SYM-AM-21-108

EXCERPT FROM THE PROCEEDINGS OF THE EIGHTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

Design and Development of Data-Driven Risk Assessment Through the Integration of Federal Acquisition Data with Open, Internet Sources

May 11–13, 2021

Published: May 10, 2021

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Design and Development of Data-Driven Risk Assessment Through the Integration of Federal Acquisition Data with Open, Internet Sources

Ningning Wu—is Professor of Information Science at the University of Arkansas at Little Rock. She received a BS and an MS degree in electrical engineering from the University of Science and Technology of China and a PhD in information technology from George Mason University. Wu's research interests are data mining, network and information security, and information quality. She holds certificates of the IAIDQ Information Quality Certified Professional (IQCP) and the SANS GIAC Security Essentials Certified Professional. [nxwu@ualr.edu]

M. Eduard Tudoreanu—is Professor of Information Science at the University of Arkansas at Little Rock. Tudoreanu has expertise in human–computer interaction, information quality, advanced visualization of complex data, and virtual reality. He worked on visual data analysis and has extensive experience in software development and user interface design. Professor Tudoreanu was the founding Technical Director of the Emerging Analytics Center. He has been the keynote speaker at ABSEL 2010 and served as a panelist for the National Science Foundation and Missouri EPSCoR. He earned his Doctor of Science degree in computer science in 2002 from Washington University in St. Louis. [metudoreanu@ualr.edu]

Sowjanya Manasa Sistla—is an enthusiastic graduate student who always has a zeal for learning. She earned a bachelor's degree in computer science. She worked at Accenture, where she started learning to analyze data and learned how deeply it is having an impact in the current world of technology, where almost everything is becoming dependent on data. Her work on researching data, finding credible sources for the data, and analyzing data to give estimates on the given input has been an amazing experience and a great opportunity to learn data analytical skills. [ssistla@ualr.edu]

Abstract

In acquisition planning and management, all information needs to be considered to ensure the successful execution of a project. Key pieces of information, with high data quality, must be used for accurate risk management that requires a good understanding and quick identification of the potential factors that might affect a program. This paper investigates effective data science techniques for a quick identification of high risks posed by natural disasters to a supply/production chain. The paper proposes a tool for helping a project manager analyze the integrity of a supply chain and find alternative potential contractors in lower risk areas. We focus on assessing the risks related to natural disasters at a contractor's location, business type, and business relationships and networks. The system described here seeks to first identify critical contractors based on the natural disaster risk of the places they perform their main activity for a contract and the uniqueness of their business types. Second, we design a risk assessment framework and a visualization framework that enable an acquisition expert to assess a project's risk by identifying contractors that may become critical and require additional redundancy. The visualization tool also can be used to determine additional potential subcontractors that are located in lower risks areas of the country.

Key words: visualization dashboard, data analytics, risk assessment, natural disasters

Introduction

The goal of this work is to assess and assign the risk level of a contractor based on its location and business properties. We built a tool to help identify alternative contractors for the high-risk ones. In addition to designing a risk assessment framework, our work relies on visualization tools to enable an acquisition expert to assess a program's risk by identifying the potential of failure for various contractors. Our approach also allows an expert to compare multiple acquisition programs to identify whether the same contractors or the same high-risk

geographic areas occur in more than one project, thus becoming a single potential point of failure or delay for multiple programs.

Assessing the natural disaster risk of an area is a challenging task because multiple metrics and dimensions must be considered for a comprehensive picture. National Centers for Environmental Information (NCEI) has been collecting natural disaster data for each U.S. county since 1950. The data cover a wide range of natural disasters, including flood, tornado, hurricane, blizzard, high wind, flash flood, hail, dust storm, and so on. However, there is no standard way to categorize a disaster by its intensity and damage level. Even though some disaster types, such as tornadoes and hurricanes, do have a categorization system for their intensity, it is often difficult to assess an incident's impact to the local communities without other supporting information. Furthermore, it is hard to compare the impacts of different disaster types. One challenge of this study is to identify appropriate techniques to categorize the severity levels of natural disasters. New data sources will be researched to complement the NECI data for a better risk categorization system. One possible source is the Federal Emergency Management Agency (FEMA), which can provide information about the recovery efforts after a natural disaster strikes.

Critical contractors are those who provide unique products and services or have some special characteristics so that if they failed, it would be hard to find alternatives to fill their roles. Critical contractors could be a weak link in a supply chain, and their early identification will help the project manager control the risks effectively. The study will leverage the North American Industry Classification System (NAICS) code along with other business information for identifying the critical contractors.

NAICS is the standard used by federal agencies to describe the business specialization of a company (Office of Management and Budget, 2017). A NAICS code can be attached to many products and many companies, and so it can be used to identify potential alternative contractors for a program or a collection of programs. Because there might be many companies with the same NAICS code, to effectively identify critical contractors, acquisition experts may need to find more business-related information using the Internet or a data science–based tool such as the business risk indicators by Gill et al. (2019) or the Internet text processing framework by Wu et al. (2018).

We considered two case studies to demonstrate our framework. The first one focused on https://www.fpds.gov/ and extracted places of performance for various types of business used by the Navy over a period of 10 years. The second case study focused on awards with a high number of contractors in high-risk areas. These data were extracted from historical records from a data set we built based on https://www.usaspending.gov/. Based on the natural disaster risk assessment framework we developed, each U.S. county is assigned a risk level of high, medium, or low, depending on the historical NCEI weather data of the area as well as the FEMA assistance programs that had been utilized in the past. Then, the database was searched for contracts with a high percentage of contractors located in the high-risk areas. Contracts with a high concentration of risky contractors are prone to fail and thus require an effective risk management plan. One approach to reducing the overall risk of a contract is to replace high-risk contractors with low-risk ones.

The technique we present here allows an expert to find projects with a high number of contractors in high-risk areas. Once that has been accomplished, we provide two mechanisms for an expert to find possible alternative contractors. One is to find geographical areas with low risk of natural disasters and where there is a concentration of industries that could provide products and services for a given project. The second is to find actual contractors that have

been involved in federal awards and are in areas with low disaster risks, regardless of the concentration of related industries.

Related Work

Previously, policy-makers and researchers have recognized the need to employ data as a multifaceted means of increasing the agility of the acquisition process (Krzysko & Baney, 2012). To this end, research has looked at automatic means of dealing with the heterogeneous acquisition data sources from text processing (Zhao et al., 2015), systems engineering (Cilli et al., 2015), and business (Gaither, 2014) perspectives. Our paper is different both in content and in approach—in content in that we rely on big data to identify hidden risk factors, and in the approach in that our expertise in information visualization, data quality, data governance and policy (chief data officers), and in data science provides a value-based perspective.

Gill et al. (2019) took a step further and employed data science techniques to determine a model that would assist an expert in source-selection decisions. Using a set of decision forests approaches, they distilled a number of risk indicators to predict which contractors are most likely to succeed in their federal contract obligations. Their method also uses publicly available data, but it focuses primarily on the financial and business aspects of the contractors. Our approach seeks to include natural disaster risk into the risk analysis model.

Tudoreanu et al. (2018) investigated employment data in an attempt to correlate changes in employment with negative modifications to contracts. Such correlations can be exploited to infer hidden and undisclosed contractors that are part of the defense acquisition network. Hidden contractors may pose the risk of becoming a weak stress point of a project and would affect the overall outcome of the project.

Wu et al. (2018) proposed a framework based on data science approach that aims to utilize the online information to assess and improve acquisition database quality as well as to find the hidden patterns to further acquisition research. The main component of the framework is a web-search and text mining module, whose main function is to search the internet and identify the most credible and accurate information online.

Apte et al. (2016) explored the use of big data analytics techniques to explore and analyze large data sets that are used to capture information about Department of Defense (DoD) services acquisitions. The paper described how big data analytics could potentially be used in acquisition research. As the proof of concept, the paper tested the application of big data analytic techniques by applying them to a data set of Contractor Performance Assessment Report System (CPARS) ratings of 715 acquired services. It also created predictive models to explore the causes of failed services contracts. Since the data set used in the research was rather small and far from the scope of big data, the techniques explored by the paper mainly focus on traditional data mining techniques without taking into account big data properties.

Black et al. (2014) studied the quality of narratives in CPARS and their value to the acquisition process. The research used statistical analysis to examine 715 Army service contractor performance reports in CPARS in order to understand three major questions: (1) To what degree are government contracting professionals submitting to CPARS contractor performance narratives in accordance with the guidelines provided in the CPARS user's manual? (2) What is the added value of the contractor performance narratives beyond the value of the objective scores for performance? (3) What is the statistical relationship between the sentiment contained in the narratives and the objective scores for contractor evaluations?

Previous Research Results

This paper builds upon the natural disaster risk model by Wu et al. (2020), and we review the main points here. A natural disaster risk was calculated by integrating two sources to data, NCEI and FEMA. Weather data of all U.S. counties between the years 1950 and 2018 from https://www.ncei.noaa.gov/ and natural disaster assistance data from https://www.fema.gov/ covering 1953 to 2020 was used to create a weighted disaster score (WDS). That score uses the declared assistance programs to assess a natural disaster's intensity and damage level. The score itself is then analyzed and simplified to a three-level scale—namely high, medium, and low—based on the distribution of the WDS. Figure 1 shows the distribution of risk levels both geographically and by location of federal contractor.

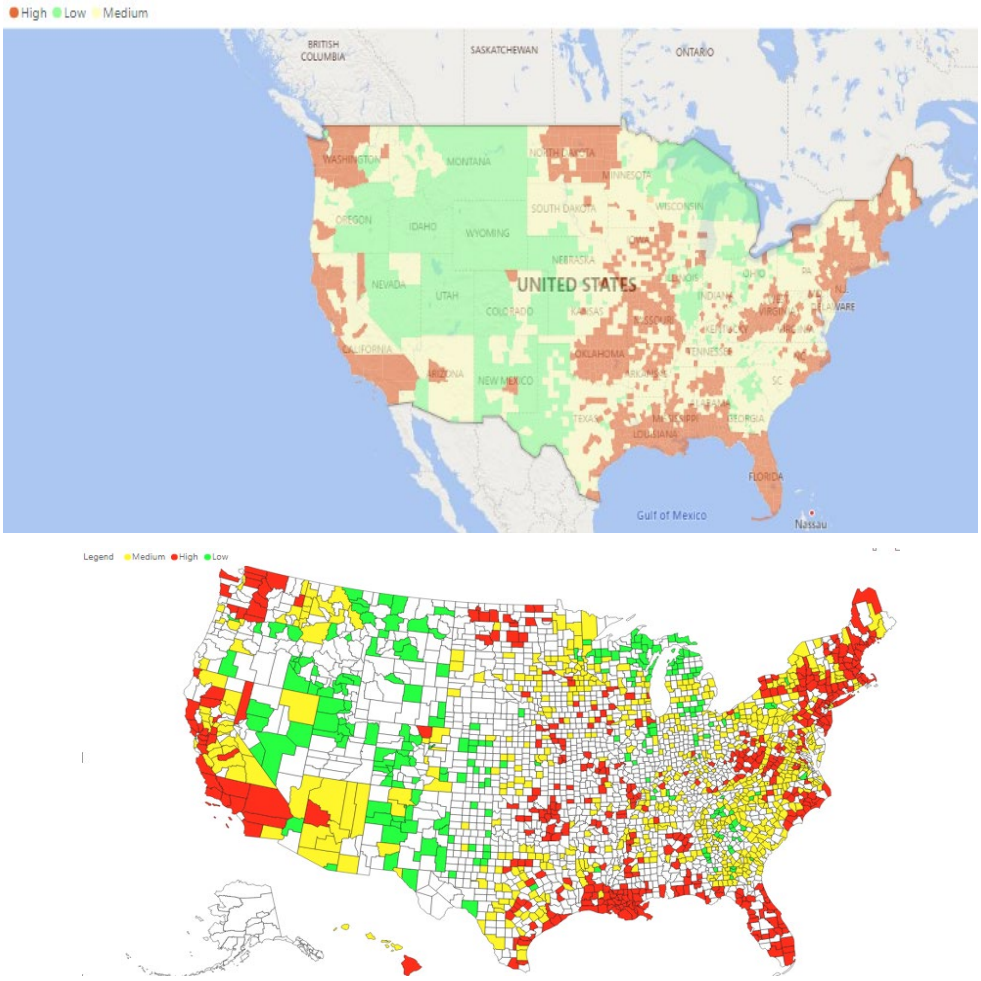

Figure 1. Natural Disaster Risk Level (Low, Medium, High), Displayed Geographically (Top) and by Location of Federal Contractor (Bottom) for the United States[1]

Research Results

The work presented in this paper can be broken down into three main thrusts:

- Find awards with a large number of contractors in high-risk areas;
- Show alternative areas that have the desired industrial concentration, industry type, and lower natural disaster risk level;

[1] The index takes into count both the number of occurrences of disasters and their magnitude.

- Show alternative individual contractors that are of the desired industry type and satisfy other geographical requirements a project might have.

All of these are meant to provide an acquisition expert the ability to assess the risk posed by natural disasters to a project or a set of projects. Furthermore, our goal is to also assist an expert in selecting alternative subcontractors in order to mitigate delays and failures posed by natural disasters.

Determine Projects with a Large Number of Contractors in High-Risk Area

The framework we propose is to maintain a database with the information from NCEI and FEMA. That database is periodically updated, and a new WDS is computed for every county in the United States. Specifically, we focus on FEMA disaster declarations and determine the number of disasters in an area during a period and the types of assistance programs declared. The data are cross-referenced with the NCEI to only consider natural disasters and eliminate man-made ones. WDS is calculated as follows:

$$s = \sum_{i=1}^{4} w_i \times n_i$$

where n_i is the number of a specific type of the assistance programs, and w_i is the corresponding weight for the type. The weight for each assistance program is defined as follows:

- Disaster mitigation: 0.25
- Public assistance: 0.50
- Housing assistance: 0.75
- Individual assistance: 1.0

The next step is to search for projects that have more contractors located in high-risk areas of natural disasters than in other types of areas. For https://www.usaspending.gov/, the process involves locating sub-awardees for each project, and it is relatively straightforward given the organization of the publicly available data. For https://www.fpds.gov/, the archives are provided in a different format, and we use a combination of data fields to match project to subcontractors. A different processing pipeline is used for each of the two data sources. For https://www.usaspending.gov/, database queries and Tableau are employed to obtain the dashboard, while for https://www.fpds.gov/, a Python and Jupyter Notebook pipeline was developed.

Alternative Geographical Areas for an Industry Type

We developed a technique to allow an acquisition expert to better inform their source-selection decision with regard to natural disaster risk. This thrust of our approach aims at providing the means to discover regions, in particular counties, where there exists some concentration of industry with a given profile. The level of concentration is to be set by an acquisition professional, and it involves the existence of a certain number of historical federal contracts that have been awarded in that region. The industry type can be specified through NAICS codes.

A program manager can examine the historical distribution of federal contractors for a given industry type (i.e., NAICS code). Figure 2 shows the distribution for a sample code that includes ventilation, heating, air-conditioning, and commercial refrigeration equipment manufacturing. In this scenario, the manager would then look at the areas that have medium or low risk of natural disaster and that also have been used in the past in many federal contracts. One possible view may be to show only those areas that have had an above-average number of federal awards (see Figure 3, top). If the selection is not sufficient, the analyst could lower the

requirements on the total number of past federal contracts, which shows more potential regions with a lower natural-disaster risk (see Figure 3, bottom).

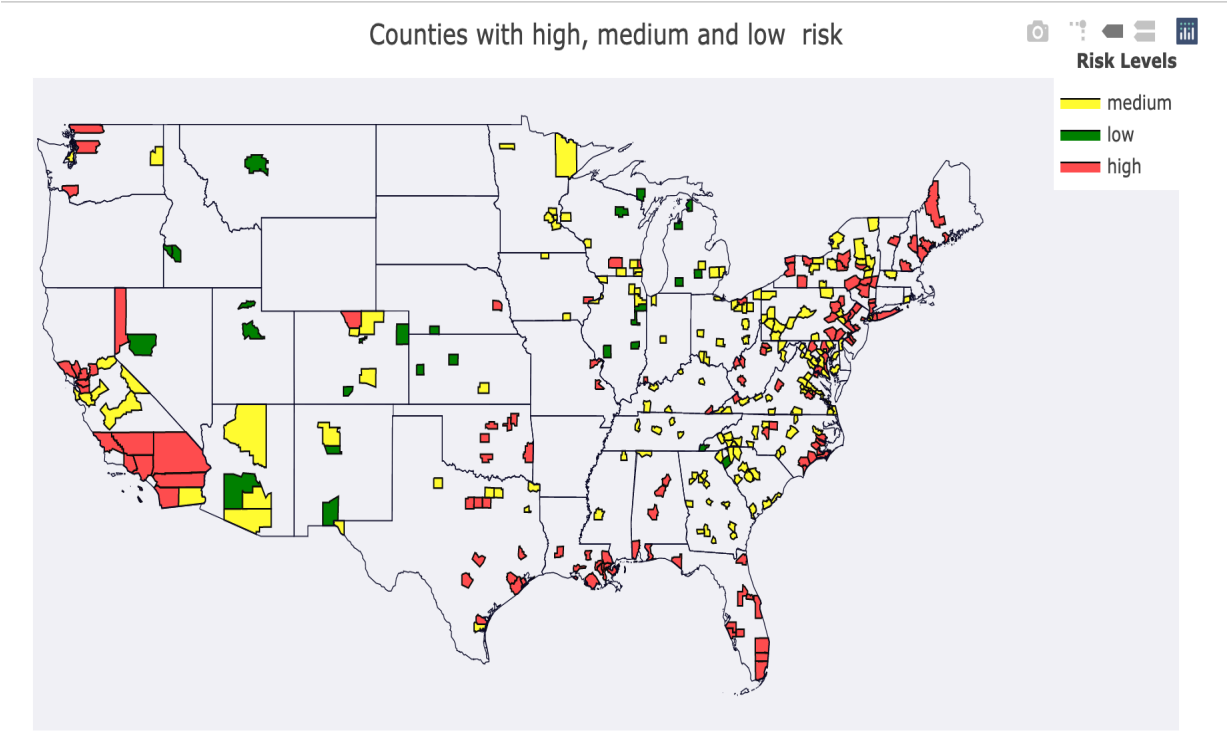

Figure 2. Distribution of Contractors from Historical FPDS Records for NAICS Code 33341, Department of the Navy[2]

[2] The color represents the natural disaster risk level for the place of performance.

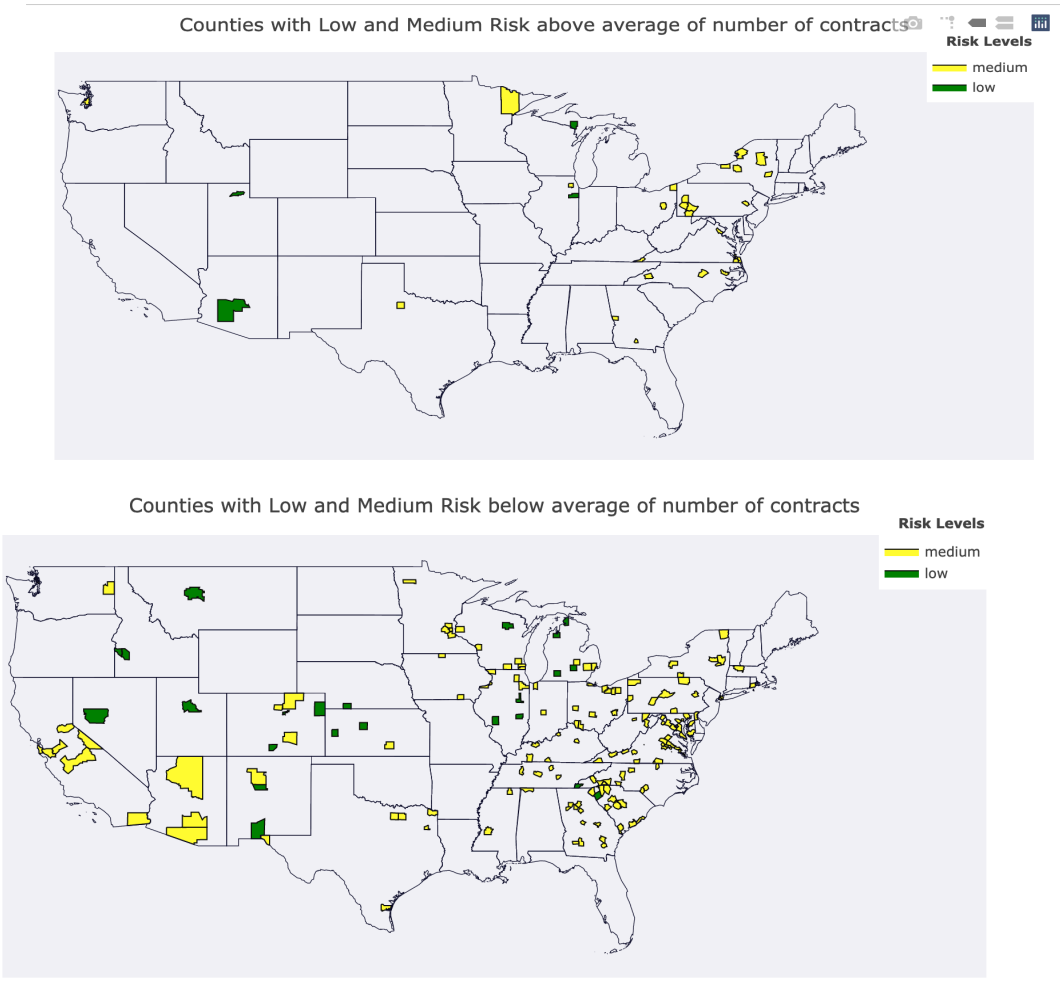

Figure 3. Regions in the Lower Risk of Natural Disasters That Have an Above Average of Past Federal Contracts (Top) and Below Average of Past Federal Contracts (Bottom)[3]

Alternative Contractors and Multiple Award Comparison

The data in https://www.usaspending.gov/ are categorized based on the types of spending, which include contracts, grants, loans, and other financial assistance. The information on contracts is organized into two tables: one for prime contracts and the other for subcontracts. The subcontract table was used in this study as it contained the detailed information of each award, including the award ID and funding agency's information, as well as the business name, address, DUNS number, and NAICS code of both primary contractor and each subcontractor. In addition, the address where each subcontractor physically performed the work is also listed in the table. Combining with the information from our natural disaster risk assessment framework, we are able to identify the awards with a high percentage of subcontractors located in the high-risk areas. Furthermore, we are able to identify the subcontractors of a same business type based on their NAICS codes and partition them by their locations and the natural disaster risks of their locations.

For illustration purposes, two contracts with a higher-than-average concentration of contractors in high-risk areas were randomly selected and are apparent in Figure 4. One of the awards has some of the largest numbers of contractors; the other has a relatively small number of contractors. For each business type identified by a unique NAICS code, we partitioned the corresponding contractors into

[3] Data is for NAICS code 33341, Department of the Navy contracts.

two groups: one located in a high-risk area, the other in an area with either a medium or low risk level. The contractors in the second group can be used to find alternatives to high-risk contractors when needed and are the subject of Figure 5.

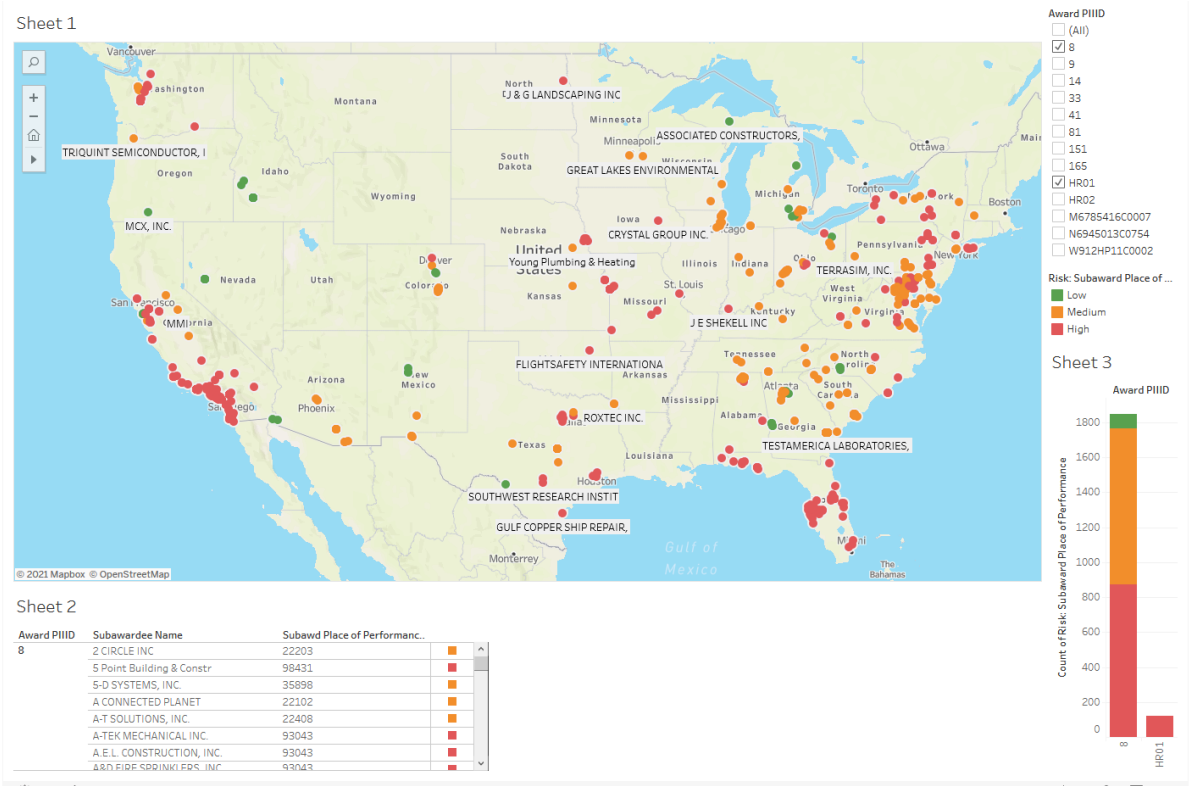

Figure 4. Dashboard Showing Sub-Awardees for Two Awards as Seen in the Checkboxes in the Top Right[4]

The first dashboard, Figure 4, is designed to provide a view of the distribution of subcontractors for a project. It consists of a map in which the places of performance are shown on a map and color-coded to reflect the natural disaster risk level. A list of awards of interest is provided on the top-right side. The bottom-right shows a bar chart, which gives an idea of the size of an award in terms of number of subcontractors. Stacked bars show how many contractors operate in areas with high, medium, or low risk. The final view, bottom-left, provides a list of contractors, places of performance of their services, and a color-coded risk level. Note that the same contractor may be producing goods or performing services in multiple places.

The interaction with this dashboard allows an expert both to analyze a single award and to compare multiple ones. The panels are linked and selecting or clicking on one element in one panel results in changes in all the others. The map allows the selection of geographical regions and the analysis of the subcontractors in that area. The bar chart can be used to focus on subcontractors from any selected project and allows the user to focus further on those subcontractors who operate in the desired risk category: high, medium, or low. Finally, the individual list of subcontractors provides fine-grained details and allows the user to examine one contractor at a time in terms of location, map position, and natural disaster risk level.

[4] The four panels are map, award selector, bar chart showing number and risk type for awardees, and individual contractor list. Panels are linked, and selection in any one of them determines what is displayed in the other three.

In addition to analysis and comparison, a program manager may need to be able to find alternatives for some subcontractors in order to mitigate the risk posed by natural disasters. To this end, our framework can search existing and past federal projects to find such alternatives. The results are provided to a user visually through the dashboard shown in Figure 5. The data are automatically filtered to eliminate high-risk areas of performance, and the sheer number of potential solutions can be further narrowed down using NAICS codes. For this paper, we defaulted to a five-digit NAICS code, which can provide relatively closely related businesses. Just as the previous panel, the map is the main feature of the dashboard, and it is augmented by a NAICS code selection panel and a list of subcontractors.

The user of this panel would start by focusing on one or more likely related NAICS codes. Codes can be added into the analysis if a suitable subcontractor replacement is not found, or they can be removed if there are too many candidates. The map offers a way of finding subcontractors who are close to a geographical area, to waterways, or to highways if transportation is an issue to be considered by the project. Finally, the list of subcontractors provides both the location of the subcontractor and the place(s) of performance of the service or work.

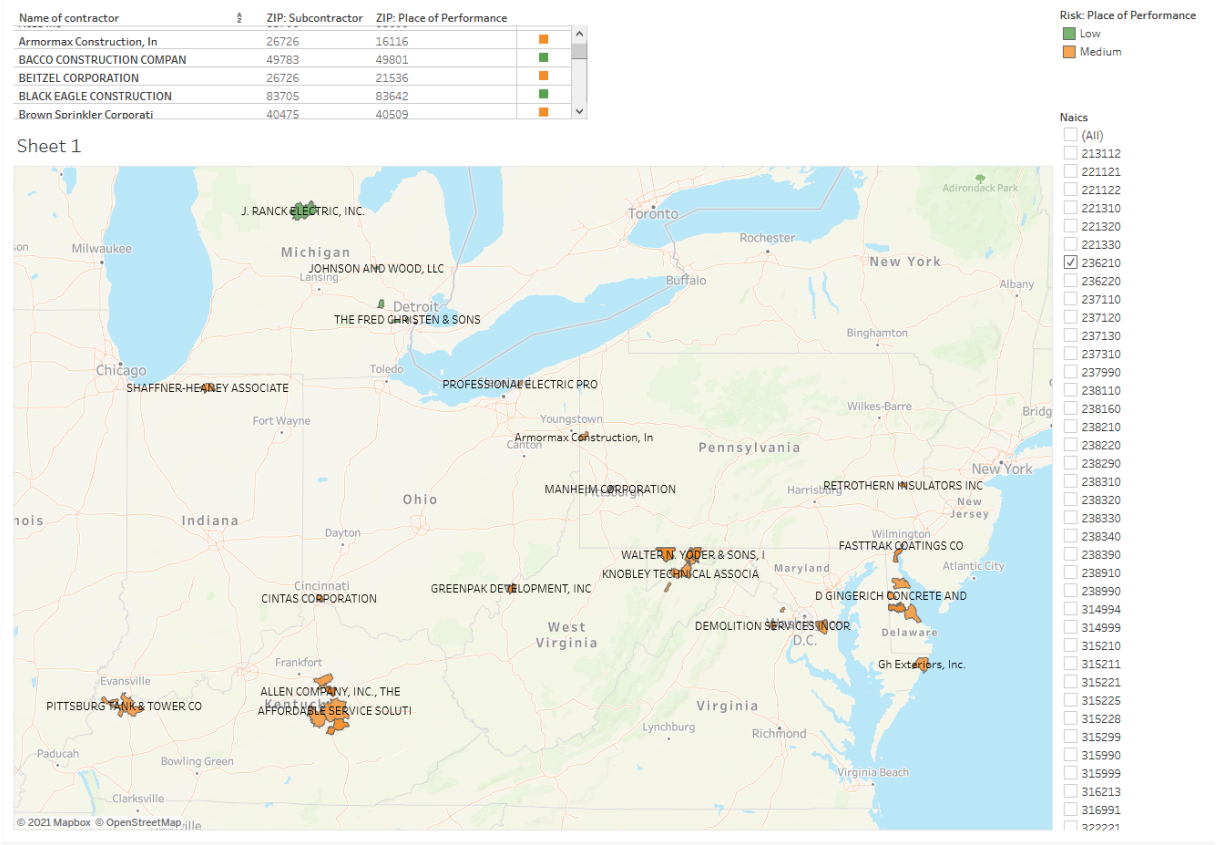

Figure 5. Dashboard for Finding Alternative Contractors by Name (Top Panel), NAICS Codes (Left Panel), or Geographical Location (Center Map)

Conclusion and Future Work

The paper presented a framework for allowing a user to manage the risk posed by natural disasters to a project or program. We developed visual tools to allow a program manager or planner to identify potential alternative subcontractors that operate in areas with a lower risk of natural disasters. Visual tools can also be used to analyze the existing risk of

natural disasters for an existing project. The tool can provide various levels of details from an overview of the where and how many high-risk subcontractors are operating to individual listing of each contractor. Furthermore, a domain expert can compare multiple awards and look for common potential risk among those awards. The expert can also analyze concentrations of various industry types in locations that are both prone to natural disaster and relatively safe from them. Finally, all of these visual, user-friendly tools are supported by data science approaches to organize, manage, integrate, and transform the underlying data. The research results would be helpful for the acquisition management and planning to control the risks of natural disasters and their impacts to a project.

Our future work will focus on two directions. First, investigate other risk factors in order to develop a comprehensive, data-driven risk assessment framework that can be applied quickly to both large and small acquisition and purchasing projects. Second, explore the use of deep-learning and website crawling to provide more focused and real-time information about contractors to a domain expert.

References

Apte, U., Rendon, R., & Dixon, M. (2016). Big data analysis of contractor performance information for service acquisition in DoD: A proof of concept. *Proceedings of the 13th Annual Acquisition Research Symposium*, *2*, 398–418. https://calhoun.nps.edu/handle/10945/53512

Black, S., Henley, J., & Clute, M. (2014). *Determining the value of Contractor Performance Assessment Reporting System (CPARS) narratives for the acquisition process* (NPS-CM-14-022) [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. https://calhoun.nps.edu/handle/10945/42583

Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015). A systems engineering perspective on the revised Defense Acquisition System. *Systems Engineering*, *18*(6), 584–603.

Gaither, C. C. (2014). Incorporating market based decision making processes in defense acquisitions. *International Journal of Defense Acquisition Management*, *6*, 38–50.

Gill, D., Muir, W. A., & Rendon, R. G. (2019). Predicting federal contractor performance issues using data analytics. *Proceedings of the 16th Annual Acquisition Research Symposium, 2*, 16–27. https://calhoun.nps.edu/handle/10945/62901

Krzysko, M., & Baney, B. (2012). The need for "acquisition visibility." *Journal of Software Technology*, *15*(1), 4–9.

Office of Management and Budget. (2017). *North American Industry Classification System*. Executive Office of the President. https://www.census.gov/eos/www/naics/2017NAICS/2017_NAICS_Manual.pdf

Tudoreanu, M. E., Franklin, K., Rego, A., Wu, N., & Wang, R. (2018). Searching hidden links: Inferring undisclosed subcontractors from public contract records and employment data. *Proceedings of the 15th Annual Acquisition Research Symposium*, *2*, 491–499. https://calhoun.nps.edu/handle/10945/58788

Wu, N., Tudoreanu, M. E., Siruvolu, S., Wang, R., & Talburt, J. R. (2020). Analysis of contractor data in federal acquisition databases. *Proceedings of the 17th Annual Acquisition Research Symposium*, 1–11. https://calhoun.nps.edu/handle/10945/64767

Wu, N., Tudoreanu, M. E., & Wang, R. (2018). Leveraging public data for quality improvement and pattern discovery of federal acquisition data. *Proceedings of the 15th Annual Acquisition Research Symposium*.

Zhao, D., MacKinnon, D., Douglas, J., Gallup, D., & Shelley, P. (2015). *Lexical Link Analysis (LLA) application: Improving web service to Defense Acquisition Visibility Environment* (NPS-AM-15-129). Naval Postgraduate School. https://dair.nps.edu/handle/123456789/2653

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net