificates and building permit records, the study applies Machine Learning techniques to structure, classify, and standardize regulatory information, illustrating how such data can be organized to support earlier feasibility analysis. The proposed framework shifts regulatory analysis from expert-dependent interpretation to a system-level information flow, potentially enhancing predictability and reducing rework at the front end of projects. Because of its exploratory nature, the findings of the present study provide a structured basis for improving interpretability and consistency in regulatory decision-making. In this sense, the research contributes to the Lean Construction literature by extending Lean thinking to regulatory processes and proposing a transferable, data-driven framework for complex urban development contexts.

KEYWORDS

Lean construction, urban regulations, decision-making, machine learning.

INTRODUCTION

In the Architecture, Engineering, Construction, and Operation (AECO) industry, early-stage decisions play a critical role in shaping project feasibility, cost, and schedule performance. Among these decisions, regulatory assessments related to land use and permitting conditions strongly influence the viability of development alternatives and the reliability of subsequent planning activities. In this sense, the AECO industry increasingly relies on diverse tools to support rapid decision-making in the early stages of projects (Van Oijstaeijen et al., 2023). However, regulatory processes are frequently characterized by fragmented information, limited transparency, and iterative rework, generating uncertainty and delays that contradict Lean Construction (LC) principles. From a Lean perspective, these processes can be understood as

¹ Assistant Professor, Facultad de Ingeniería, Universidad San Sebastián, Concepción 4081339, Chile; caroll.francesconi@uss.cl, orcid.org/0009-0008-0667-1805

² Professor, Facultad de Ingeniería, Universidad San Sebastián, Concepción 4081339, Chile; eric.forcael@uss.cl, orcid.org/0000-0002-3036-4329

³ Lecturer, Department of Construction and Risk Prevention, Universidad Técnica Federico Santa María, Concepción 4603255, Chile; reinaldo.valdebenito@usm.cl, orcid.org/0009-0003-8384-032X

part of the production system, where deficiencies in information flow and standardization generate waste and affect downstream project performance.

It is well known that LC conceptualizes production as a system focused on value generation through reliable flow and waste reduction (Garcés et al., 2025). While Lean principles have been widely applied to construction planning, design management, and production control, their application to early-stage regulatory processes remains limited. At the same time, advances in digitalization and data analytics create opportunities to structure regulatory information more systematically. In this context, the research objective of this study is to propose a digital Lean-oriented framework that conceptualizes regulatory decision-making as a production system supported by data-driven classification of administrative records using machine learning (ML) techniques. Using publicly available secondary data from urban land-use certificates and building permit records, the study adopts an exploratory approach to support early-stage decision-making. Although prior studies have explored Lean applications in planning and control, limited attention has been given to structuring and classifying regulatory information to support decision-making before construction. This study illustrates how regulatory data can be systematically organized to improve consistency in early-stage feasibility assessment.

LEAN CONSTRUCTION

Lean Construction (LC) emerged as a production management theory focused on minimizing waste and maximizing value through reliable flow and coordination of activities (Howell, 1999; Koskela, 2020). Rooted in principles from Lean Manufacturing and the Toyota Production System, LC conceptualizes production as the management of material, information, and decision flows required to support downstream processes (Liker, 2021; Womack et al., 1991). This perspective is encapsulated in the TFV (Transformation-Flow-Value) theory, which posits that production must be managed not just as a set of tasks, but as a continuous flow that generates specific value for the end customer (Tzortzopoulos et al., 2020). On the other hand, clear messaging and consistent information sharing are crucial for LC adoption (Romo et al., 2025).

However, fragmented information, lack of standardization, and variability are recognized as primary sources of waste (specifically, “non-value-adding” activities such as waiting and rework) that affect system performance (Hines & Taylor, 2000; Shingo & Dillon, 1989). While Lean principles have been widely applied to on-site production and the Last Planner® System (Ballard, 2000), their application to administrative and early-stage regulatory processes remains an emerging frontier (Tzortzopoulos et al., 2020). Addressing these processes is vital because regulatory uncertainty acts as a bottleneck, destabilizing downstream planning and cost reliability (Alarcón, 1997; Ballard & Howell, 2003).

DIGITALIZATION AND REGULATORY PROCESSES IN LEAN CONSTRUCTION

Digitalization has been increasingly recognized as a Lean enabler when applied to support flow, transparency, and standardization rather than as an end in itself. Despite a lack of a methodology for the development and implementation of lean information concepts (Redeker et al., 2019), digital tools can improve coordination and information availability in planning and production systems, in which digitalization serves as a Lean enabler by increasing transparency and reducing the information waste inherent in complex urban bureaucracies. However, regulatory processes have received limited attention within Lean-oriented digitalization efforts.

Regulatory decision-making generates large volumes of administrative data that are often heterogeneous and weakly structured. From a Lean perspective, this lack of structure contributes to variability, rework, and delays. Advances in data analytics and artificial

intelligence provide opportunities to organize and classify regulatory information, supporting more consistent interpretation and earlier identification of feasibility constraints.

While the LC literature has advanced in optimizing on-site production and design management, it often views regulatory requirements as static external constraints, ignoring their role as dynamic, destabilizing factors in the production flow. This narrow perspective overlooks how regulatory uncertainty causes downstream unreliability. Without applying Transformation-Flow-Value (TFV) theory to administrative processes, research misses how early “information waste” affects the project lifecycle and customer value.

In summary, digitalization, often seen as a Lean enabler, mainly serves as a coordination tool rather than fixing the structural fragmentation of regulatory data. Redeker *et al.* (2019) highlight the lack of methodologies for creating lean information concepts capable of managing high-dimensional, heterogeneous records, where Lean approaches can positively impact information flow (Fernandes et al., 2024). However, current methods depend heavily on expert judgment, leading to variability and rework. Thus, this research proposes a data-driven framework to transform regulatory analysis from reactive to proactive and to standardize information flow.

METHODOLOGY

This section outlines the steps taken to conduct the research: Study Design, Data Sources, Data Preprocessing, and Analytical Approach. The methodology comprises analyses that incorporate both numerical and categorical variables derived from building permit records in a mid-sized city in southern Chile. Numerical variables include the number of housing units (reported on each permit issued by the municipality), the number of building floors, and the total building surface area (in square meters). Categorical variables correspond to land-use classification (e.g., residential, non-residential) and regulatory destination categories. These variables were preprocessed through normalization (numerical) and encoding (categorical) to enable their integration within the clustering model. The overall methodological workflow is summarized in Figure 1.

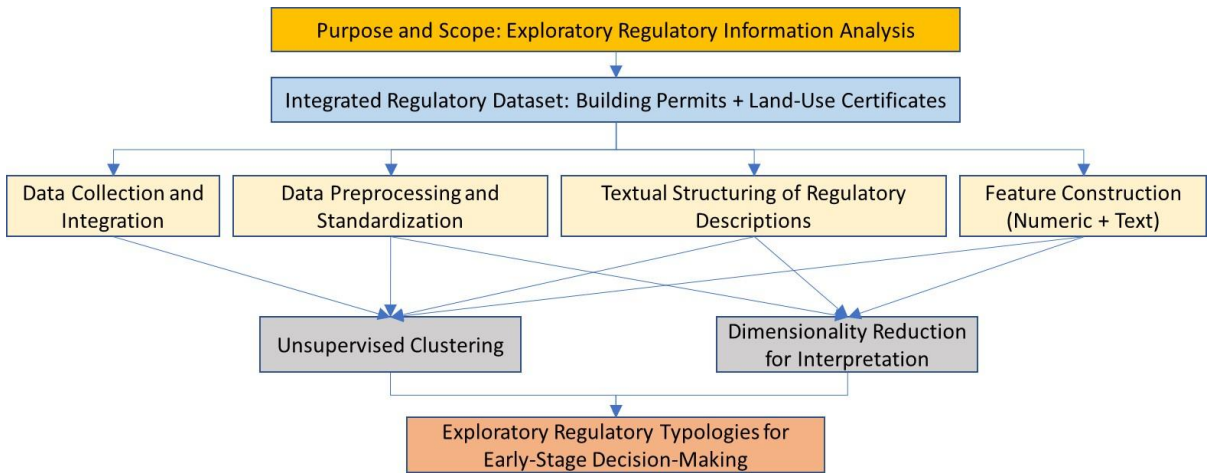


Figure 1: The overall methodological workflow.

STUDY DESIGN

This study adopts a descriptive, exploratory research design to illustrate how urban regulatory information can be structured and analyzed from a LC perspective. The methodological approach conceptualizes regulatory processes as information production systems, emphasizing information flow, variability, and standardization rather than performance evaluation. The

analysis is illustrative and supports early-stage regulatory decision-making in a representative Latin American urban context.

MACHINE LEARNING FRAMEWORK

In the era of rapid proliferation of operational data, a balanced approach that integrates advanced technologies and sound data governance is crucial for maintaining high data quality in the modern industrial context (Fu et al., 2024), where the AECO industry has not been absent. To structure the heterogeneous regulatory data, this study utilizes an unsupervised ML pipeline. Specifically, the K-means clustering algorithm (Likas et al., 2003) was employed to partition the administrative records into distinct groups based on feature similarity. Because the dataset contains high-dimensional features, including quantitative metrics and vectorized text descriptions, Singular Value Decomposition (SVD) (Akritas & Malaschonok, 2004) was employed to reduce dimensionality. SVD projects complex regulatory data into a two-dimensional space, preserving the most significant variance while enabling the visual interpretation of the resulting clusters (Jolliffe, 2002). This combination shifts regulatory analysis from subjective expert interpretation to a standardized, data-driven flow.

DATA SOURCES

The study relies on secondary administrative records obtained from two complementary sources. The first dataset comprises building permit records retrieved from official public databases, covering the period 2010-2024 and comprising more than 2,500 permits for the selected city. These records include quantitative, categorical, temporal, and geospatial attributes related to building characteristics and intended use. The second dataset comprises urban land-use certificate records from the local municipal planning authority, covering the same temporal period and containing more than 40,000 records. This dataset contains detailed textual descriptions related to intended activities, regulatory conditions, and zoning-related observations. Together, both datasets provide a comprehensive representation of formal regulatory decisions and their underlying informational content.

DATA PREPROCESSING

Data preprocessing focused on reducing information variability and ensuring analytical consistency in line with Lean standardization principles. Records were filtered to a standard spatial and temporal scope (2010–2024) and cleaned to remove incomplete or inconsistent entries. Categorical attributes were normalized, numerical variables were scaled for comparability, and geospatial information was standardized. Textual regulatory descriptions were cleaned, normalized, and transformed into structured representations using term-frequency-based methods, thereby enabling the incorporation of qualitative regulatory information into the analysis while preserving the original intent.

ANALYTICAL APPROACH

In unsupervised learning, or “learning without a teacher”, the aim is to extract properties directly from a probability density without relying on a supervisor or “teacher” to give correct answers or error degrees for each observation (Hastie et al., 2009). In the present study, an unsupervised analytical approach was adopted to explore patterns within the integrated regulatory datasets. In this context, unsupervised learning is particularly suited to exploratory urban research in which predefined categories of regulatory complexity do not yet exist. Clustering techniques were applied to group regulatory cases based on their quantitative, categorical, and textual characteristics. Dimensionality reduction methods were used exclusively to support the visual interpretation of the resulting clusters. The analysis is exploratory and illustrative, aiming to identify regulatory typologies and information structures rather than predict outcomes or assess efficiency or performance improvements.

RESULTS

DESCRIPTIVE ANALYSIS OF REGULATORY RECORDS

In terms of peculiarities in construction production (Vrijhoef & Koskela, 2005), bulkiness, rootedness, idiosyncratic needs, one-of-a-kind production, temporary organization, or client-driven business logic can influence the early stages of a project, where regulatory records play a crucial role. In this sense, the descriptive analysis reveals substantial heterogeneity in the regulatory records across the study period (2010-2024). The dataset comprises 2,514 building permit records for a representative Latin American city (population approximately 1,000,000), integrating numerical attributes (e.g., number of units, floors, built-up area) with categorical descriptors of intended land use and destination.

Most records correspond to low-density residential developments characterized by single-unit, low-rise construction with limited built-up area. In contrast, a small subset of permits accounts for a disproportionately high share of housing units, floors, and total surface area. This marked imbalance indicates a highly skewed regulatory structure, in which a large volume of standardized, repetitive permits coexists with a limited number of large-scale developments.

From a LC point of view, this variability constitutes a critical source of informational complexity at early project stages. While small-scale permits follow relatively predictable and stable approval patterns, large-scale projects introduce greater uncertainty, coordination requirements, and interpretative effort, increasing the cognitive and procedural burden on regulatory decision-making.

EXPLORATORY CLASSIFICATION OF REGULATORY PATTERNS

An unsupervised clustering approach was applied to explore latent structural patterns within the regulatory dataset. Figure 2 presents the two-dimensional projection obtained through singular value decomposition (SVD), illustrating a clear separation between regulatory typologies based on a reduced set of integrated attributes.

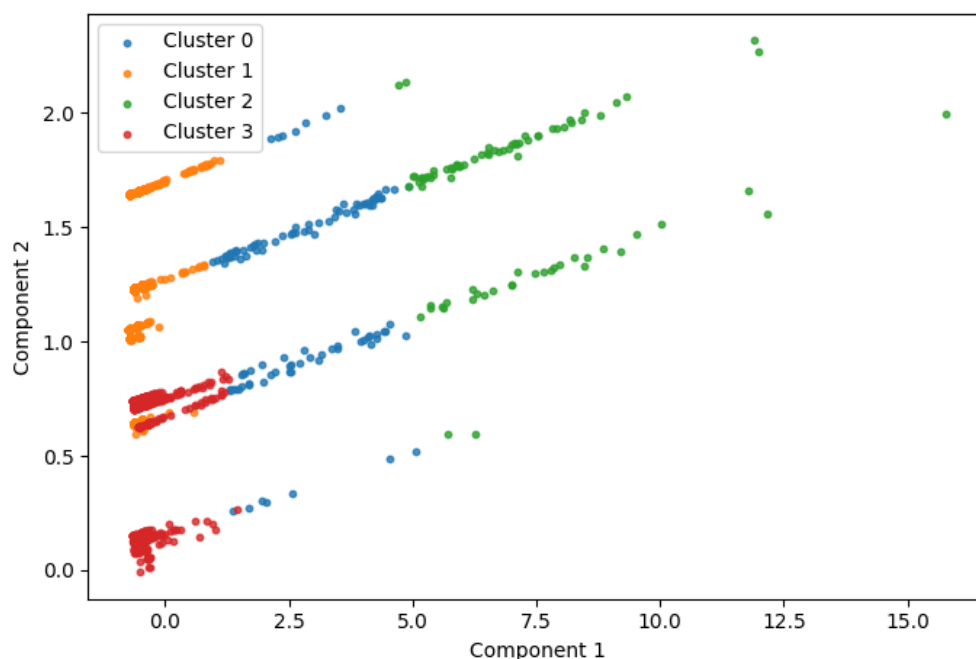


Figure 2: Two-dimensional SVD projection of regulatory permit clusters (61.2% explained variance).

The first two SVD components account for 61.2% of the total variance, indicating that a low-dimensional representation captures a substantial portion of the structural information in the regulatory records. The observed separation supports the internal coherence of the identified clusters and suggests that key differences between regulatory cases can be meaningfully summarized without relying on predictive or supervised assumptions.

The projection reveals distinct groupings associated with different regulatory profiles, ranging from small-scale residential permits to high-density, multi-story developments. This representation facilitates interpretability by highlighting structural contrasts in permit characteristics rather than individual project details.

Table 1 summarizes the main attributes of each identified cluster, including their relative frequency, temporal coverage, and representative quantitative and functional characteristics.

Table 1: Summary of identified regulatory permit clusters

Cluster	Records (n)	Period	Median Units	Median Floors	Median Surface (m ²)	Dominant Use	Dominant Destination
0	118	2010-2024	85.5	7.0	7,313.0	Residential	Single or block buildings
1	1,491	2010-2024	1.0	2.0	86.0	Residential	Detached housing
2	86	2010-2023	209.5	17.5	15,662.5	Residential	Single or block buildings
3	819	2010-2024	1.0	2.0	284.0	Non-residential	Commercial uses

Cluster 1 represents the dominant regulatory pattern, characterized by single-unit, low-rise residential permits with limited built surface. This cluster reflects highly standardized, repetitive approval processes that tend to generate stable, predictable information flows. Cluster 3 groups non-residential permits, primarily associated with commercial uses and moderate surface areas, introducing functional diversity into the regulatory landscape.

In contrast, Clusters 0 and 2 capture low-frequency but high-impact regulatory cases. These clusters correspond to large-scale, multi-story developments with substantially higher numbers of units and built surface. Although numerically marginal, they concentrate a significant share of regulatory complexity and represent structurally distinct cases that require greater analytical effort and coordination during early-stage evaluation. The temporal distribution of clusters (Figure 3) indicates persistent regulatory patterns. Cluster 1 remains dominant throughout the entire study period, while non-residential permits (Cluster 3) exhibit episodic peaks, suggesting periods of functional diversification in land-use approvals.

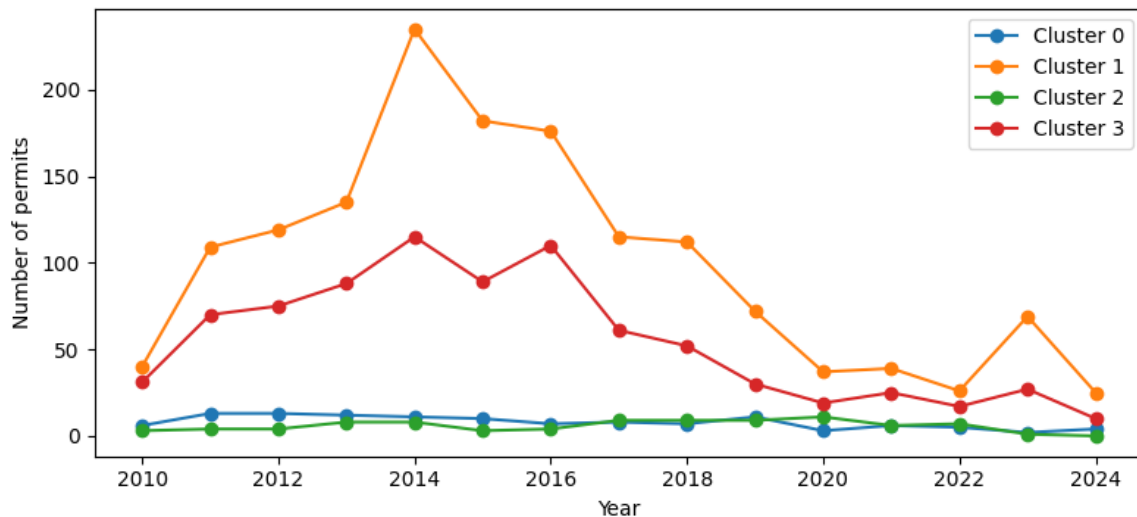


Figure 3: Temporal distribution of permit clusters.

Large-scale development clusters (Clusters 0 and 2) appear sporadically across the study horizon, reinforcing their role as exceptional regulatory events rather than recurrent patterns. The spatial visualization of regulatory clusters (Figure 4) further highlights differentiated territorial footprints within the urban area. Small-scale residential permits exhibit a dense, continuous spatial distribution, whereas large-scale developments are spatially concentrated and confined to specific zones.

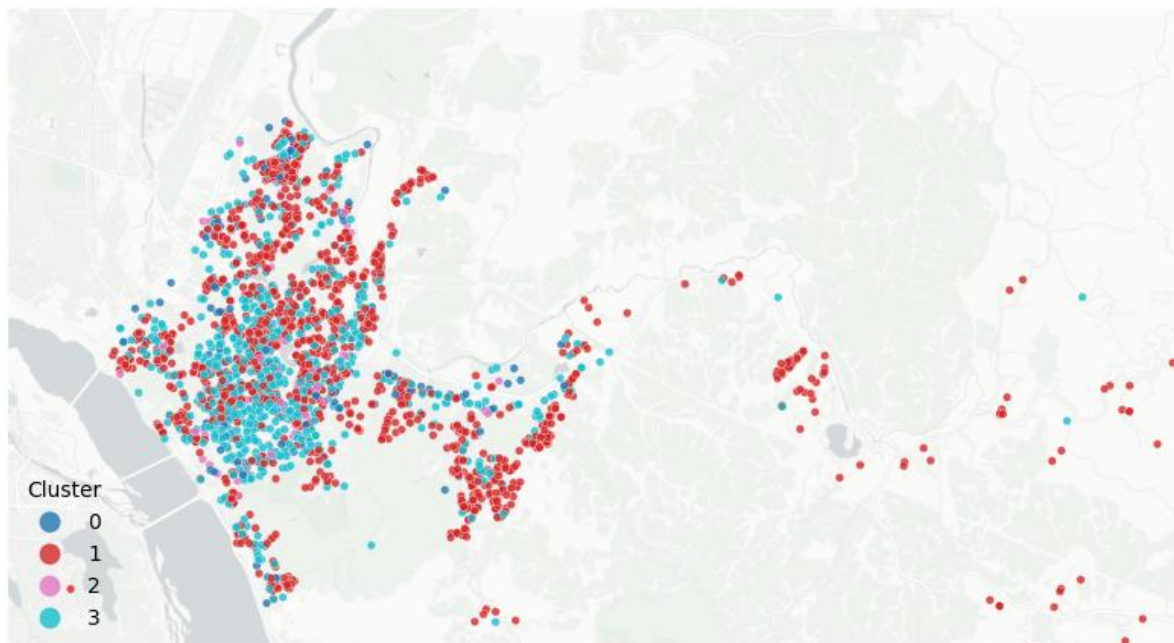


Figure 4: Spatial distribution of permit clusters

This spatial differentiation reinforces the interpretation of regulatory clusters as distinct informational typologies. Rather than isolated administrative records, the clusters emerge as structured signals that reflect recurring patterns of regulatory demand and complexity, supporting a Lean-oriented reading of regulatory data in early-stage project decision-making.

From a LC perspective, the classification results indicate improved information flow and standardization in the early stages. By structuring regulatory data, the framework enables more precise differentiation among regulatory scenarios, thereby supporting more reliable feasibility-related decisions.

LEAN-ORIENTED INTERPRETATION

From a Lean Construction perspective, the identified clusters represent distinct regulatory information patterns rather than isolated administrative cases. The dominance of small-scale residential permits reflects highly standardized and repetitive information flows with low variability, consistent with stable early-stage regulatory processes.

In contrast, large-scale development clusters, although infrequent, concentrate substantially higher informational complexity due to their scale and heterogeneity. These cases introduce greater uncertainty and interpretative effort during early-stage project assessment, challenging standardized regulatory processing.

Non-residential permits exhibit intermediate characteristics, introducing functional diversity without the structural complexity of large-scale developments. Overall, the results suggest that regulatory information can be organized into typologies with different Lean implications, supporting its use as structured input for early-stage decision-making.

DISCUSSION

The findings demonstrate that regulatory records, when treated as structured information rather than isolated administrative artifacts, reveal consistent and interpretable patterns. By utilizing K-means clustering and SVD for dimensionality reduction, this study operationalizes the Transformation-Flow-Value (TFV) theory, converting raw administrative data into a structured flow that serves as a reliable input for downstream planning. The clustering results show that a limited set of quantitative and categorical attributes is sufficient to distinguish dominant regulatory typologies, supporting the notion that early-stage regulatory decision-making operates through relatively stable information structures.

From a LC perspective, the observed imbalance between highly standardized permits and low-frequency, high-complexity cases highlights a key source of informational variability. This variability, particularly evident in the high-impact developments found in Cluster 2, represents a significant production bottleneck where information waste and rework are most likely to occur if not managed proactively. While most regulatory cases follow predictable patterns, a small subset accounts for a disproportionate share of complexity, uncertainty, and interpretative effort. This duality is rarely made explicit in traditional regulatory analysis but becomes visible through exploratory, data-driven approaches. The fact that the SVD projection captures 61.2% of the total variance suggests that a substantial portion of what is often perceived as “regulatory noise” is, in fact, a set of identifiable signals that can be standardized to improve system predictability.

Importantly, the analysis does not aim to optimize or evaluate regulatory performance. Instead, it demonstrates how regulatory data can be reorganized into meaningful informational typologies that may support earlier and more informed project decisions, reducing reliance on purely experiential or ad hoc interpretations. In this sense, this shift from expert-dependent interpretation to a system-level information flow aligns with the principles of regulatory intelligence, facilitating the synchronization of urban land-use constraints with Lean construction control objectives.

In practical terms, identifying these clusters supports a “regulatory triage” approach, where resources are allocated based on each case’s complexity profile. For example, Cluster 1, which involves highly standardized, repetitive patterns with minimal variation, can be processed through automated or expedited approval processes. In contrast, projects in Cluster 2 (those with significant impact) should trigger an early-stage intensive coordination phase involving multi-disciplinary reviews. This strategy minimizes “over-processing” by preventing simple and low-risk projects from unnecessary delays caused by procedural demands suited to more complex developments.

Moreover, the implications for performance are most pronounced in enhanced planning reliability. By aligning a new project proposal with these historical clusters, developers can substitute arbitrary lead-time assumptions with empirical, data-driven estimates for regulatory approval. This reduction in variability is vital because regulatory uncertainty significantly destabilizes downstream production control and schedule reliability. Therefore, the ability to forecast whether a project belongs to a high-complexity “bottleneck” cluster (Cluster 2) or a standardized flow (Cluster 1) enables practitioners to strategically add buffers to the master schedule, thus shielding the production system from the volatility usually associated with urban land-use approvals.

CONCLUSIONS

This article presents an exploratory, Lean-oriented framework for analyzing regulatory information in the early stages of the construction production system. By extending Lean thinking to the regulatory sphere, this framework provides a foundation for regulatory intelligence that aligns administrative workflows with project delivery objectives. Using secondary administrative data from a representative Latin American city, the study illustrates how regulatory records can be structured, classified, and interpreted using unsupervised learning techniques. The main contribution lies in both theoretical and practical terms: theoretically, by extending Lean Construction thinking beyond project execution to the regulatory sphere through the conceptualization of regulatory processes as information production systems; and practically, by enabling a structured, data-driven approach that supports more consistent and interpretable decision-making in the early stages. The framework emphasizes information flow, variability, and standardization, rather than performance outcomes. However, as this is an exploratory study, the findings are limited by the focus on a single geographic context and the lack of quantitative metrics for waste reduction. Therefore, future research should explore integrating these data-driven typologies into BIM environments to improve transparency and real-time decision-making. The proposed approach remains transferable and adaptable, laying the groundwork for future research on digital regulatory intelligence and decision support in the early stages of construction.

ACKNOWLEDGMENTS

The authors thank the National Research and Development Agency (ANID) of the Government of Chile, ANID-Subdirección de Capital Humano/Doctorado Nacional/Grant numbers: 2026-21263366 and 2026-21263454, and Fondecyt Regular/Grant number: 1251708. This work was also supported by Vicerrectoría de Investigación y Doctorados de la Universidad San Sebastián-Fund numbers USS-FIN-26-CNGI-06 and USS-FIN-26-CNGD-16.

REFERENCES

- Akritis, A. G., & Malaschonok, G. I. (2004). Applications of singular-value decomposition (SVD). *Mathematics and Computers in Simulation*, 67(1–2), 15–31. <https://doi.org/10.1016/j.matcom.2004.05.005>
- Alarcón, L. (1997). Lean Construction. In L. Alarcón (Ed.), *Lean Construction*. Taylor & Francis.
- Ballard, G. (2000). *The Last Planner System of Production Control*. The University of Birmingham.
- Ballard, G., & Howell, G. (2003). Lean project management. *Building Research & Information*, 31(2), 119–133. <https://doi.org/10.1080/09613210301997>
- Fernandes, M., Correia, D., & Teixeira, L. (2024). Lean maintenance practices in the improvement of information management processes: a study in the Facility Management

- division. *Procedia Computer Science*, 232, 2269–2278. <https://doi.org/10.1016/j.procs.2024.02.046>
- Fu, Q., Nicholson, G. L., & Easton, J. M. (2024). Understanding data quality in a data-driven industry context: Insights from the fundamentals. *Journal of Industrial Information Integration*, 42, 100729. <https://doi.org/10.1016/j.jii.2024.100729>
- Garcés, G., Forcael, E., Osorio, C., Castañeda, K., & Sánchez, O. (2025). Systematic review of Lean Construction: an approach to sustainability and efficiency in construction management. *Journal of Infrastructure Preservation and Resilience*, 6(1), 1–28. <https://doi.org/10.1186/s43065-025-00119-1>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). Unsupervised Learning. In *The Elements of Statistical Learning* (pp. 485–585). Springer. https://doi.org/10.1007/978-0-387-84858-7_14
- Hines, P., & Taylor, D. (2000). *Going Lean*. Lean Enterprise Research Centre.
- Howell, G. A. (1999). What is Lean Construction - 1999. In I. D. Tommelein & G. Ballard (Eds.), *7th Annual Conference of the International Group for Lean Construction (IGLC)* (pp. 1–10). International Group for Lean Construction.
- Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Science and Business Media.
- Koskela, L. (2020). Theory of Lean Construction. In P. Tzortzopoulos, M. Kagioglou, & L. Koskela (Eds.), *Lean Construction - Core Concepts and New Frontiers* (1st ed., pp. 1–12). Routledge.
- Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451–461. [https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)
- Liker, J. (2021). *Toyota Way: 14 Management Principles from the World's Greatest Manufacturer* (2nd ed.). McGraw-Hill.
- Redeker, G. A., Kessler, G. Z., & Kipper, L. M. (2019). Lean information for lean communication: Analysis of concepts, tools, references, and terms. *International Journal of Information Management*, 47, 31–43. <https://doi.org/10.1016/j.ijinfomgt.2018.12.018>
- Romo, R., Orozco, F., Moreno, F., Francesconi, C., Valdebenito, R., & Forcael, E. (2025). Lean Construction Barriers From the Organizational Culture Lens. *33rd Annual Conference of the International Group for Lean Construction (IGLC 33)*, 1321–1332. <https://doi.org/10.24928/2025/0284>
- Shingo, S., & Dillon, A. P. (1989). *A study of the Toyota Production System: From an Industrial Engineering Viewpoint*. Productivity Press.
- Tzortzopoulos, P., Kagioglou, M., & Koskela, L. (2020). *Lean Construction* (P. Tzortzopoulos, M. Kagioglou, & L. Koskela (eds.)). Routledge. <https://doi.org/10.1201/9780429203732>
- Van Oijstaeijen, W., Silva, M. F. e, Back, P., Collins, A., Verheyen, K., De Beelde, R., Cools, J., & Van Passel, S. (2023). The Nature Smart Cities business model: A rapid decision-support and scenario analysis tool to reveal the multi-benefits of green infrastructure investments. *Urban Forestry & Urban Greening*, 84, 127923. <https://doi.org/10.1016/j.ufug.2023.127923>
- Vrijhoef, R., & Koskela, L. (2005). Revisiting the Three Peculiarities of Production in Construction. *13th Annual Conference of the International Group for Lean Construction*, 19–27. <https://iglc.net/Papers/Details/345>
- Womack, J. P., Jones, D. T., & Roos, D. (1991). *The Machine That Changed the World: The Story of Lean Production* (1st ed.). HarperCollins Publishers Inc.