BUILD AMERICA CENTER

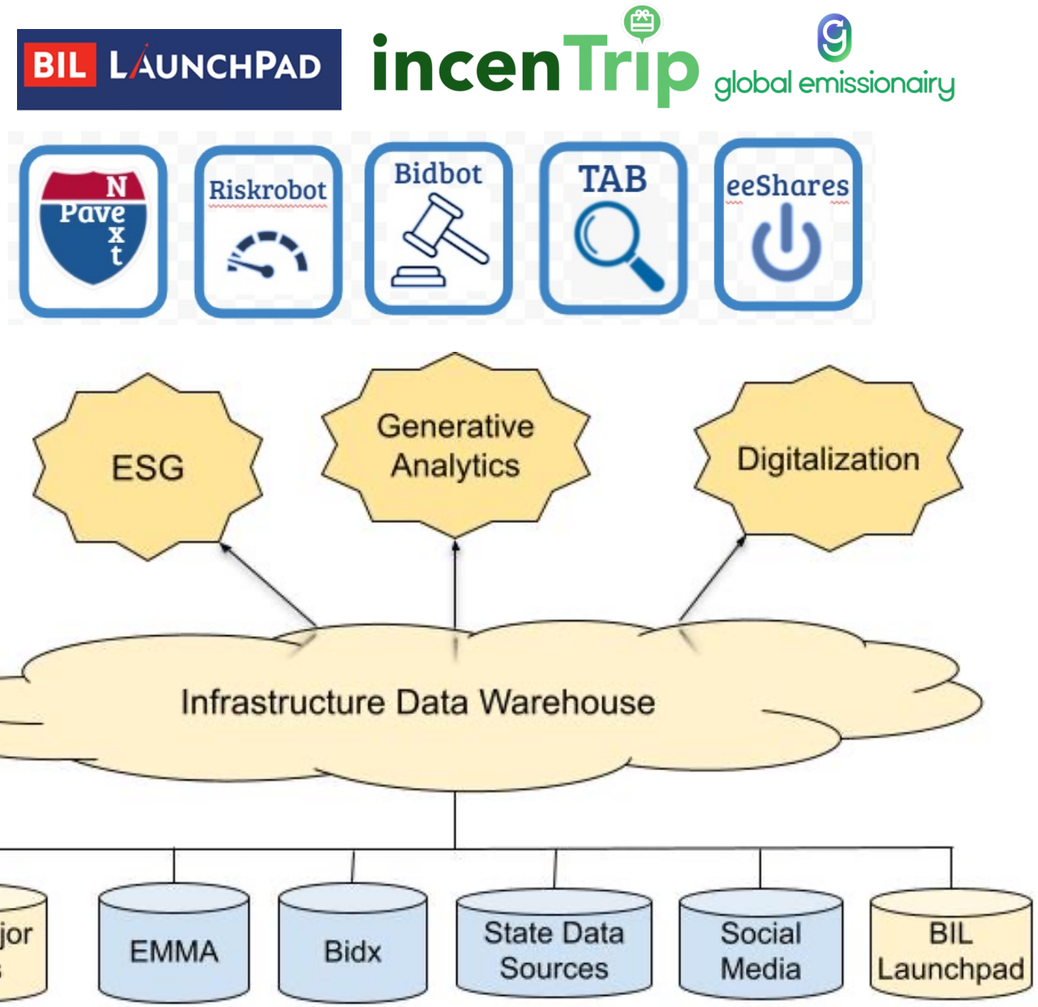
INNOVATIVE FINANCING AND DELIVERY
OF TRANSPORTATION INFRASTRUCTURE

The Power of Data: Transforming Decision-Making in Transportation Projects

Qingbin (QC) Cui

Our Research Platforms

The Build America Center (BAC) will mobilize the use of innovative financing, funding, and project delivery solutions to foster new approaches to transportation infrastructure development and delivery through creation of a knowledge hub, spurring innovation with cutting edge research, collaborative partnerships and the development of an academy to deliver innovative education and training plus tailored technical assistance.

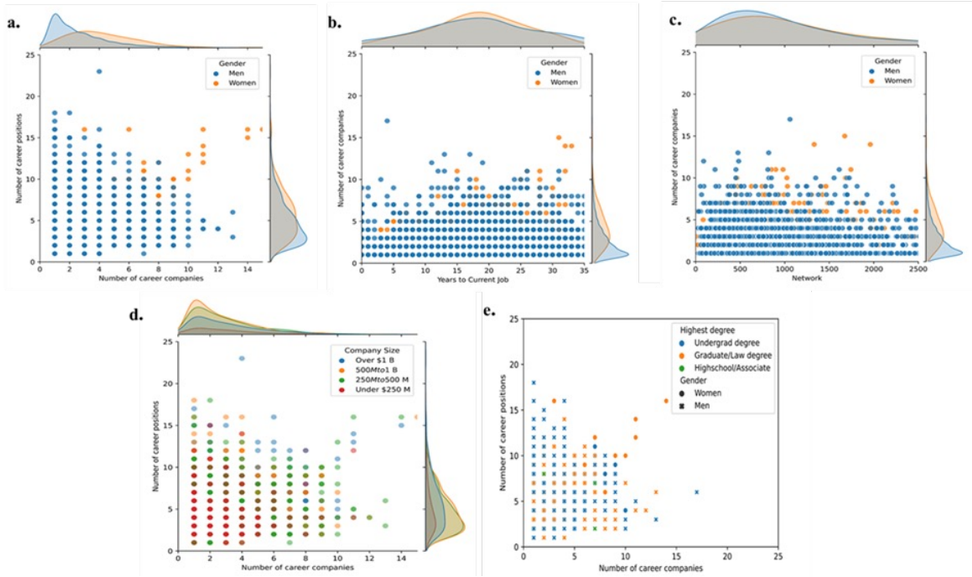


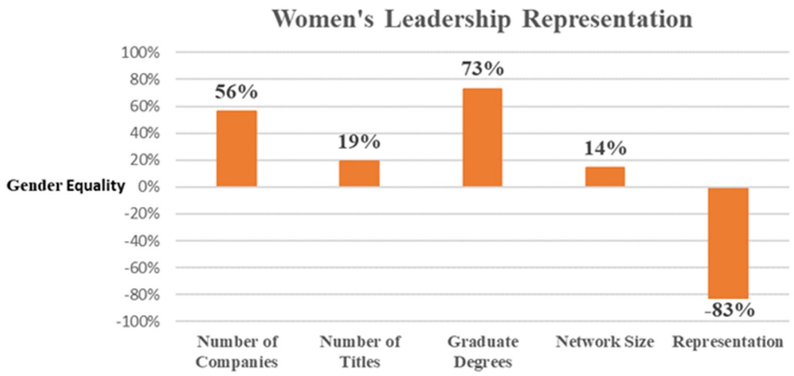
BUILD AMERICA CENTER
INNOVATIVE FINANCING AND DELIVERY OF TRANSPORTATION INFRASTRUCTURE

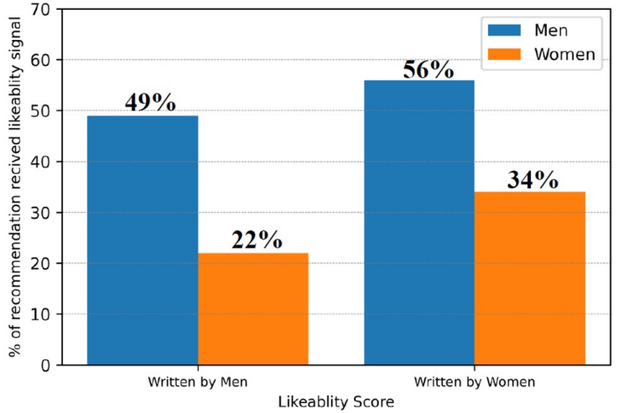
FEARLESS INNOVATION

Data Driven Project Management: Examples

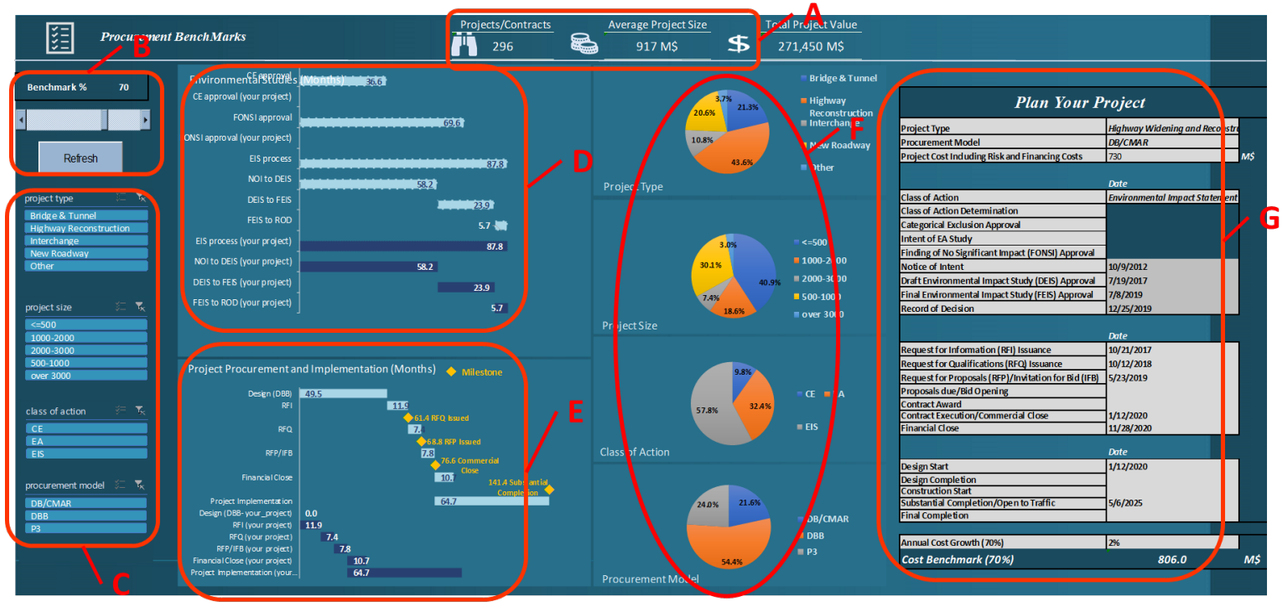
- Data Centric Risk Management
- Benchmarking and Planning
- Price Manipulation Detection
- Public Perception and Engagement
- Bidding Intelligence
- Data Driven Career Paths and Workforce Development
- Life-Cycle Emission Analysis
- And many other applications

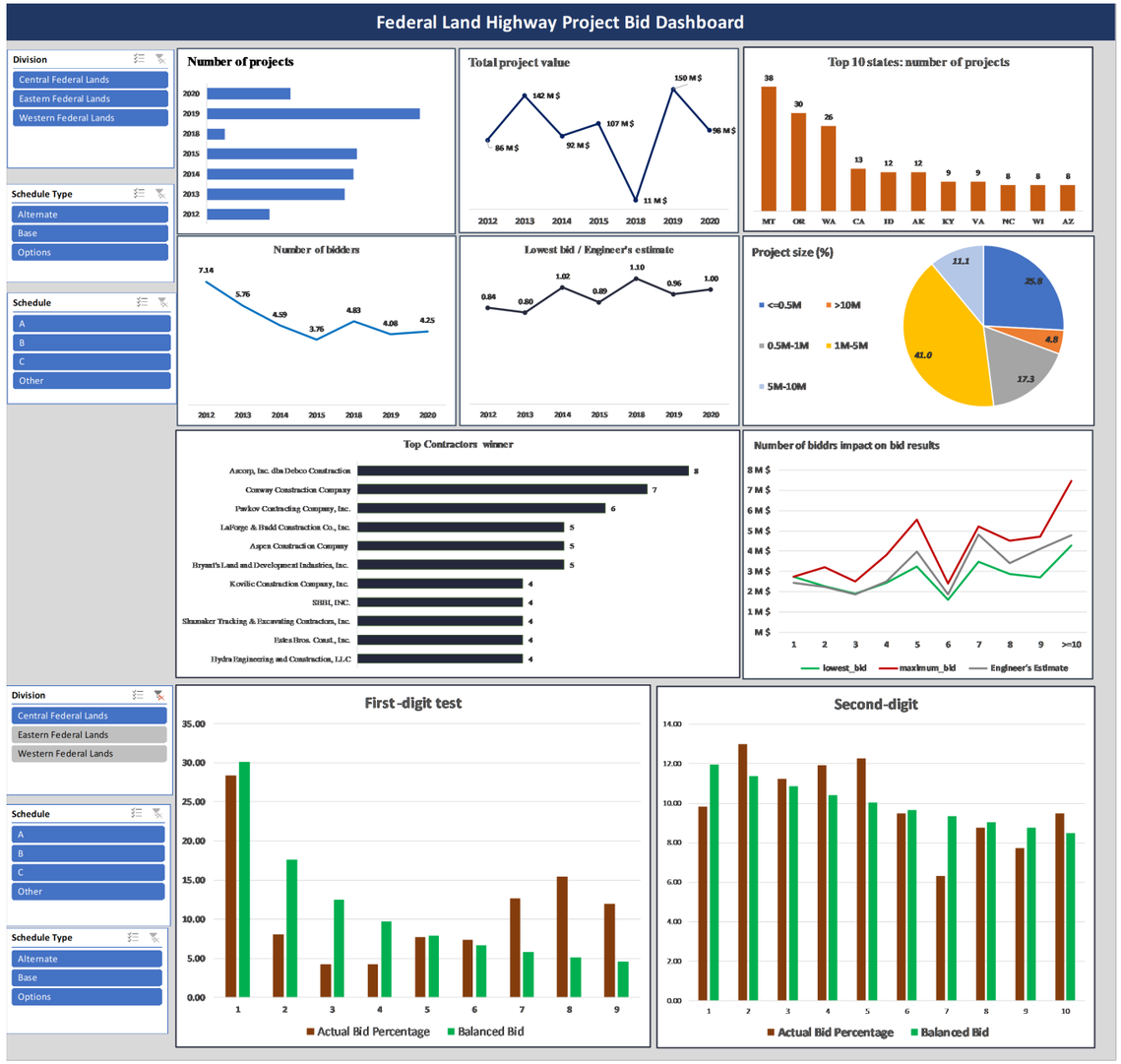




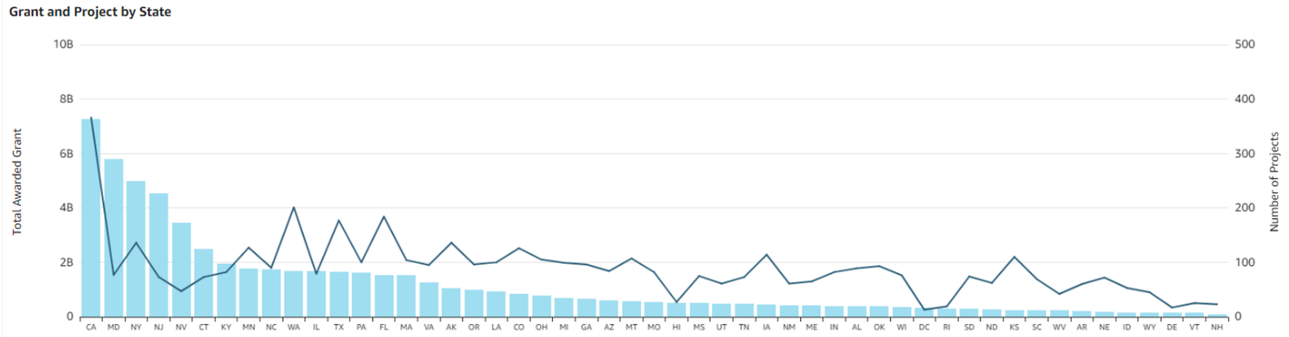


Dashboard and Benchmarking





Total Awarded Grants	Total Awarded Projects	Total Awarded Organizations	Total Awarded Local and Tribal Governments
$59.68B	4,542	2,995	1,058



Combating Unethical Practice

NEWS COLUMNS ANTITRUST CHRONICLE CPI MEDIA

US: AG now takes the lead in state's asphalt antitrust lawsuit

CPI - January 11, 2017

The state Attorney General's Office is now taking the lead in an antitrust lawsuit claiming the price of asphalt in West Virginia markets has been inflated through noncompetitive practices.

In October, the private firm Bailey & Glasser filed lawsuits on behalf of Charleston, Parkersburg, Beckley and Bluefield. A couple of days later, the state Department of Transportation jumped on board the lawsuit.

The matter arose a few times during Patrick Morrisey's re-election campaign for Attorney General. A legislative committee overseeing the state Department of Transportation has aggressively asked leaders in the Department of Transportation how and why the lawsuit was filed without the Attorney General. And lawyers for the asphalt companies filed a motion to dismiss, saying the only way the state could legally file a claim is through the Attorney General.

Throughout, Morrisey and his office had said there was nothing that could be said publicly while the matter was under investigation.

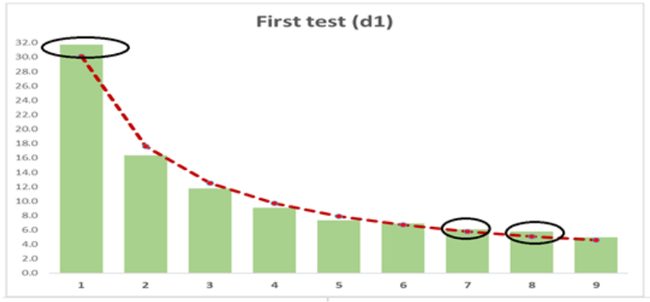
WV annual TAB model results.

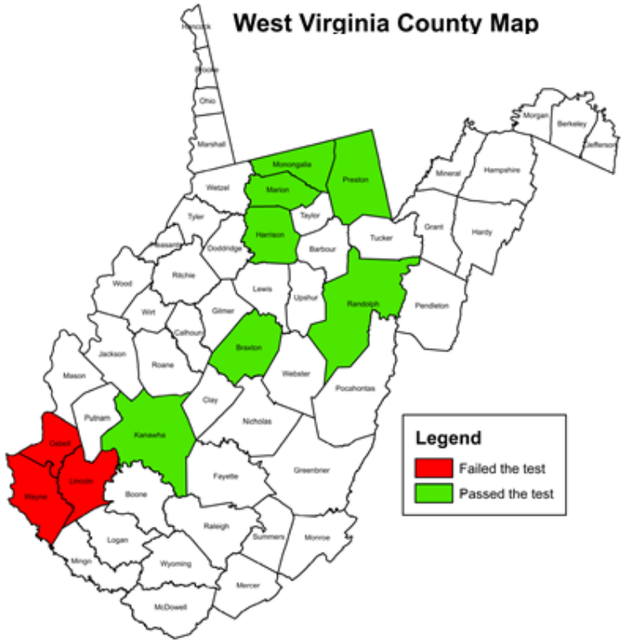
year	Number of records	Test 1 (d1)	Test 2 (d2)	Test 3 (d1d2)	Framework result
2011	1403	1.56	0.87	1.84*	Retain
2012	2030	1.74*	1.58	2.21**	Reject
2013	1543	1.90*	0.90	2.38**	Reject
2014	1874	2.04**	1.30	2.56**	Reject
2015	1661	1.66	1.41	2.09**	Reject
2016	1717	1.62	1.49	2.36**	Reject
2017	2106	2.61**	1.51	2.89**	Reject
2018	2057	1.99*	2.11**	2.43**	Reject
2019	2169	2.21**	1.41	2.63**	Reject
2020	2039	1.21	1.46	2.09**	Reject

WV company TAB model results.

Company	Number of records	Test 1 (d1)	Test 2 (d2)	Test 3 (d1d2)	Framework result
West Virginia Paving, Inc.	2556	3.19**	1.03	3.39**	Reject
J. F. Allen company	1325	1.25	0.53	1.56	Retain
Mountaineer Contractors, Inc.	1049	0.49	0.73	1.21	Retain
Marshal Asphalt Wear CRSE, SG, TY I	1810	2.83**	1.46	3.09**	Reject
Marshal Asphalt Base CRSE, SG, TY I	1630	5.02**	1.47	5.42**	Reject
Marshal Asphalt Base CRSE, SG, TY II	1756	1.94*	0.81	2.88**	Reject

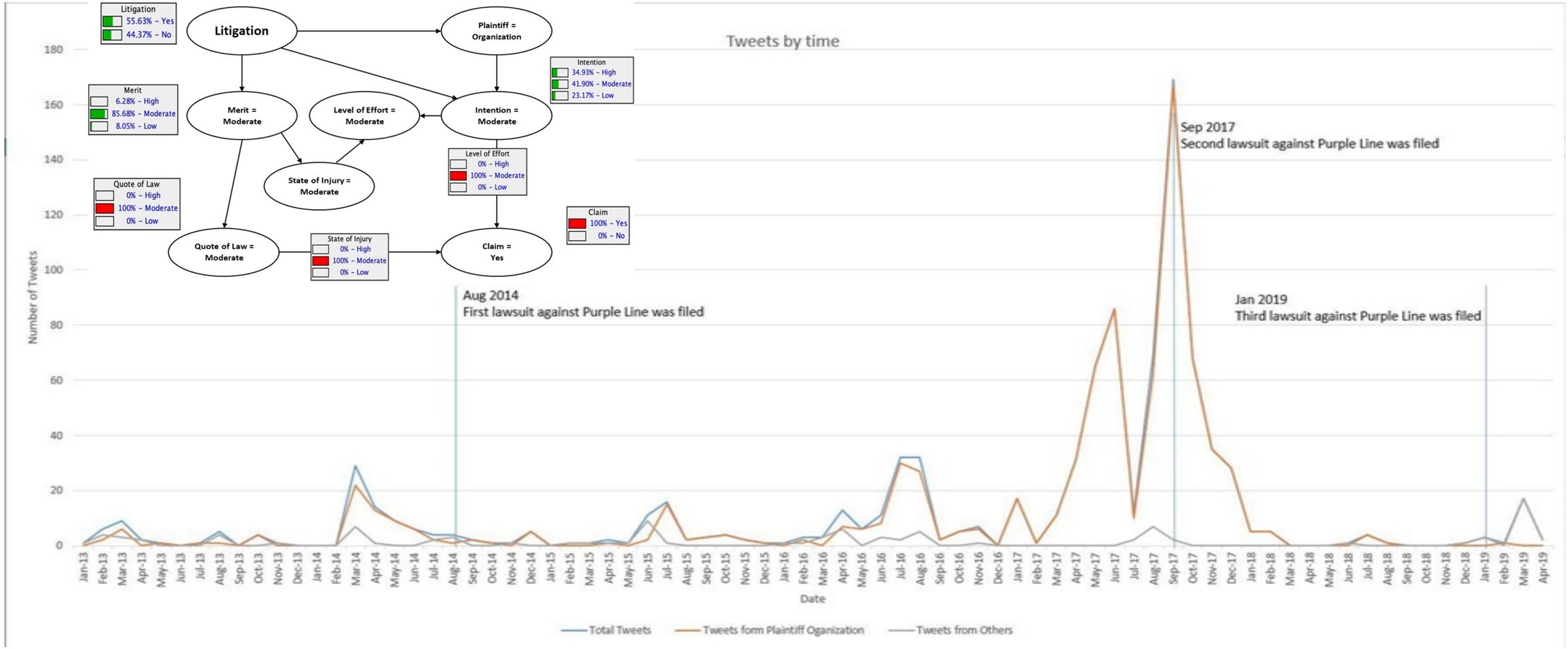
Note: ** and * denote a significant difference from Benford's law distribution at 1% and 5% levels of significance, respectively. Critical values of Kuiper test are 2.001 and 1.747 regardless of sample size, respectively (Louter 1970).





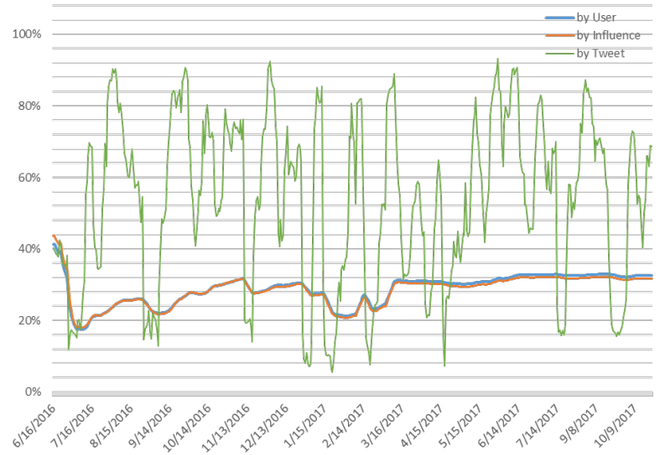
https://github.com/benfordlaw/Irregular-Bidding-detection

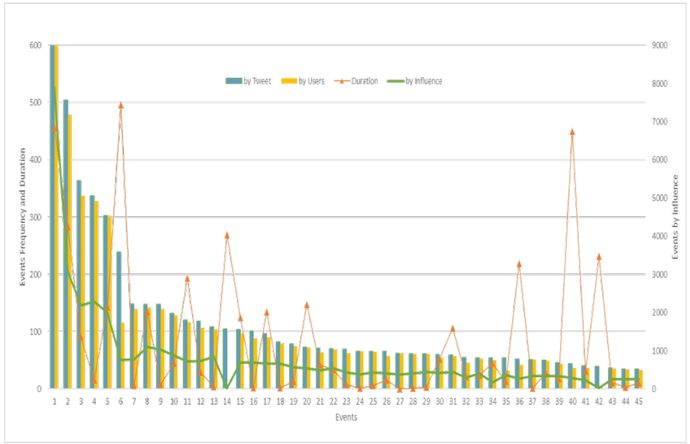
Project Litigation Intelligence

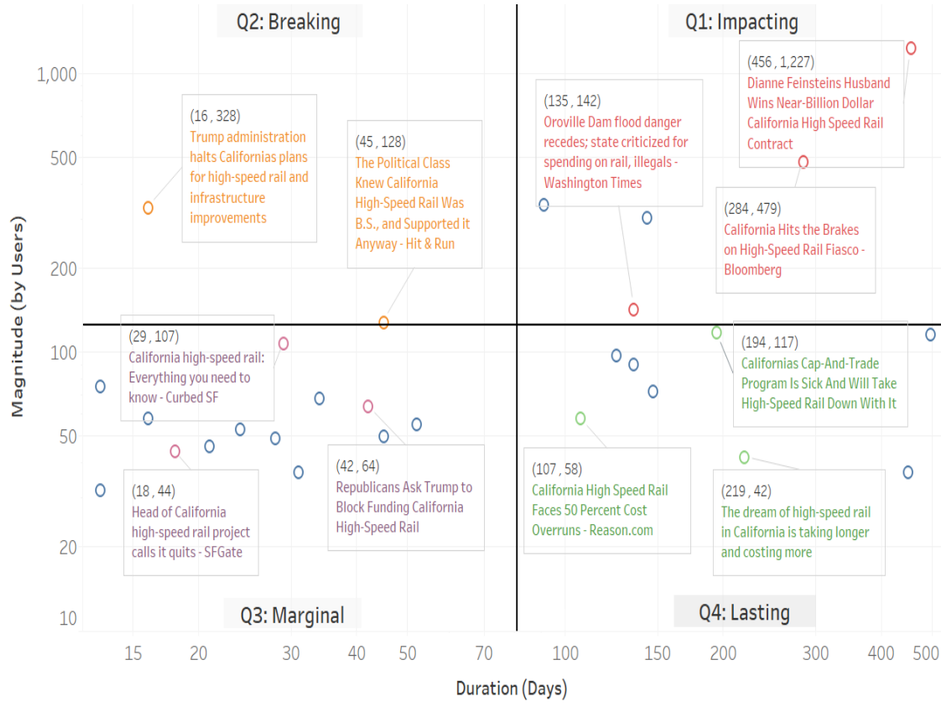


Understanding Public Perception

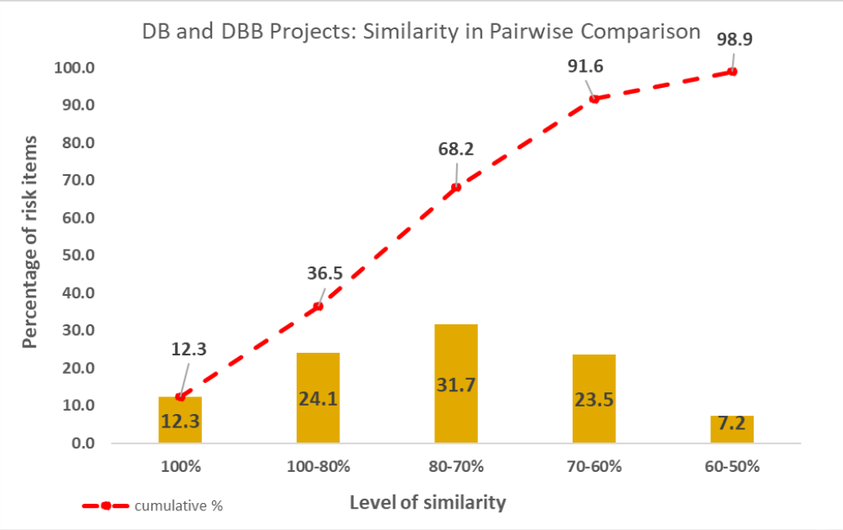
- Public Acceptance is driven by social media events
- Event Influence
  - Breaking
  - Impacting
  - Lasting
  - Marginal
- User Analysis
  - Opinion Leaders
  - Opinion Followers

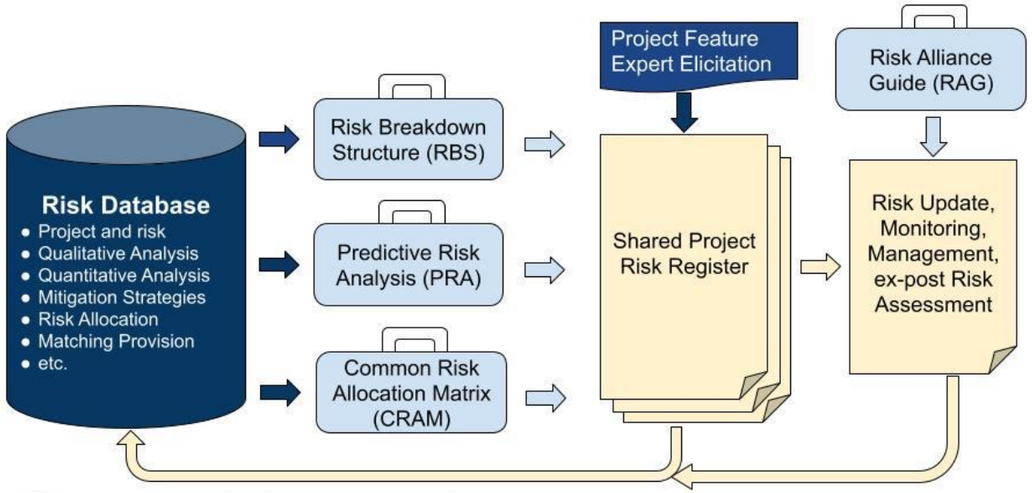






Data Centric Risk Assessment

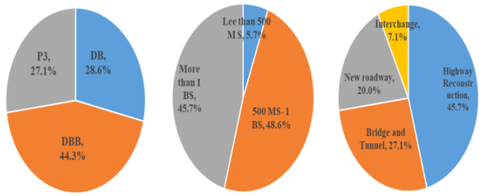




Data Entry Form

Project Name	
Location	
Type	All
Delivery Metho	All
Size	All
Number of Risks	10 20 30
Selection Criteri	Frequency

Reset | Generate Risk Template



Data Profile

Initial Risk Template			Number of projects		70
#	Type	Risk	Prevalence	Probability	Cost (M$)
1	right of way	additional right of way required	100.00	0.41	4.55
2	utilities	Utility Cost, Cooperation from utility owners for both r	91.43	0.49	13.69
3	structures an	hazardous materials	87.14	0.39	3.43
4	construction	construction impacts due to lack of right of way and tin	78.57	0.41	0.87
5	right of way	right of way Cost	60.00	0.43	1.08
6	structures an	drainage	60.00	0.37	5.32
7	right of way	different site conditions	57.14	0.51	4.55
8	construction	delay in obtaining permits and approvals	51.43	0.69	0.06
9	utilities	utility relocation may not happen in time	50.00	0.43	6.01
10	construction	Operations and maintenance during construction	48.57	0.29	8.34
11	managemen	schedule expectations and constraints	42.86	0.58	12.70
12	design	design exceptions approvals	41.43	0.55	17.81
13	structures an	contractor access, staging coordination and constructab	40.00	0.75	0.63
14	utilities	difficult utility relocations and conflicts	40.00	0.78	-0.30
15	partnerships	local agency coordination	38.57	0.67	15.05
16	environmen	environmental mitigation costs	35.71	0.61	0.89
17	environmen	environmental permits	35.71	0.53	0.11
18	contracting a	Contract delay, change orders	35.71	0.41	4.45
19	construction	accurate and timely traveler information during/after c	35.71		
20	construction	Surface condition impacts construction cost and schedu	34.29	0.34	2.69
21	managemen	higher real estate cost due to development, annexatio	34.29	0.13	0.60
22	right of way	right of way acquisition	31.43	0.10	3.63
23	managemen	market condition	31.43	0.30	10.88
24	construction	discovery of unknown utilities during construction	30.00	0.47	0.51
25	partnerships	delay in agreements	30.00	0.51	0.25
26	managemen	traffic management	28.57	0.44	0.31
27	managemen	construction funding shortfall	28.57	0.29	16.08
28	environmen	noise mitigation	28.57	0.42	26.25
29	structures an	additional mitigation for section 4(f) impacts	27.14	0.30	13.50
30	design	changes in design standards and requirements	27.14	0.31	0.21

Risk Register: An Example

- General Risk
  - Market and Inflation
  - Labor Shortage
- Task-oriented
  - Delay in NEPA Approval
  - Financing
- Linked to WBS
  - Construction
  - Design
  - Environmental
  - Utilities
  - Structure and Geotech

APPENDIX B: RISK REGISTER

FDOT District 5 Wekiva Parkway 2014 Risk Assessment Update Risk Register										Initial Risk Quantification							
Risk Information											Cost Risk Information (Millions $)			Schedule Risk Information (Months)			
Record #	Risk ID	Status	Functional Assignment	Risk Name	Risk Description	Modeling Notes	Cost Threat/ Opportunity	Schedule Threat/ Opportunity	Prob. (%)		Low Cost	Most Likely Cost	High Cost	Low Schedule	Most Likely Schedule	High Schedule
-	MGT 40.03-2A2B2C	Dormant	Management / Funding	Schedule delay from not obtaining TIFIA	For 2A-C, pursuing TIFIA for the sections with combined 1A1B-base schedule assumes TIFIA and without it there may be delay to construction start of 18 months for Segments 2. Schedule for 1A1B will not change. **Base scenario will assume TIFIA Funding. Can run separate non-TIFIA schedule scenario if required.**	Correlate with risk on material availability for 2		Threat	50%				18	18	18
2	CTR 40.01-1A	Active	Contracting and Procurement	Market conditions – 1A	This is related to shortage of skilled labor, materials, equipment due to demand from other projects in the area. I-4 will have started and other Wekiva sections. Structural steel, Buy-American and reimbursable utilities (with Duke Transmission for some segments); escalation on materials for conditions six months from now. Quantification is 5 percent of 40 percent of construction cost to account for premium on materials with 50 percent probability.	Construction	Threat		50%	$0.825	$1.100	$1.375			
3	CTR 40.01-1B	Active	Contracting and Procurement	Market conditions – 1B	This is related to shortage of skilled labor, materials, equipment due to demand from other projects in the area. I-4 will have started and other Wekiva sections. Structural steel, Buy-American and reimbursable utilities (with Duke Transmission for some segments); escalation on materials for conditions six months from now. Quantification is 5 percent of 40 percent of construction cost to account for premium on materials with 50 percent probability.	Construction	Threat		50%	$0.825	$1.100	$1.375			

Data Changes Project Risk Management

Transforming Status Quo

- How trustworthy are experts in Risk Assessment?

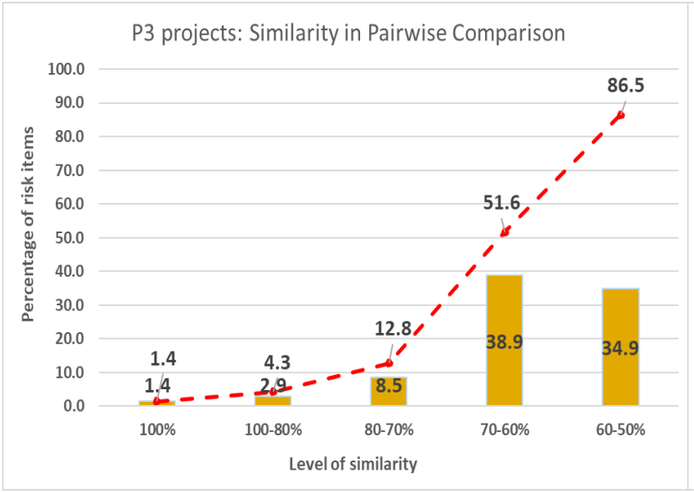
1. Are risks unique?
2. How did experts perform?
3. What is the Value of Using Historical Data to Predict Project Risks?

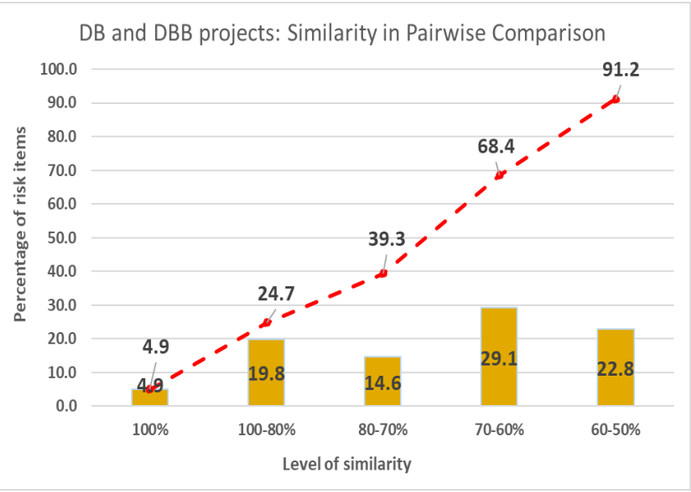
Issue	Challenge	Solution	Method
Incompatibility	Knowledge sharing across agencies	Risk Breakdown Structure	Natural Language Processing
Interdependency	Under/overestimate risk consequence	Predictive and Generative Modeling	Big Data Analysis
Performance	Lessen learned & best practice	Ex-post Analysis	Risk Lifecycle Modeling
Transparency	Stakeholder collaboration	Risk Database	Collective Intelligence

Risk Uniqueness

- Research Design
  - Similarity at Document level
  - Similarity at Risk Item level
  - Similarity at Risk Evaluation level
- Data: Risk Registers
  - 70 Major Transportation Projects
  - 6000+ Risk Items
  - Structural/Unstructured Data Format
- Research Methodology: Natural Language Processing
  - TF-IDF
  - Word2vec

- **82 %** of words and phrases are similar in risk registers
- Pairwise comparison of risk registers shows **53%** of risks are similar

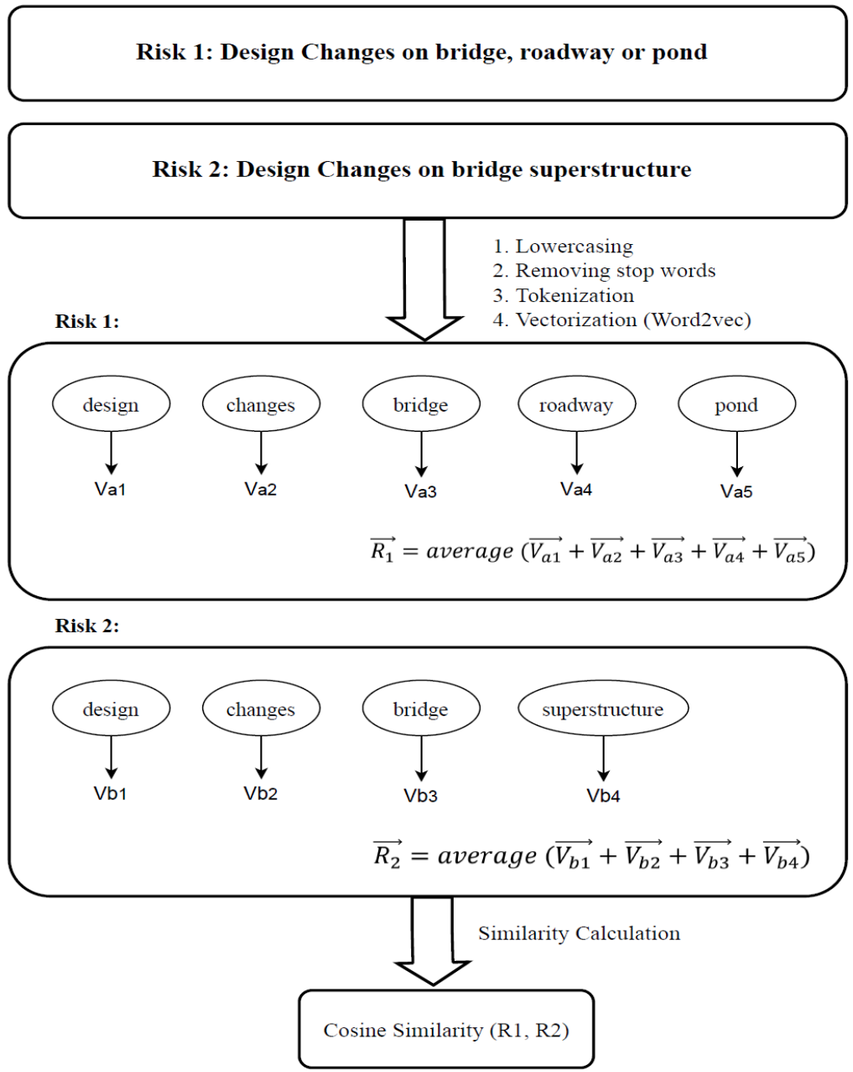




Source: Erfani, A., Cui, Q., & Cavanaugh, I. (2021). An Empirical Analysis of Risk Similarity among Major Transportation Projects Using Natural Language Processing. Journal of Construction Engineering and Management, 147(12)

Risk Item Similarity

Risk (1)	Similarity	Risk (2)
Owner directed changes and design views	1	Owner directed changes and design views
Determination of secondary impacts to wetlands	0.95	Determination of wetlands impacts
Utility Relocations	0.85	Utility Relocations and conflicts
Federal agencies may take longer than expected to review and issue a permit	0.81	Permits or agency actions delayed or take longer than expected
Handling of Contaminated Materials	0.78	Unanticipated Hazardous Materials or Contaminated Soils
Unforeseen Utilities	0.72	Unknown Utilities
Right of way Acquisition Cost	0.64	Right of Way Acquisition is costlier than expected
Construction Change Orders & Incentives	0.62	Change orders
Concrete delivery	0.59	Materials delivery constraints
Disposal of Regulated Material	0.56	Unidentified Hazardous materials Found in Construction



Semantic similarity using NLP

Risk Evaluation Uniqueness

- Similarity at Risk Evaluation level
- Comparing the assessment of similar risks with different terminologies show a large similarity across their assessment in (1-5) Likert scale

$$Similarity = 1 - Distance\ index = [1 - \left(\frac{|x_1 - x_2|}{4}\right)] * 100$$

- A Good Database might cover more than 70% of risks in a risk register

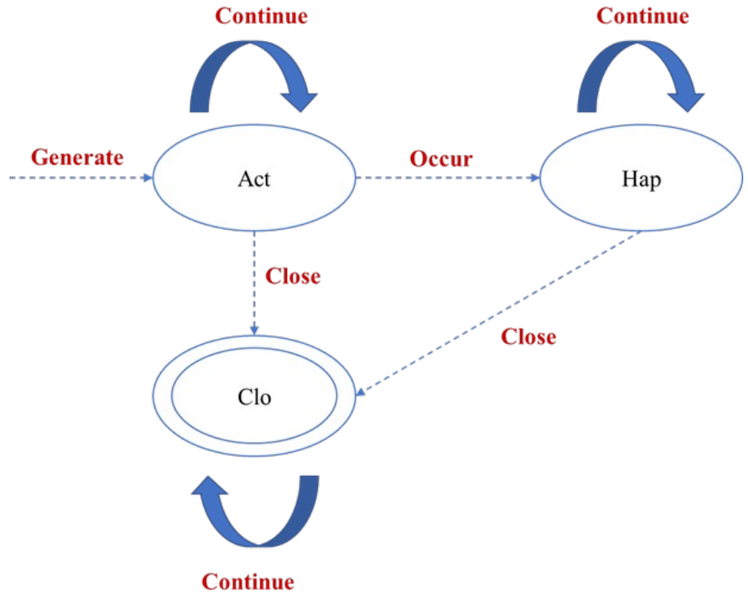
- Risk consequences and probabilities were evaluated in similar patterns
- Cost impact shows very common evaluation

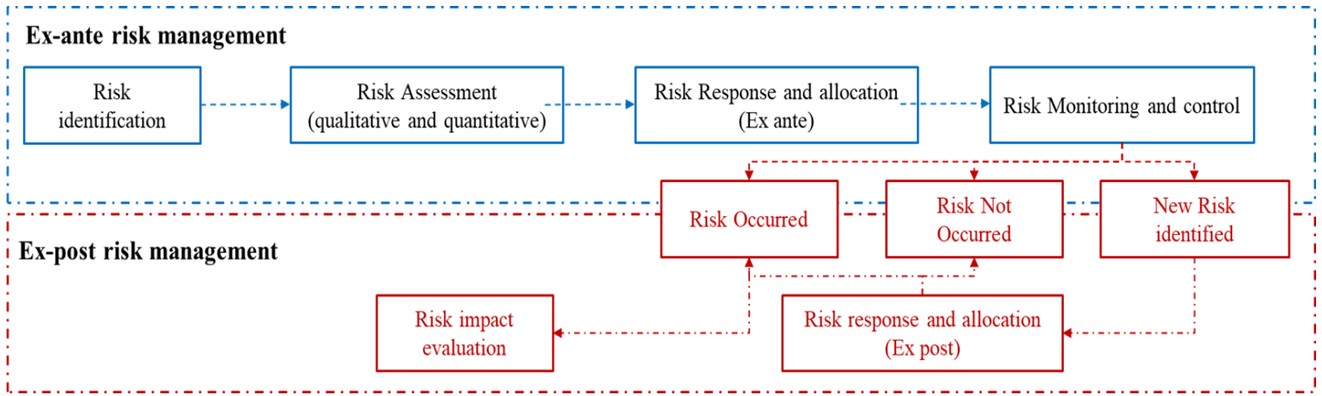
Cosine similarity level	Probability	Cost	Schedule	Probability + Cost	Probability +Schedule
Group A: DBB/DB Projects					
At least 0.5	58.9%	96.6%	60.6%	72%	41%
At least 0.7	68.2%	100%	75%	83%	50%
At least 0.8	73.7%	100%	85.7%	70%	57%
Group B: P3 Projects					
At least 0.5	62.0%	72.0%	66.5%	71%	55%
At least 0.7	63.7%	93.1%	73.9%	81%	57%
At least 0.8	60.3%	96.7%	65.6%	71%	51%

Source: Erfani, A., Cui, Q., & Cavanaugh, I. (2021). An Empirical Analysis of Risk Similarity among Major Transportation Projects Using Natural Language Processing. Journal of Construction Engineering and Management, 147(12)

Ex-Post Risk Assessment

- **Research Design**
  - Risk Life Cycle Automata
  - Risk State: Active (Act); Happening (Hap); Closed (Clo)
  - Risk Transition: Generate; Occur; Continue; Close
- **Data: Registers and Annual Updates**
  - 11 Major Transportation Projects
  - Total $15.6 Billion ($583M - $4922M)
- **Risk Methodology**
  - Performance Indicator Definition
  - Performance Analysis



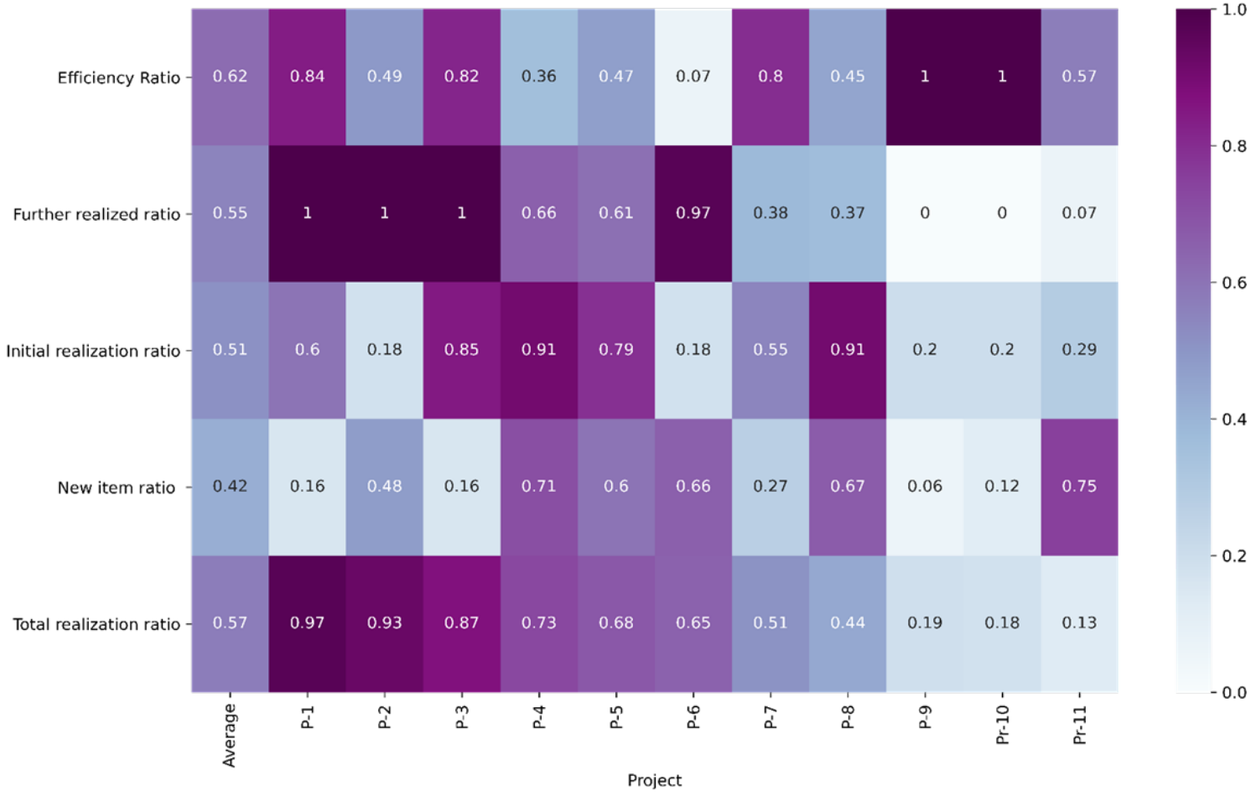


Source: Erfani, A., Ma, Z., Cui, Q., & Baecher, G. (2023). Ex post Project Risk Assessment: Theory, Method, and Empirical Study. Journal of Construction Engineering and Management, 149(2).

Risk life-cycle examples

	Risk Register Year(1)	Risk Register Year(2)	Risk Register Year(3)	Risk Register Year(4)	Corresponding Language
Risk A	Risk is identified	Risk is active not happened yet	Risk start happening	Risk is Closed (Realized)	Act, Act, Hap, Clo
	Generate → Act → Continue → Act → Occur → Hap → Close → Clo				
Risk B	----------	Risk is identified	Risk is active not happened yet	Risk is Closed (Dismissed)	Ø , Act, Act, Clo
	Inactive → Generate → Act → Continue → Act → Close → Clo				
Risk C	Risk is identified	Risk is active not happened yet	Risk is Closed (Dismissed)		Act, Act, Clo, Clo
	Generate → Act → Continue → Act → Close → Clo → Continue → Clo				

Ex-Post Risk Assessment



Overall performance metrics

$$\text{Total realization ratio} = \frac{\text{Number of } \mathbf{realized} \text{ risks}}{\text{Number of } \mathbf{identified} \text{ risks}}$$

$$\text{Total dismissed ratio} = \frac{\text{Number of } \mathbf{dismissed} \text{ risks}}{\text{Number of } \mathbf{identified} \text{ risks}}$$

$$\text{Initial realization ratio} = \frac{\text{Number of } \mathbf{realized} \text{ risks in year 1}}{\text{Number of } \mathbf{identified} \text{ risks in year 1}}$$

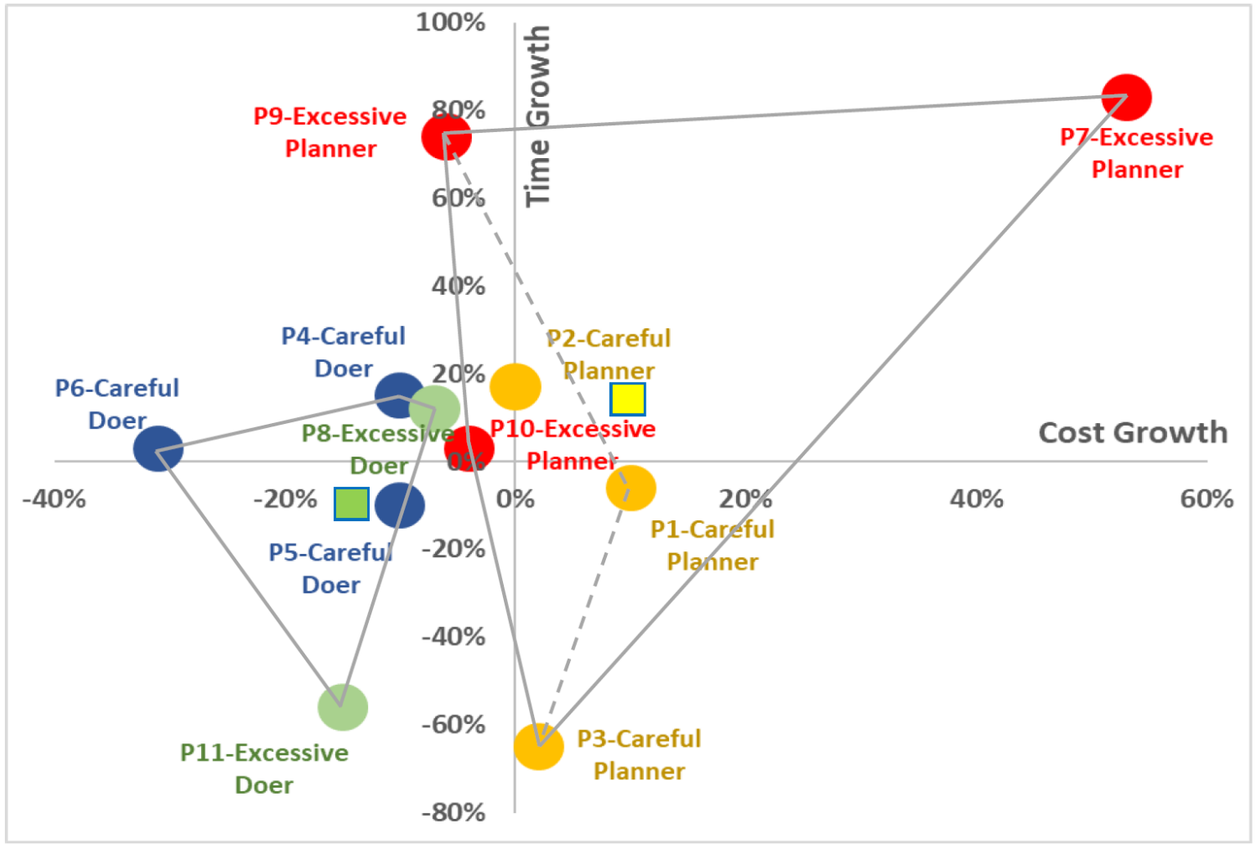
$$\text{Initial efficiency ratio} = \frac{\text{Number of } \mathbf{realized} \text{ risks in year 1}}{\text{Number of } \mathbf{realized} \text{ risks}}$$

$$\text{New item ratio} = \frac{\text{Number of } \mathbf{identified} \text{ risks after year 1}}{\text{Total Number of } \mathbf{identified} \text{ risks}}$$

$$\text{Further realized ratio} = \frac{\text{Number of } \mathbf{realized} \text{ risks from risks after year 1}}{\text{Number of } \mathbf{identified} \text{ risks after year 1}}$$

- 57% of identified risks are realized (ranging from 0.13 to 0.97)
- 62% of identified risks at the first step are finally realized
- 55% of those added risks are happened finally

Risk Management Style



Careful Planner identifies most risks at the initial phase and most of these identified risks ultimately occur with consequences;

Excessive Planner identifies the majority of risks at the initial phase but most of these identified risks are ultimately closed without occurrence;

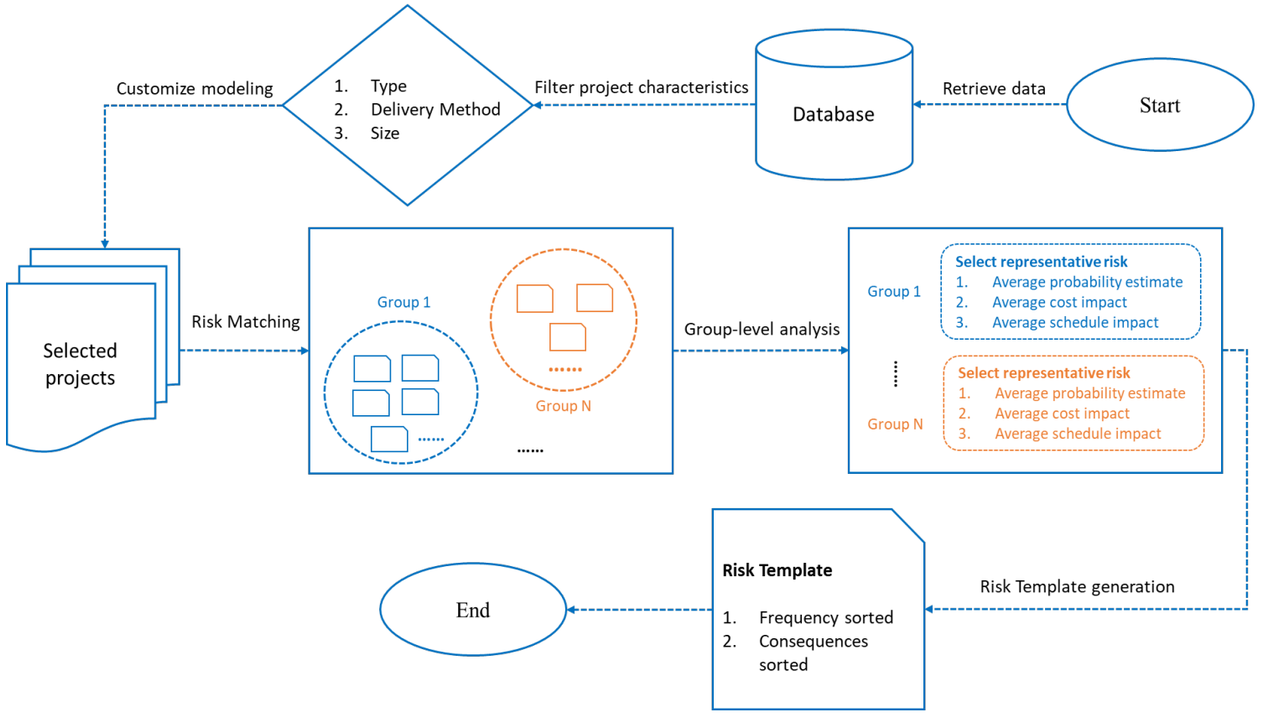
Careful Doer actively updates risks during project implementation and most of these identified risks ultimately occur with consequences;

Excessive Doer actively updates risks during project implementation and most identified risks are closed without occurrence.

Source: Erfani, A., Ma, Z., Cui, Q., & Baecher, G. (2023). Ex post Project Risk Assessment: Theory, Method, and Empirical Study. Journal of Construction Engineering and Management, 149(2)

Value of Risk Data

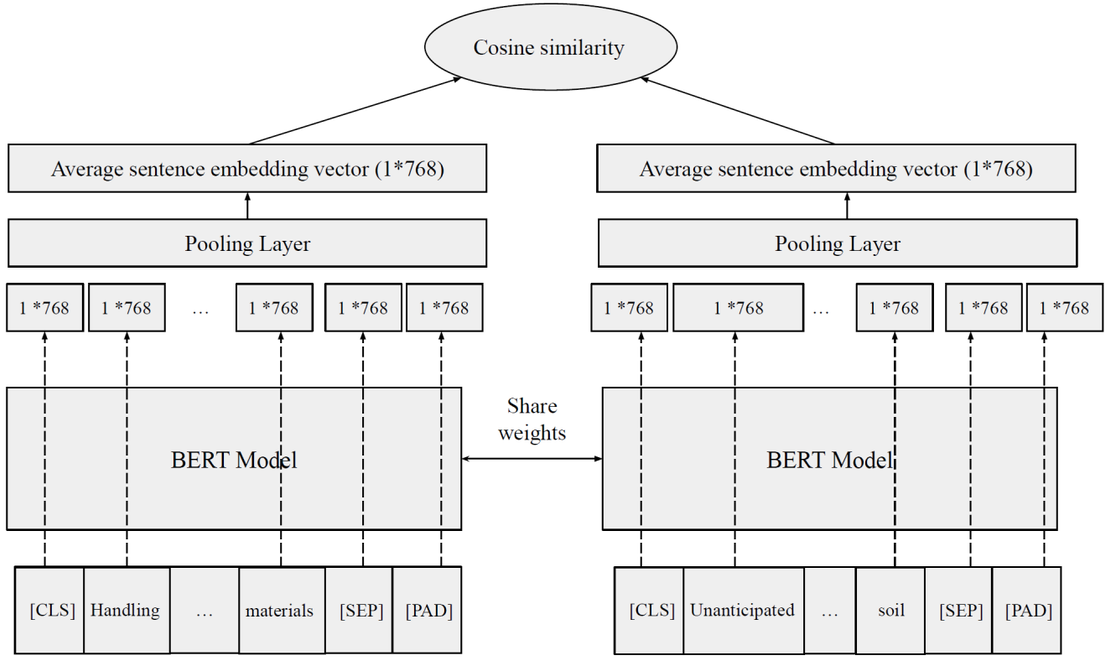
- **Predictive Risk Modeling**
  - Step 1: Customize modeling: User define project characteristics
  - Step 2: Risk matching : Group similar risks with different languages
  - Step 3: Group-level analysis: Finalize the grouped risks as one risk item
  - Step 4: Risk register generation: Sort the risks based on frequency or consequences
  - 66% of risk items in five random testing projects are covered in predictive risk model



Project ID	Recall	Precision	F_1-score
A	45.7%	45.7%	45.7%
B	55.3%	52.5%	53.8%
C	82.0%	95.3%	88.2%
D	92.3 %	40.0%	55.8%
E	71.4 %	16.7%	27.0%
Overall	66.4%	53.4%	59.2%

Source: Erfani, A., & Cui, Q. (2022). Predictive risk modeling for major transportation projects using historical data. Automation in Construction, 139, 104301

Value of Risk Data



Risk 1: Handling of contaminated materials

Risk 2: Unanticipated hazardous material or contaminated soil

Table IV. Risk co-occurrence out of 70 risk registers

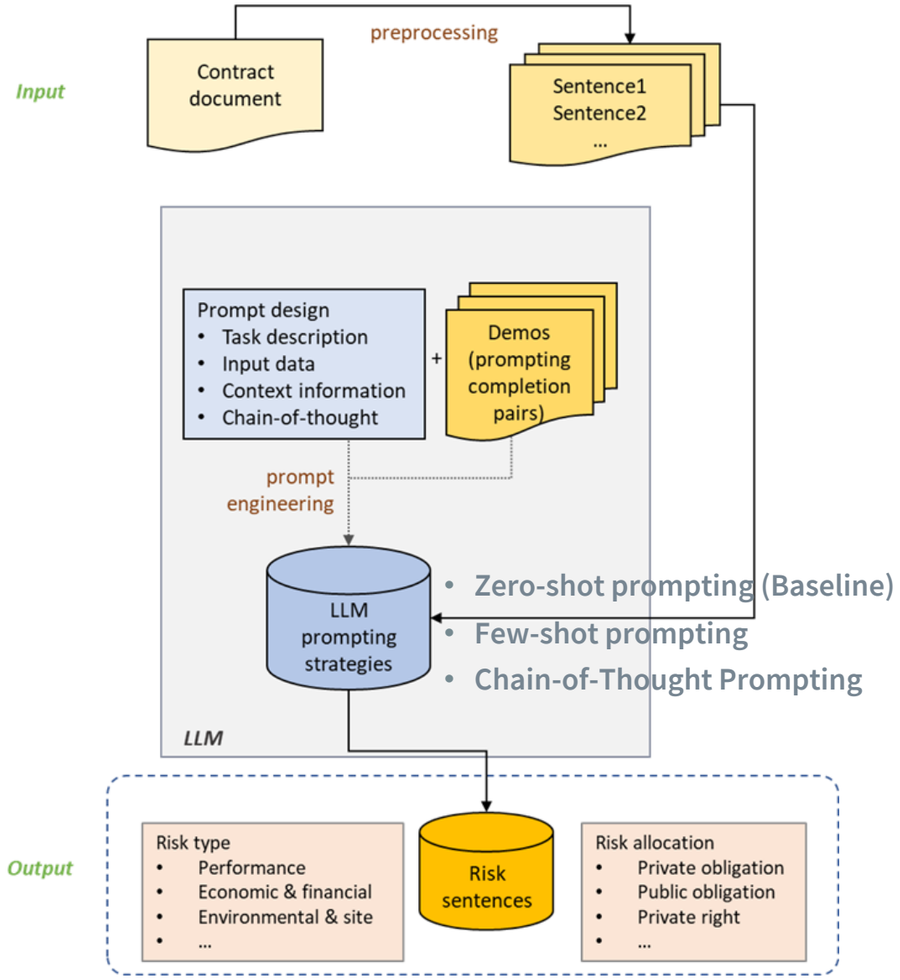
Risk (1)	Risk (2)	Rate
Right of way plan	Utility relocation	40
Delay in procurement	Utility relocation	36
Contractor Access	Utility relocation	36
Right of way plan	Delay in procurement	34
Design changes	Utility relocation	33
Contractor Access	Right of way plan	33
Contractor Access	Different site and subsurface condition	32
Contractor Access	Delay in procurement	32
Hazardous Materials	Utility relocation	32
Right of way acquisition issues	Utility relocation	31
Design changes	Right of way plan	31
Contractor Access	Hazardous Materials	30
Contractor Access	Design changes	30
Contractor Access	Construction Excavation	30
Hazardous Materials	Right of way plan	30

Unified Risk Breakdown Structure

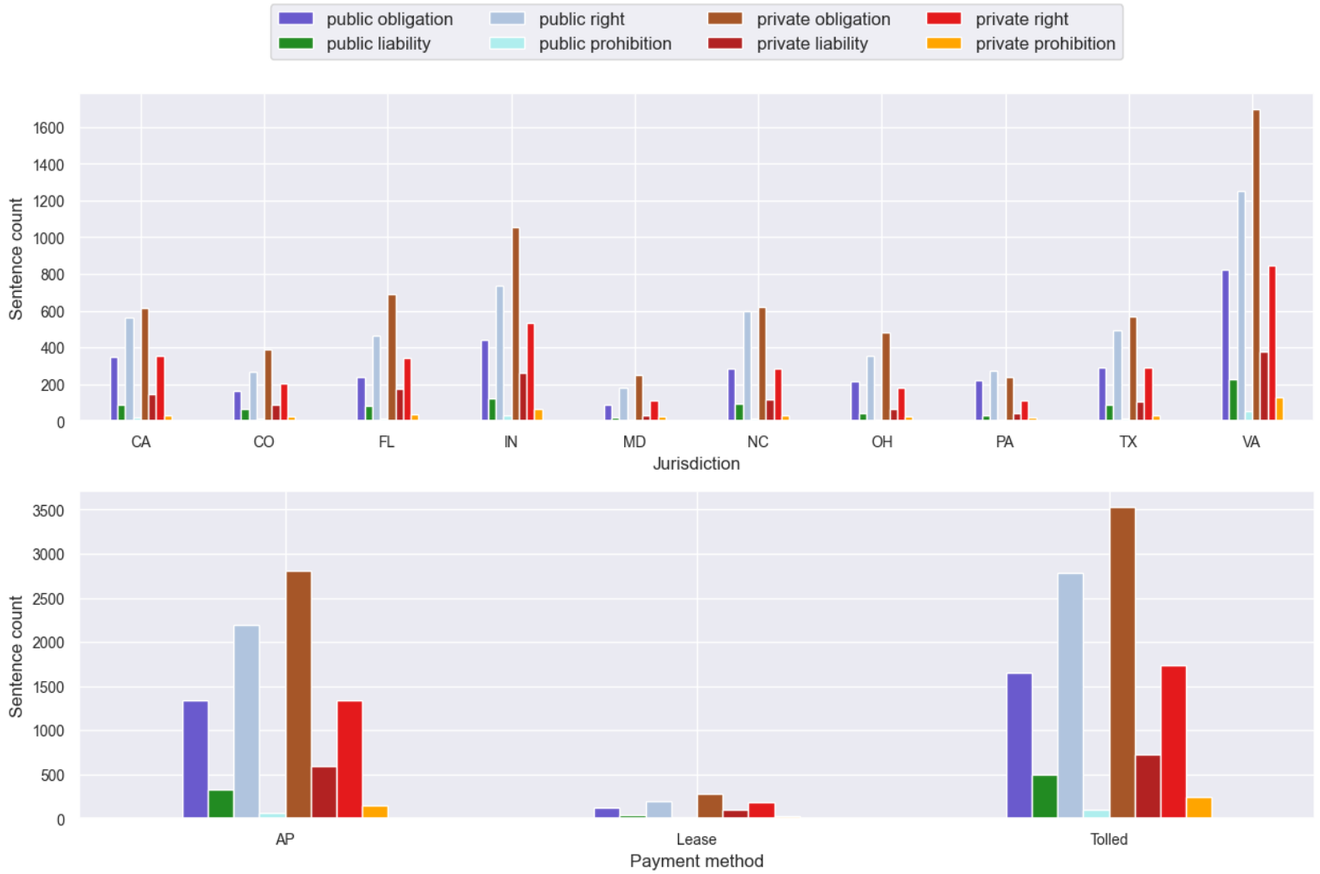
Co-occurrence based Risk Interdependency

Erfani, A., Cui, Q*, Baecher, G, and Kwak, Y (2023). Data-driven approach to risk identification: A common risk breakdown structure. IEEE Transactions on Engineering Management. In print

LLM for Contract Risk Management



- **Distribution of risk allocation sentences**



BUILD UPON SUCCESSFUL PARTNERSHIP

Qingbin Cui, Ph.D.
cui@umd.edu
https://bac.umd.edu/