

ABSTRACT

The rapid digital transformation within the infrastructure sector has led to the generation of large volumes of data across the life cycle of highway projects, particularly in India. Advanced tools such as Geographic Information Systems (GIS), Building Information Modeling (BIM), sensor networks, and the National Highways Authority of India's (NHAI) Data Lake initiative have significantly expanded data collection capacities. Examples include traffic volume counts, pavement condition surveys, environmental clearances, and project cost estimates. Yet, key omissions such as real-time sensor feeds, standardized documentation, and predictive maintenance data limit the usability of this otherwise vast dataset. Much of this data remains poorly structured, inaccessible, or detached from real-time decision-making, resulting in missed opportunities for performance improvement and project efficiency.

This research addresses this challenge by examining how data quality dimensions, stakeholder perspectives, and knowledge management barriers influence the usability of infrastructure data, and by developing an integrated framework to strengthen data-to-decision pathways. Stakeholders within NHAI were mapped and engaged through structured surveys and interviews, with the semiotic framework guiding the analysis of syntactic, semantic, empiric, and pragmatic dimensions. Structural Equation Modeling confirmed that data structure and accessibility exert the strongest influence on usability, while semantic richness alone does not guarantee effective usage. To address knowledge integration, 35 knowledge management barriers were identified and validated, then analyzed through Social Network Analysis to reveal interdependencies and systemic effects.

Based on the initial analysis of data usage assessment and knowledge management barriers, a novel Data-Information-Knowledge-Decision (DIKD) integration framework was

developed. The framework visualizes how data transitions through layers of interpretation and context to inform decisions. It identifies “active,” “inactive,” and “missing” data pathways, providing a structured method to diagnose weaknesses in data use. A Composite Integration Score (CIS) was introduced to quantify integration effectiveness, allowing organizations to evaluate their performance on a standardized scale. When applied to actual NHAI project datasets, the framework revealed moderate integration levels, 62.16% in the pre-construction phase and 69.34% post-construction on a 0-100 scale, where 100 represents complete integration. These scores indicate that while digital tools are in place, substantial gaps remain in data transformation and knowledge application.

Gap analysis using developed framework further revealed systemic inefficiencies in data-to-decision pathways. The framework's application provided actionable diagnostics, pinpointing where and why data fails to support timely decisions. It emphasizes the need for coordinated governance, cross-functional communication, and role-based data responsibility to ensure that data collection efforts translate into organizational intelligence.

This research makes significant contributions to both theory and practice by introducing an adaptable and empirically validated framework that integrates data science principles into infrastructure decision-making. The DIKD framework addresses critical disconnections between data generation and decision-making by aligning stakeholder needs, data quality expectations, and organizational workflows. It enables infrastructure agencies to benchmark their data integration maturity, identify systemic inefficiencies, and pursue targeted improvements. While grounded in the Indian highway context, the framework is designed to be transferable across infrastructure systems globally, promoting transparency, efficiency, and long-term sustainability through a holistic, stakeholder-sensitive approach.

Keywords: Infrastructure data integration, Decision-making frameworks, Knowledge management barriers, Social network analysis, Data quality, Semiotic framework.