Research Statement

My research interests are in line with adding value to the existing isolated data of the civil infrastructure industry to reduce manual and duplicate efforts/waste in data collection, processing, and utilization by introducing new knowledge, computational intelligence infrastructure systems and methods that can support automated data extraction and mining. The adoption of advanced digital technologies such as Building Information Modeling (BIM), Geographic Information System (GIS), or Lidar throughout the project life-cycle has enabled a large portion of the project life cycle data to become available in digital format. The improved efficiency in sharing and utilizing of these complex machine-readable datasets, will in turn, translate into increased productivity, efficiency in project delivery and accountability. In the last three years of my doctoral program, I have conducted ample research to address various aspects of digital project delivery varying from theoretical to implementation research including: 1) life-cycle data linkage, 2) data terminology discrepancy among different sources, 3) natural language based partial model extraction engine and 4) data and information flow through transportation assets' life cycle. In my future research, I plan to continue work on enhancing the utilization efficiency and the value of digital project data and broad my research focus to the areas of intelligence systems and data mining from unstructured data sources of texts, images, sounds or videos.

Life-cycle data space linkage - Aiming to address the heterogeneity of data structure I developed a framework for linking life-cycle data. Asset management requires data that are generated from various upstream design and construction phases to condition survey. In the conventional practices, discipline divisions in a highway agency store and manage data individually in their own data inventory. To allow decision makers in asset management effectively inherit and reuse digital data created by upstream partners, I developed a novel framework that enables the interconnection of heterogeneous life-cycle digital data sources (BIM models, project inspection systems, and asset management systems) and translate into meaningful information. The platform includes several data translators that are able to convert data proprietary formats into connected graph networks of data. This proposed platform has been published on the Journal of Automation in Construction (Le & Jeong 2016). The findings from this study are expected to provide an effective and efficient means to facilitate seamless digital data exchange throughout the life cycle of a highway project. I plan to leverage the application of this platform from management of a single asset to integrated urban and infrastructure management to allow for concurrent collaboration between construction sectors (buildings, pipeline, railway, water supply, etc.). Once local data sets can be instantly accessed by other related disciplines, better decision making with holistic and long-term benefits would be achieved.

Inconsistency of data terminology - Data terminology discrepancy is a big hurdle to integration or sharing of digital data among multiple disciplines, partners, geographic regions. The lack of common understanding to the same or similar data presented in different terms can lead to the extraction of wrong data or misinterpretation. To enable semantic transparency for commonly used technical terms among highway agencies across the United State, I have developed a computational infrastructure that supports automated development of a machine-readable dictionary of American-English civil engineering terms. The proposed platform leverage Natural Language Processing (NLP) techniques and machine learning to extract English-American roadway terms and their meanings from natural language technical documents for instance roadway design manuals and specifications. The present algorithm would accelerate the process of removing the current bottleneck in machine readable dictionaries which are required for an unambiguous data sharing, integration or exchange. In future research, I'm interested in implementing the proposed method in developing an national mapping of technical terms among highway agencies across the United States which would have significantly beneficial impacts to integration of state historic project data to support data-driven infrastructure management.

Natural language based data retrieval engine - Regarding to data acquisition, I have recently developed a successful proposal (PI: Dr. Jeong) which is awarded for mostly $300,000 by National Science Foundation (NSF) to develop a computational partial model extraction platform that allow users to use plain English data requirements to query civil infrastructure digital data. Simple and easy extraction of desired data from large and complex machine-readable digital infrastructure data critically decides the degree of reusability of up-stream digital models and their associated project data. However, the state-of-practices on digital data retrieval in the civil infrastructure domain shows that the traditional ad-hoc data query, which is manual and error-prone, has imposed big burdens on professionals. Users are required to have deep understanding of data structure, meanings behind each data label and a query language. This research proposes a novel approach for a fast

and unambiguous reuse of digital models for the civil infrastructure industry by developing an automated data retrieval engine which is capable of recognizing user intention from their natural language queries (e.g., words, phrases, questions) and extracting the desired data from heterogeneous digital datasets. I'm currently the lead researcher in an interdisciplinary project team of Linguistics, Machine Learning and Construction Engineering and Management. In order to enable computer systems to understand users data requirements in natural language, I have been translating domain knowledge in design manuals, guidelines and specification into an extensive machine-readable dictionary for the civil infrastructure using my recently developed NLP-based method as mentioned above. Upon finishing this digital dictionary, I will utilize recent advances in Natural Language Processing (NLP) techniques, machine-learning based semantic measure methods to develop the data retrieval system. NLP will be utilized to process and interpret users natural language inputs. This research is expected to make transformative impacts on digital date exchange and sharing in the civil infrastructure industry, and promote and accelerate the industry transition to the digital project delivery as digital models and their associated data can be readily and seamlessly reused through the project life cycle. This research will fundamentally transform the way data users interact with and query digital modeling data and information in the civil infrastructure domain. I plan to extend research to speech language rather written format that allow user to communicate with digital models using natural speech language.

Data and information flow across the life-cycle of transportation assets - In the spectrum of implementation research that can be ready for highway agencies, a research proposal mainly contributed by myself is funded by Iowa Highway Research Board and Mid-West Transportation Center for $180,000 to enhance the understanding of data and information workflow during the life-cycle of transportation assets by capturing the industry knowledge and experience and developing a business process map and a data sharing map. To accomplish that objective, I have conducted series of working group discussions with various participant professionals from various divisions from different project phases, disciplines involved during the life-cycle of various transportation asset including signs and guardrails. Based on these discussions I have identified the workflow that require data sharing during the project life cycle and captured data exchange requirements specifying what data to be shared by whom and to whom. For each type of transportation assets, I have developed current practice and ideal process map and exchange requirement documents that show what data, who and when to be transferred to whom. This project is expected to provide a better understanding on data and information flow throughout the lifecycle for various transportation assets. This understanding will allow researchers and professionals to identify current roadblocks in digital data transferring, find a method for leverage existing data and reduce data-recreation and significantly enhance the process of data collecting and sharing between project participants. This current research is focused on the current digital data flow. In my future research, I plan to target NCHRP for a project that develop a national guide that can be used by highway agencies to evaluate their maturity in digital based project delivery, identify requirements and tools that need to help them facilitate seamless digital data transferring throughout the project life-cycle.

Future research - In addition to continue my research on data sharing and retrieval, my research long term goals are smart building and civil information model (smart BIM/CIM) and big data analytics. Smart BIM/CIM with integrated domain knowledge with can performed self-automated reasoning which can answer what if questions. My previous and current research on NLP related research provide profound impact to construction industry where natural language documents are still play major role in data and information sharing and communication among project stakeholders including project contracts, project inspection reports. Im enthusiastic to continue pursue my research career in line with the interests in developing both theoretical and applicable platform and computational infrastructure to assist researchers and professionals dig into information included in text documents, for instance translating federal and state design guidance, rules and requirements documented specifications in text documents into an extensive resources of digital constraints to support automated compliance checking, detection of disclosure statements in contractual documents that involves risk to contractors, process RFI (Request for Information). I plan to integrated these digital knowledge into BIM and CIM models to create smart digital models. The system can answer what if questions in design, planning of the project (design, planning, decision making, cost scheduling). I also plan to expand my research areas to big data analytics with focus on alternative data mining and alternative information source such texts, images, voices or videos to reduce data collection effort for decision making in construction job site. One-call center data is just message transferring to participant operators, in urban cities where numerous underground facilities utilities such as electricity, water pipe and sewer, gas, internet, cables using natural lan-

guage speed, excavator, location, locates, excavating, digging. Constructing BIM information models using current method such as Lidar are still have low accuracy and time-processing for those is costly. One-call center receive request and send it to participant members who are operators and owners of facilities who is responsible for locate/marking their utilities by paints or flags. Research is need to allow for construction of a digital library of those utilities. Average speed to answer takes hours and days. Potential sponsors: NSF, NIST, Federal Highway Agency, State DOTs. Benign involved in some project, work with dot, meeting with some other people from many state DOTs, their great interest in transferring to digital data project delivery. Final paragraph overall good expressions of my research