**InfraLex: A digital dictionary for unambiguous data sharing in the civil infrastructure domain**

**Summary of research contribution**

The data consistency among project phases, stakeholders, or geographic regions (counties, states, etc.) is a major barrier to digital exchange of life cycle project data in the civil infrastructure industry. However, research to address the issue of terminology inconsistency in the construction industry has been very limited. This research aims to fill that gap by proposing a novel methodology for translating text-based domain knowledge into an extensive lexicon, namely InfraLex, for the domain of civil infrastructure. The lexicon formally organizes civil infrastructure technical terms in a lexical hierarchy manner in which semantic relations (synonyms, hyponyms, hypernyms, attributes, etc.) among terms are explicitly represented. The InfraLex dictionary enables computer systems to understand the meaning of a piece of data; hence, data mismatch when integrating isolated datasets will be eliminated. The study is, therefore, intended to make contributions in facilitating seamless data exchange of digital data across the life-cycle of an infrastructure project. Thanks to the explicit explanations of technical terms, InfraLex is also expected to become a fundamental resource for natural language processing studies in the civil infrastructure domain.