



ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Acquisition Requirements and Root Cause Analysis: A Data-Centric Perspective with Data Governance, Data Analytics, and Data Quality

September 14, 2020

Dr. Richard Wang

University of Arkansas at Little Rock

Disclaimer: This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. HQ0034-18-1-0011. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Approved for public release; distribution is unlimited.

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



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 the Acquisition Research Program (ARP) via email, arp@nps.edu or at 831-656-3793.



ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Abstract

There is a cyclical relationship between the acquisition community and contractors that does not function in the same manner as the free market, where pressures self-regulate supply and demand. Fluctuations do not exist in the same manner in this closed acquisition environment. However, this closed system does interact with external systems that are part of the free market, such as the labor market. This research analyzes data from employment and Federal contracts and provides a characterization of how acquisition contracts affect local employment. Employment variations can potentially be used to infer undisclosed subcontractors participating in multiple awards. Such subcontractors should be accounted for by Integrated Product Teams (IPTs) during the development of more accurate Integrated Master Plans (IMPs). Many Federal and DoD contracts are performed by a team of contracting entities, where some prime contractors rely on subcontractors to execute specific parts of the contract. For many reasons, including national security, privacy, or competitive advantage, some of these subcontractors are not publicly disclosed. This study presents a large-data analysis process for understanding the boost to local employment that the start of a large contract may provide. The results show that locations of large Navy awards rank above 70% of other locations in the country in terms of the magnitude of employment increases. A secondary analysis of negative contract modifications revealed that employment decreases when small contractors have relatively large reductions in DoD contracts.



THIS PAGE LEFT INTENTIONALLY BLANK





ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Acquisition Requirements and Root Cause Analysis: A Data-Centric Perspective with Data Governance, Data Analytics, and Data Quality

September 14, 2020

Dr. Richard Wang

University of Arkansas at Little Rock

Disclaimer: This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. HQ0034-18-1-0011. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.



THIS PAGE LEFT INTENTIONALLY BLANK



Table of Contents

Introduction	1
Related Work	5
Methodology.....	7
Big Data Process.....	7
Analysis of Negative Contract Modification.....	12
Results	15
Big Data Analysis.....	15
Negative Contract Modification	17
Conclusion and Future Work.....	21
References.....	23



THIS PAGE LEFT INTENTIONALLY BLANK



Introduction

The Defense Acquisition System (DAS) functions best when relevant information is known and shared among the entire Integrated Product Teams (IPTs) as it is spelled out in DoD Directive 5000.01. Information relevant to proper identification and management of risk leads to higher quality Integrated Master Plans (IMPs). Such information may be as important in the context of the Adaptive Acquisition Framework and the use of the more flexible middle tier pathway. However, across the complex array of acquisition projects, there are a myriad of overlapping supply and service chains, some involving the same contractors or subcontractors.

The context of Augustine's Law (1997) as it relates to acquisition in the US military suggests a cyclical relationship between the acquisition community and contractors that does not function in the same manner as the free market, where pressures self-regulate supply and demand. Fluctuations do not exist in the same manner in this closed acquisition environment. However, the ecosystem of DoD contracts often involves multiple entities, both large enterprises and small businesses, who work together to support the warfighter. This ecosystem does interact with outside systems that are more based on traditional supply and demand. Federal contracts can be awarded to single entities, but also to multiple contractors. Furthermore, contractors rely on other entities, henceforth subcontractors, to perform specific parts of the contract. For reasons of security, confidentiality, or competitive advantage, some of these subcontractors are not publicly disclosed. One overarching research question is whether enough publicly available data exists to allow for the discovery of undisclosed subcontractors.

The open society and transparent government of the USA, through contributions from local and federal government as well as private companies, makes information available on a wide variety of topics, from air quality to social interactions, from Federal contracts to employment status. It is unlikely that a single such data repository would allow information about hidden subcontractors to be determined, though the



combination of multiple data sources might do just that, especially if the same undisclosed contractor is participating in multiple contracts.

The data-driven approach taken in our research is to analyze a large number of contract events, specifically the start, end, and any modifications. The reasoning is that one such event, for example securing a contract to perform work for the federal government, may lead, at least in some cases, to a boost in employment in those locations closely related to the performance of the work. Similarly, the ending of an award or a negative modification may result in a drop in employment. When a large number of events are processed together, it may become possible to hone in the location of an undisclosed contractor as shown in Figure 1. This process could potentially be used to determine the likelihood of a location to be home to an undisclosed contractor for a given industry. A number of factors will influence the ability to detect the correlation between awards and employment, but two are worth considering next. First, the larger the award or award modification, the more likely it is to produce an effect on the employment data. Second, employment variations can be better detected in smaller cities and rural areas than in large metropolitan regions.

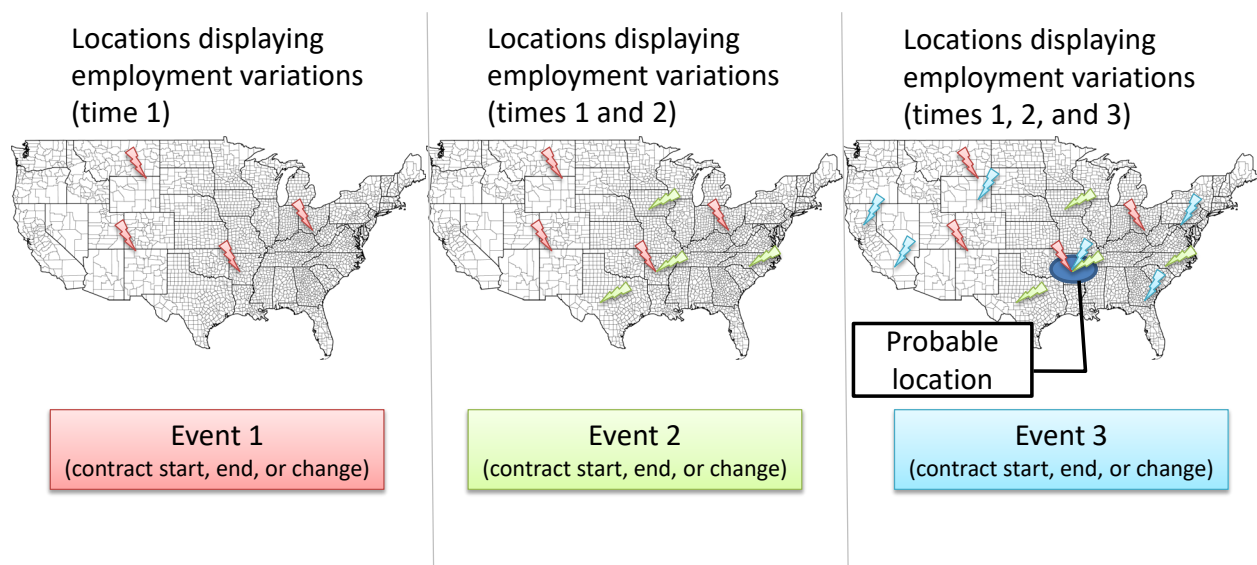


Figure 1: Determining possible locations for subcontractors based on repeated correlations between contract events and employment variations. Undisclosed contractors who participate in many awards, and thus pose more risk to acquisition, are more likely to be discovered.

The research focuses on the narrower issue of determining the relationship between known contract events (beginning, end, modification) and employment variations at or around the same time both

- a) at the disclosed locations from the contract, which provides a known relationship that can be reliably assessed, and
- b) in the rest of the country, which will provide a measure of the existing noise in the data.

That noise can be further reduced by examining only industries related to each individual contract being studied and by considering employment trends at a local level in the context of the country as a whole.

The report presents a process to analyze two large open data sources, one about federal contracts and the other about employment in the USA. Based on the datasets derived through the process, this research determined that a promising metric may be to consider the magnitude of employment changes at a given location and for a given classification of industry type. On the datasets for the year 2016, the metric shows that locations that benefit from large contracts are in the top 30% of all locations in terms of magnitude of employment. Thus, through repeated elimination, it may be possible to narrow the location of an undisclosed contractor to a manageable number of places in the country. Those places can then be searched manually for web or social media presence of businesses capable of being such an undisclosed contractor.

Secondarily, a small-scale analysis reveals that reductions in contracts, which are large relative to the contractor's size, are correlated to a drop in employment in more than two thirds of the cases. Our research used two types of metrics, one linked to absolute employment levels for a county and industry type and the other, location quotient (2008), providing a relative measure of a county's employment in an industry sector relative to the nation. The second metric would be able to register a drop even if a particular location had an actual increase in employment, but that increase was not as large as the national trend. The location quotient showed a decline for 75% of the contract cuts studied.



THIS PAGE LEFT INTENTIONALLY BLANK



Related Work

Our previous ARP Symposium work (Tudoreanu, Franklin, Rego, Wu, & Wang, 2018) used a manual correlation process and revealed that reductions in contracts (modifications that cut the amount of money originally allotted) that were large relative to the contractor's size correlated to a drop in employment in more than two thirds of the cases. Location quotient (LQ), which provides a relative measure of a region's employment in an industry sector relative to the nation, showed a correlation for 75% of the studied contracts' reduction. The main difference between this research and the past is the addition of positive contract modifications (including contract start date) into the analysis, the use of data science to be able to handle and process tens of thousands of Navy contracts, and the examination of the noise present in employment data.

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 & Barney, 2012). To this end research has looked at automatic means of dealing with the heterogeneous acquisition data sources from text processing (Zhao, MacKinnon, Douglas, Gallup, & Shelley, 2015), systems engineering (Cilli, Parnell, Cloutier, & Zigh, 2015), and business (Gaither, 2014) perspectives. Our work is different both in content and in the approach. In content in that we seek to characterize hidden flow of funds in the supply network, and in the approach in that our expertise in data quality and data science provides a more value-based perspective.

A recent approach used the data collected by government agencies. The objective was to use administrative data, i.e., state unemployment insurance information that is from covered wages and salaried workers based on the workers' quarterly earnings. In studying domestic outsourcing, the primary constraint was how existing data is limited in estimating the number of workers who are employed by contractor companies or who provide services to firms as independent contractors. This data "can be used to document employment in contractor firms and the number of independent contractors, link contractor industries with the firms using their services." (Houseman &



Bernhardt, 2017). The paper also recommends using earnings information that includes wages, salary employment, and self-employment tax data that can be provided by the IRS.

A second paper “uses linked worker-firm administrative tax data from U.S. tax returns to explore the changing relationship between firms and independent contractors.” (Miller, Risch, & Wilking, 2017). A dataset was constructed, that used digitized tax filings from the IRS, for the tax years of 1997 to 2015. Individuals were linked to their employer via reports Form 1040, Schedule C and Schedule SE, and information reports W-2, 1099-MISC and 1099-K.

Other researchers have successfully integrated publicly available data with private information, but their goal was radically different than the scope of our research. Such data integration efforts have been employed to reconcile corporate names (Gayo, Pablos, Rodriguez, & Vafopoulos, 2013) and to provide a tool for accessing unified corporate names (Llorens, Rodriguez, & Vafopoulos, 2015). Another approach involved government data enhanced with information services from outside sources to provide additional, generalized context into the data (Felten, Robinson, Yu, & Zeller, 2009).



Methodology

Big Data Process

Two data sources were used in our analysis, Federal Procurement Data Systems - Next Gen (FPDS) and Bureau of Labor Statistics (BLS). The procurement database provides a list of federal contracts awarded over the years, including any modifications to an award. The labor statistics includes employment data on a monthly and quarterly basis, both in absolute numbers and relative to the rest of the country, through the location quotient (LQ, 2008). Due to the limited amount of time and computational resources, the primary focus was limited to the year 2016, yet even this approach resulted in some operations involving tens of millions of entries and long processing times. The methodology described here is applicable to the analysis of additional years.

Data from the two sources was copied to a local database hosted on a relatively powerful server powered by an 8-core, dual Intel Xeon processor with 48 GB of memory. Java programs and MySQL queries were used to populate the database.

The data processing steps are listed below, along with an explanation of the types of data involved in those steps. The end-goal is three-fold:

- a) to determine whether a large contract correlates with an increase in employment given the contract's industry and location,
- b) to determine the overall behavior of employment in the entire US, and
- c) to compare the employment at contract's location to all the other locations in the country at the same time (of the year) and industry in order to be able to determine a metric that has the potential to uncover hidden subcontractors.

To this end, the processing steps in the middle (numbered 3 and 4 below) follow two separate pathways, one to examine employment behavior strictly related to FPDS contracts, and one to study the employment trends in the entire country (Figure 2).



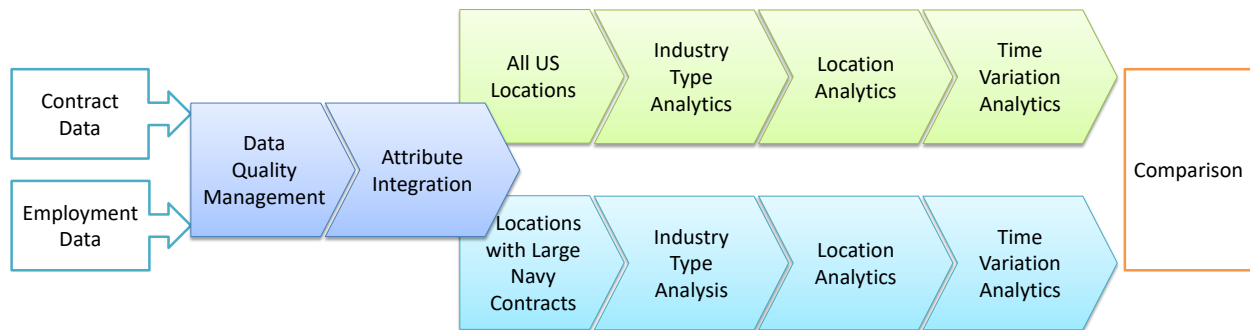


Figure 2: Overview of the processing framework.

Step 1: Select contract data from FPDS

The analysis of contract data started with FPDS records for the Department of the Navy, and it underwent two selection steps. First, only contracts and modifications with a start or end date in 2016 were considered. Second, the analysis filtered "small" contract events (begin, end, or modification events) which in this case were events with dollar amounts in the range \$-99,000 and \$499,000. That is only changes to a contract that either reduced the amount awarded by more than \$99,000 or had more than \$499,000 were considered because they have a larger chance on producing layoffs or hiring. This step resulted in over 23,000 contracts (and modifications) being selected.

While a wealth of information is available for each contract, the relevant data to be used for the rest of the analysis includes:

- dollar amount;
- zip code of the principal place of performance, which can provide the local area in which to examine employment;
- start date, used as the effective date when the contract could start affecting employment;
- NAICS code, which is a standard way of classifying various types of industry in US and Canada (NAICS, 2017). The codes in FPDS are 6 digits long. NAICS uses a somewhat hierarchical structure where related sub-industry share the first few digits of the larger industry type. For example, a four-digit ABCD code is generally a sub-type of the three-digit ABC industry.

Step 2: Employment Data from Bureau of Labor Statistics (BLS)

Employment data included both monthly and quarterly metrics. Due to the nature of our search, the year 2016 was selected because it was already in our systems. The

last quarter of 2015 and the first of 2017 were also considered for some of the steps below. The number of quarterly entries in the year 2016 alone is over 14 million (that computes to over 52 million monthly entries) for all covered locations and industry codes.

The following fields were deemed important:

- month of the year to determine the time frame;
- industry type, which was provided as NAICS (2017) codes. Three, four, five, and six digit industry codes were used (see Step 1 for an explanation of the NAICS code length);
- US county and state;
- number of people employed in the county per industry type and month. Note that not all counties have employment for all NAICS codes;

Step 3: Contract Pathway: Filter Relevant BLS Data Based on Large Contracts

From this point, two pathways were pursued in the processing. One that focused on employment relevant to an existing large contracts, henceforth termed Contract Pathway, and one that focus on the overall employment situation in all locations in the US, named All Locations Pathway.

This steps reduced the size of employment data (step 2) to only those locations that appear in at least one of the large contracts selected at step 1. The zip code of the place of performance of a contract was mapped to county and state. All monthly employment information about that location was copied in a working data store regardless of NAICS or time of year. All of 2016 and last quarter of 2015 and first of 2017 went through this transformation. For the year 2016, only about 2 million quarterly (that is 6 million monthly) entries were excluded.

Step 3: All Locations Pathway: Filter Relevant BLS Data Based on Large Contracts

While it is important to analyze the employment trend in the country as a whole, there is little reason to consider NAICS codes that are completely unrelated to any of the large contracts in the other pathway because employment performance in one part



of the economy (say agriculture) is not directly correlated to performance in another type of industry (say information technology).

We designed an easily measurable definition for what would mean for two industries to be related. The FPDS data provides a six-digit NAICS code for the product that is the subject of the contract. BLS uses the NAICS code slightly differently in that BLS identifies the type of employer. It is possible that one employer may have products in different, yet related industry types. Our definition of related industries relies on the hierarchical nature of NAICS. Thus, in addition to the six-digit code known from FPDS, we also considered the, often more encompassing, five-digit code obtained by removing the last digit, as well as the four- and three-digit ones. Thus, for each FPDS contract, we created a bundle of related four related industries by the process of removing last digits of the code.

Given our definition of related industry types, this step removed all employment information for NAICS codes that do not belong in an industry bundle from at least one of the large contracts. Note that this automatically excluded codes with fewer than three or more than 6 digits. The result was a dataset of about 4.8 million quarterly (14.4 million monthly) employment entries. No data from years 2015 nor 2017 was included.

Step 4: Contract Pathway: Calculate Employment Before and After Contract Start Date

The start date of the contract may be a possible event that can lead to observable changes in employment levels. It is unlikely that any business would start hiring before the contract begins, and thus before funds become available. One good temporal point to be considered as the reference employment is the month right before the start date. We will refer to it as the *before* employment level. Similarly, after the contract is awarded, the business may hire additional people to perform the work. The hiring may actually take some time, so we decided to examine the contract start month and two more months after for any changes in the employment. Positive dollar amounts might lead to higher employment, therefore we use the maximum employment level reported in any of the three months (start month and the immediately following two months) as our *after* employment metric.



The before and after levels were computed for each contract and for each NAICS code from the related industry bundle (see step 3 All Location Pathways for an explanation of the bundle). An entry in this dataset consists of:

- contract data (dollar amount, location, start date);
- NAICS code (either a three-, four-, five-, or six-digit code). Not all combinations of location, code, and month have reported employment data;
- local before and after employment levels.

The dataset has over 53,000 entries, which include the multiple industry bundles, with about 16,400 unique contracts/modification. Some of the contracts from step 1 were not included because the employment information was not available for that location and time of the year either before the contract start date, or after, or both.

Step 4: All Locations Pathway: Calculate Employment Before and After Each Month

A similar technique was used to calculate employment levels in every county in the US for months starting in February and ending in October 2016. The before level is the one from the month before the month being calculated (thus the first one is February), and the after level is maximum of the three month following, and including, the target month (thus the last possible target month is October because it requires November and December data).

The dataset has the following fields, with more than 3 million entries. Note that it is possible that employment level was not reported for some locations, NAICS, and month combinations.

- month;
- NAICS code;
- county;
- before and after local employment levels.

Step 5: Combining the two pathways

This final step of data processing involves comparing each contract entry from step 4, Contract Pathway, to each possible location in US (step 4, All Locations



Pathway). All possible combinations were generated as long as they had the same NAICS code and the same month. The total exceeds 52 million combinations. In addition to being able to perform a pair wise comparison, we also determined the national average before and after employment levels at the time of each contract.

Analysis of Negative Contract Modification

The data used in our analysis originated from Federal Procurement Data Systems - Next Gen (FPDS) and from Bureau of Labor Statistics (BLS). The first source provides a list of federal contracts awarded during the years, while the second published employment data. The data used in the research covered the period of 2013 through 2017.

Federal Procurement Data Systems (FPDS)

Our initial focus for FPDS were contracts for the Department of the Navy. The intention was to identify contracts with a funding reduction larger than at least a yearly salary of a qualified, skilled employee that was affecting small and medium-sized entities. From over 1.1 million entries pulled for the Navy, about 3,300 modifications to a contract were found that met the following criteria: i) the contract was reduced by more than \$100,000, and ii) the vendor had between 10 and 200 employees. Both contract modification value and number of employees for the vendor are fields in FPDS.

The correlation between FPDS and BLS data is in the process of being automated, thus for this work the analysis was performed manually. This limited our ability to process all contract modifications identified in the Navy data to a more manageable 33 instances of reductions in the amount of the contract, which amount to 1% of the Navy contract reductions for the studied period.

The selection focused on instances which had the largest drop in contract amount per unit of annual revenue. We also restricted the industry to manufacturing (NAICS code beginning with 3) and IT services (NAICS code beginning with 581). Finally, to avoid some noise in the FPDS data, the selection only included entities whose revenue was larger than \$100,000 (both NAICS code of the product and annual revenue of a vendor are fields in FPDS).



Bureau of Labor Statistics (BLS) Employment Data

The employment data provided monthly absolute employment levels and monthly location quotients for each county and type of industry. The industry codes were also using NAICS (2017) to identify the type of employers. BLS uses the NAICS code slightly differently than FPDS in that BLS identifies the type of employer, while FPDS codes the type of product provided by an employer. It is possible that one employer may have products in different, yet related industry types.

Given that the single NAICS code existing in FPDS can be too rigid, a bundle of related industries were considered for any given location and date as sampled in Figures 3 and 4. This approach is further supported by a secondary analysis we performed, which reveals that in about one third of counties (912 counties) a drop in employment in one industry was accompanied by a drop in related industries.

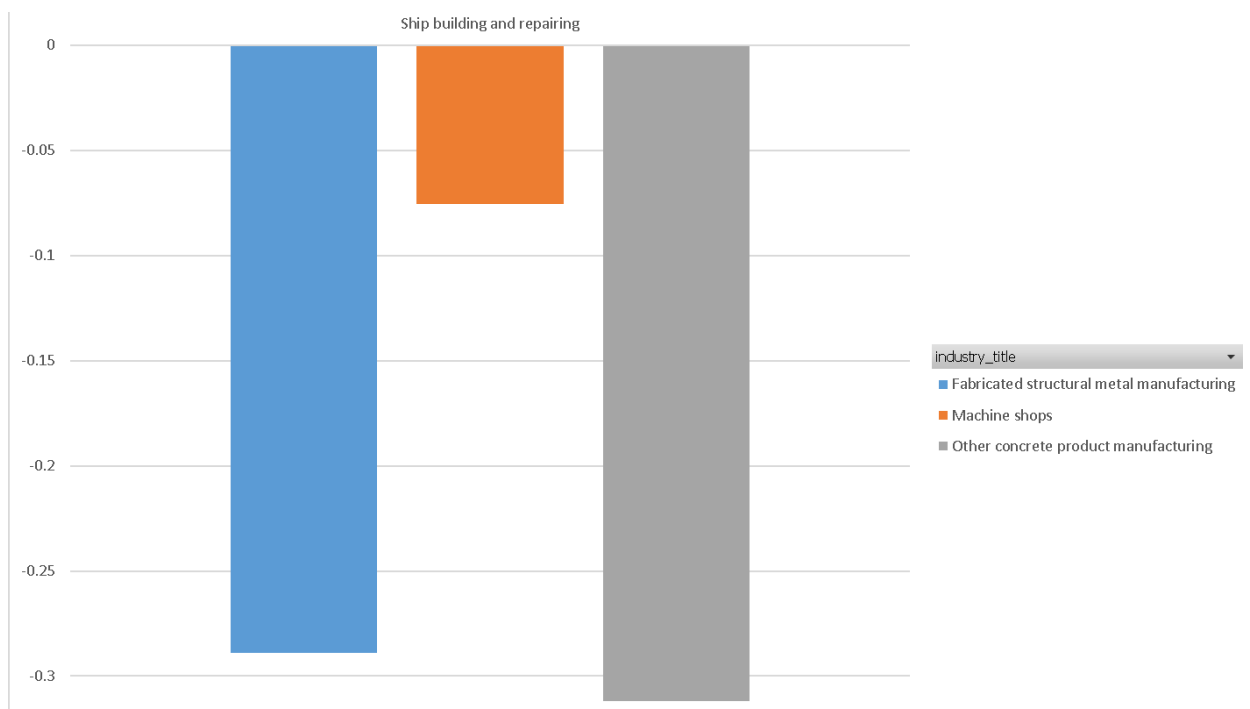


Figure 3: Sample industry-type bundle for Honolulu, HI, third quarter 2013. The bars show relative change in employment from the quarter before to the next quarter.

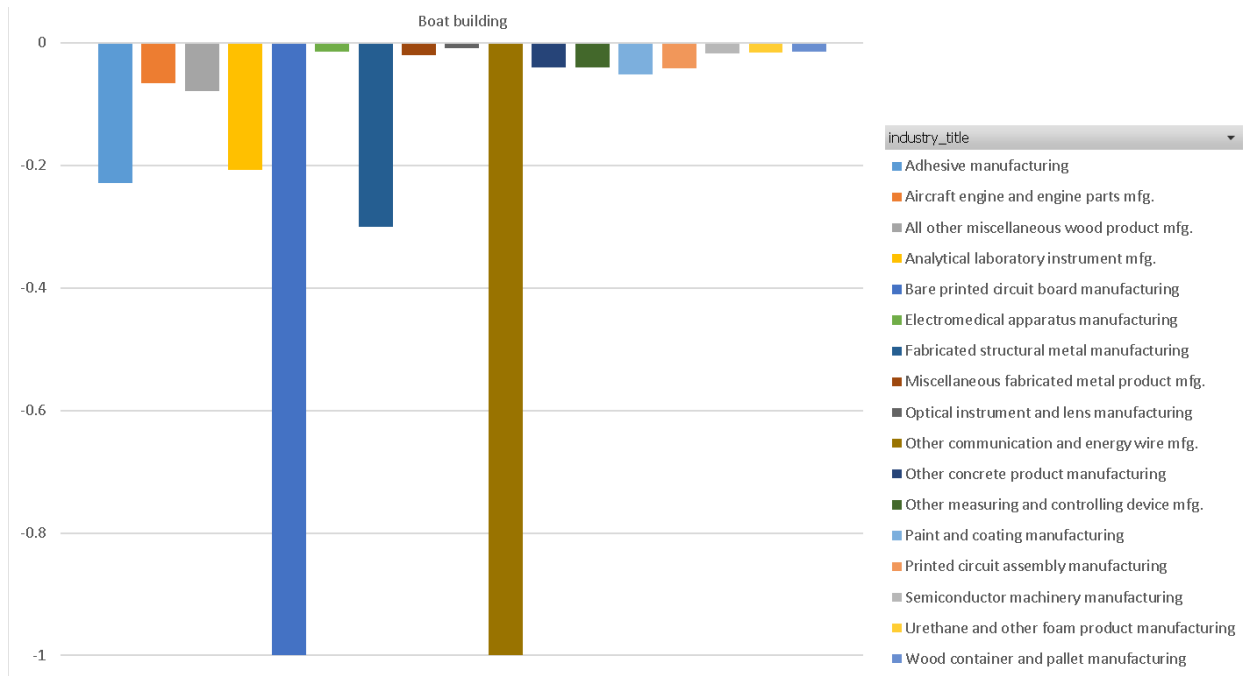


Figure 4: Sample industry bundle for Essex, MA, fourth quarter 2014. The bars show relative change in employment from the quarter before to the next quarter.

Two metrics were used to estimate any potential changes in the employment situation: absolute employment level and location quotient (2008). The latter, location quotient (LQ), measures the local economy, county in our data, in relation to the rest of the nation. As such, it can capture cases where the absolute employment moves up or down in response to national trends unrelated to any local factors (local factors such as a cut in a Navy contract). A drop in LQ value is likely explained by a worsening of the local economy unrelated to national trends. Note that LQ can decrease even when employment increases, signifying that the increase is not taking place as quickly as in the rest of the USA.



Results

Big Data Analysis

This section presents two broad metrics for employment levels. First, it just examines the whether the employment registered an increase as shown by the *before* level being smaller than the *after* numbers. Second, the magnitude of the change, whether increase or decrease, from *before* to *after* is presented. For this section, we focus on positive contract events, particularly a contracts start date and any positive modification to the dollar amount. The majority of the contracts, around 13,900, from step 4, Contracts Pathway, fall into this category.

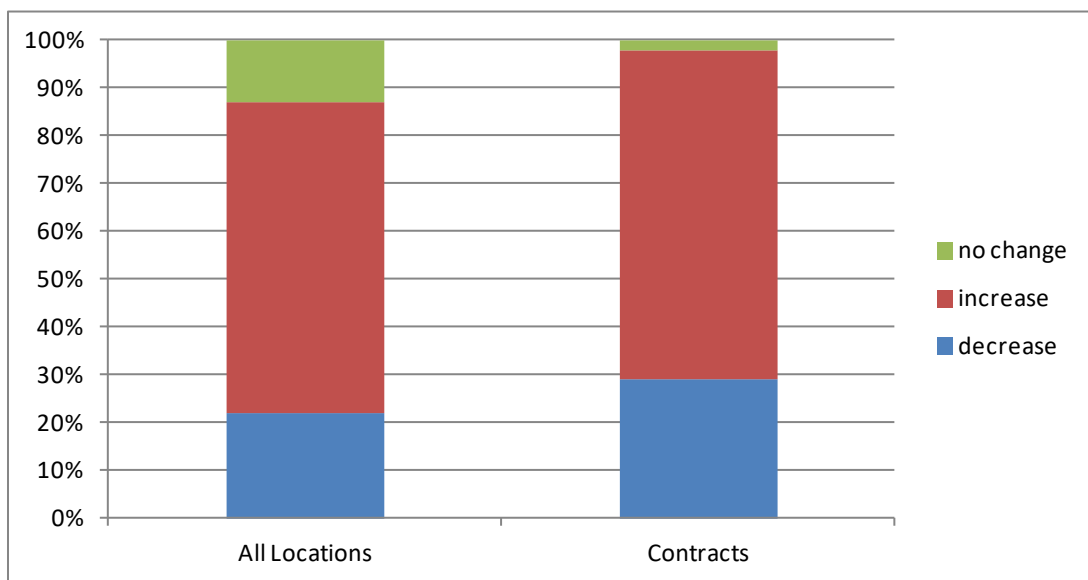


Figure 5: Relative changes in employment as recorded in the datasets produced in step 4. Both the country as whole over the year 2016 and those locations affected by awarded contracts show similar behavior in the number of instances where an increase is recorded (as a percentage of all entries in each dataset).

The country overall is experiencing an increase in employment during the studied period. Based on step 4, All Locations Pathway, in 65.2% of the counties and NAICS codes over the year, the employment records show an increase. For the locations, times, and NAICS codes related to contracts. that number is only slightly larger at



68.8% (from dataset produced at step 4, Contracts Pathway). The results are depicted in Figure 5.

A change in employment in itself does not seem to be sufficient to allow the discovery of an undisclosed contractor. Thus, a second metric, the magnitude of the employment change, was considered. Formally, the magnitude is obtained by subtracting the *before* number from the *after*. The magnitude can be negative if the employment level drops. Using the data from step 5, the magnitude in the counties related to a contract, at the start date of the contract, and for the relevant industry bundle was compared to magnitude of change for the same industries and at the same time for each and every county in the US.

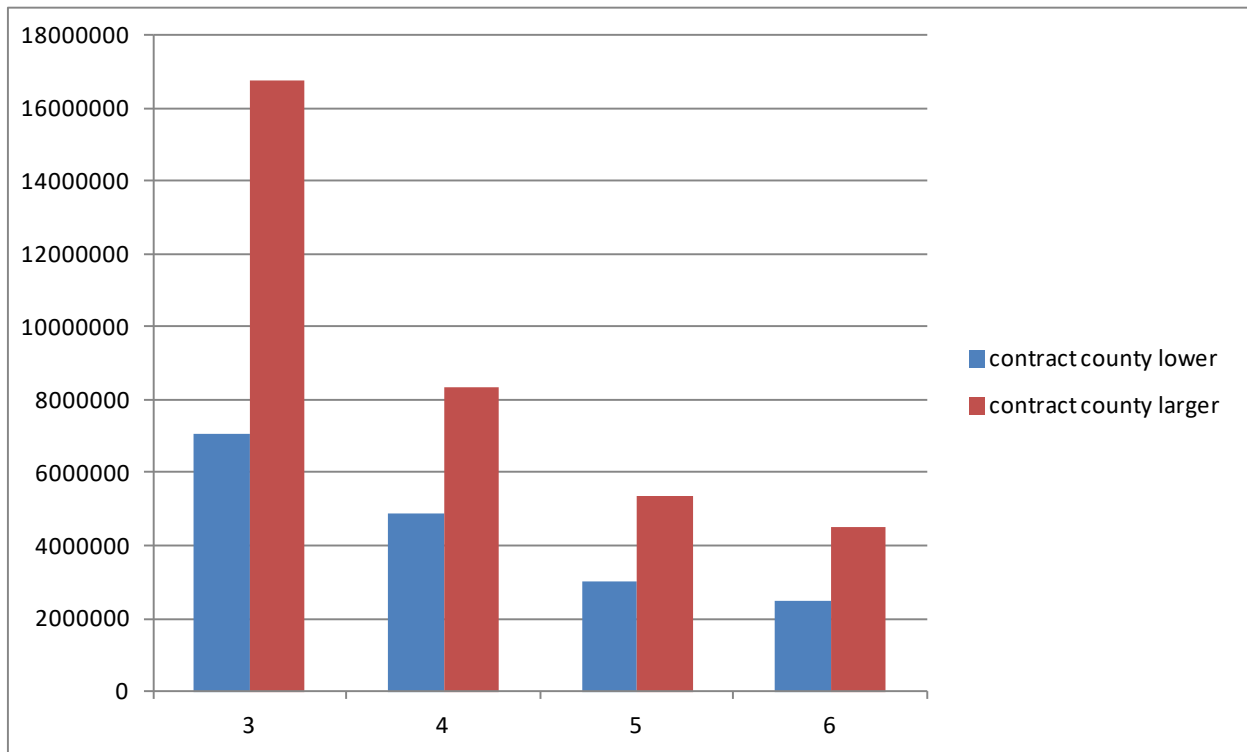


Figure 6: Comparing the magnitude of employment changes on contract locations, times, and NAICS code with all other recorded locations in the US. The bars' height shows the number of pairs in which the magnitude of employment change in the contract county is larger (red) or lower (blue) than some other county's. The x-axis breaks down the numbers by length of NAICS code.

Using the dataset from step 5, the locations of an awarded contract have a larger magnitude of employment increase when compared to all other US location (about 35 million out of the 52 million possible pairs). Figure 6 shows the result broken down by



the length of the NAICS code, and it can be seen that for three-digit NAICS code, the relative percentage between larger contract locations (red) and lower ones (blue) is the highest at 70.3%.

Negative Contract Modification

The 33 contracts studied in in this track of our research are listed in Table 1 together with the two labor metrics of interest. This analysis took into account three months preceding and three months following the contract change to determine its impact. Averages were computed for both of the three-month periods. The difference between those two averages, in percentage, is used for the rest of the analysis, both for absolute employment levels and location quotient.

Table 1: List of 33 instances of contract modifications. Aggregated data for each vendor as well as individual reductions by date are provided.

Name of Vendor / Cut Date	Percent change in employment level (the three months before the cut compared the three months after)	Percent change in employment location quotient (the three months before the cut compared the three months after)
AMC DEFENSE TECHNOLOGIES	-2.25%	-2.89%
9/10/2014	-2.69%	-3.47%
12/10/2014	-1.82%	-2.31%
ARNOLD DEFENSE AND ELECTRONICS	-18.09%	-17.38%
8/19/2015	-24.23%	-24.25%
7/15/2016	-12.63%	-11.27%
BARBER-NICHOLS INC.	-15.52%	-15.16%
9/28/2016	-15.52%	-15.16%
B-K MANUFACTURING CO.	1.29%	1.72%
6/3/2014	1.29%	1.72%
BOSTON SHIP REPAIR	-53.26%	-51.10%
12/5/2016	-53.26%	-51.10%
CD BIOSCIENCES INC	57.38%	4.10%
10/3/2014	57.38%	4.10%
CHAE&NAM UNIVERSE INC	-5.67%	-5.75%
8/23/2016	-5.67%	-5.75%
CORONET MACHINERY CORP.	-12.10%	-12.68%
9/25/2014	-12.10%	-12.68%
DUCWORKS	10.05%	7.95%
7/6/2016	10.05%	7.95%



EAG-LED	-8.24%	-10.24%
11/22/2016	-8.24%	-10.24%
FORM FIT AND FUNCTION	-5.98%	-6.59%
7/3/2013	-5.98%	-6.59%
G I INDUSTRIAL-MARINE	414.68%	1173.90%
9/16/2016	414.68%	1173.90%
G SYSTEMS LP	1.03%	-8.07%
9/23/2014	1.03%	-8.07%
H L TECHNOLOGIES	2.08%	0.48%
3/8/2013	2.08%	0.48%
LOGAN ENERGY CORP.	-2.98%	-4.69%
4/15/2015	-2.98%	-4.69%
MANTEC SERVICES INC.	0.64%	-3.28%
5/20/2013	0.64%	-3.28%
MOBILE TECHNICAL SYSTEMS	-46.62%	-47.53%
3/8/2016	-46.62%	-47.53%
OWL INTERNATIONAL INC.	-8.38%	-12.04%
8/8/2013	-8.38%	-12.04%
PACIFIC ENGINEERING INC.	-16.25%	-18.94%
4/21/2015	-16.25%	-18.94%
PELATRON	-8.53%	-10.75%
9/4/2013	-8.38%	-12.04%
5/6/2014	-8.67%	-9.57%
RIBCRAFT USA	-11.98%	-11.88%
11/5/2014	-11.43%	-10.61%
3/2/2017	-12.69%	-13.52%
SEA-LECT PLASTIC CORPORATION	129.60%	35.31%
11/20/2013	141.25%	37.46%
1/15/2015	98.89%	24.04%
1/14/2016	151.49%	45.60%
SUMMIT SOLUTIONS OF COLORADO	-6.68%	-6.88%
11/20/2014	-9.98%	-7.99%
3/26/2015	-3.39%	-5.77%
SWIFTSHIPS	-51.04%	-47.73%
8/30/2016	-51.04%	-47.73%
SYSTEMS APPLICATIONS&SOLUTIO	3.39%	-1.04%
5/12/2014	3.39%	-1.04%
WHITAKER TECHNOLOGIES	-5.67%	-8.84%
6/10/2014	-5.67%	-8.84%



In a clear majority of the cases included in the study, a reduction in a contract is correlated with a drop in employment indicators, when comparing the average employment numbers the quarter (three months) before the reduction date with the quarter after. The absolute employment level was reduced in 22 out of 33 instances, or 66.6% of time. Similarly, the location quotient displays a drop in 25 out of 33 cases (75.7%). Figure 3 shows an overview of the two metrics. The data has a few outliers, especially on the positive side of the change, where the location quotient increased significantly in one instance.

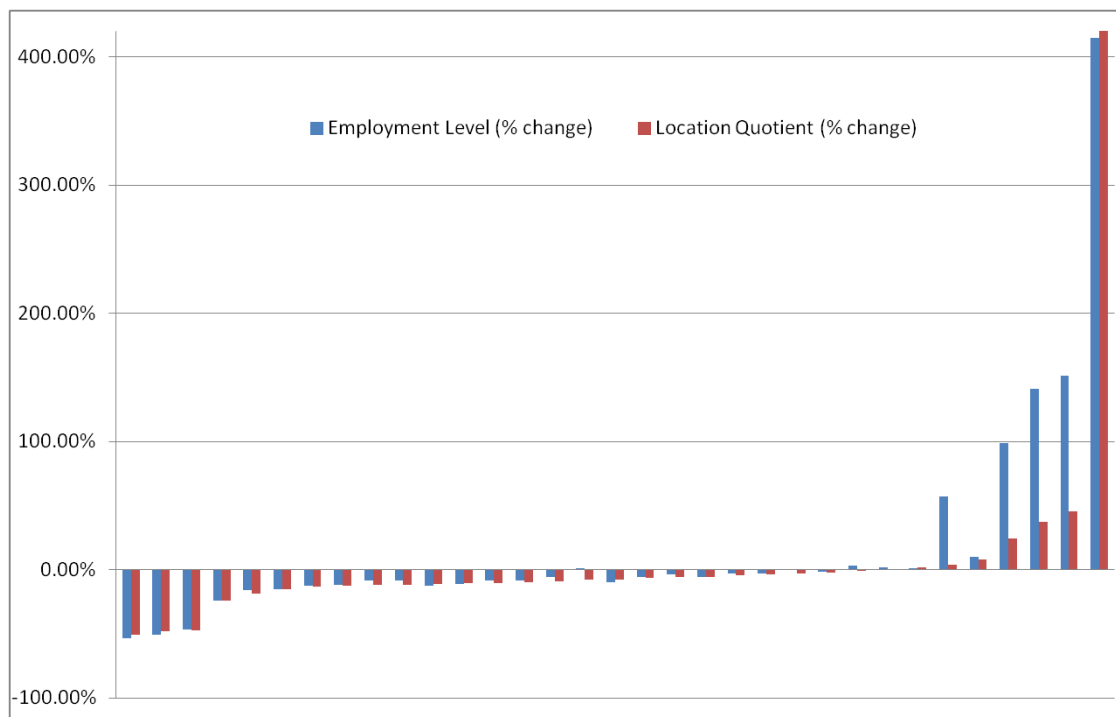


Figure 7: Percentage change in employment level (blue) and location quotient (red) before and after a reduction in a federal contract. The x-axis lists all 33 cases of contract reduction and is ordered by location quotient. The right-most red bar is not shown in its entirety as it would reach up to almost 1,200%.

The study also examined the correlation between the magnitude of the reduction and the magnitude of the change in employment indicators, but in the limited cases (33) that were available, the correlation was relatively weak. Even when removing positive outliers (the right side of Figure 7), R^2 was about 0.3 for both employment level and quotient. For the revenue of a vendor to change in employment, the correlation was only marginally better ($R^2 \approx 0.4$).

THIS PAGE LEFT INTENTIONALLY BLANK



Conclusion and Future Work

Our research employed large data analysis and found that the magnitude of employment changes is higher in 70% of places of performance for a Federal award as compared to the rest of the country for three-digit NAICS codes. This finding is promising as a metric to help uncover hidden contractors because it has the potential to eliminate more than two-thirds of US locations. It can be used repeatedly for many contracts that use products and services in related industries to assign a probability for various potential locations of undisclosed subcontractors. Such undisclosed contractors are typically employed in contracts with a larger scope to achieve confidentiality, security, or a competitive advantage. Depending on the situation, acquisition experts may need additional planning to protect such hidden contractors if security is desired, or may rely on data science to identify these contractors and avoid them becoming a weak link in the acquisition process.

We also found that large reductions in federal contracts are correlated in a majority of cases (66% or 75%, depending on the metric used) to drops in employment in a given region and industry. This finding shows that it is possible to determine the location of an undisclosed contractor by examining public employment data at the times when large contracts are reduced or simply reach the end of their period.

The main contributions of this work are the development of a data science process for joining large acquisition and employment datasets and the development of a potential metric based on using the magnitude of employment changes in three-digit NAICS code industry. Future work will focus on three main thrusts:

- a) running simulations where one contract is picked at a time, and the most likely locations of that contract are determined based on the three-digit NAICS code metric presented in this report;
- b) considering additional years, especially one where the employment is declining, which would be the opposite of 2016, a good year for employment; and
- c) determining additional probability metrics for possible locations for contractors. Good candidates may be combinations of NAICS codes and location quotients (LQ, 2018).



THIS PAGE LEFT INTENTIONALLY BLANK



References

- Augustine, Norman R. (1997) Augustine's laws. AIAA.
- Cilli, M., Parnell, G. S., Cloutier, R., & Zigh, T. (2015) "A Systems Engineering Perspective on the Revised Defense Acquisition System." *Systems Engineering*, vol. 18, no. 6, p. 584–603.
- Felten, E. W., Robinson, D., Yu, H., & Zeller, W. P. (2009). Government Data and the Invisible Hand. *Yale Journal of Law and Technology*, 160
- Gaither, Carl C. (2014) "Incorporating Market Based Decision Making Processes in Defense Acquisitions." *International Journal of Defense Acquisition Management*, vol. 6, p. 38–50.
- Gayo, J. E., & Pablos, P.O., & Rodriguez, J.M., & Vafopoulos, M. (2013). Towards a stepwise method for unifying and reconciling corporate names in public contracts metadata. The CORFU technique. *In Proceedings of the 7th Metadata and Semantics Research Conference Track on Metadata and Semantics for Open Repositories, Research Information Systems and Data Infrastructures*, Thessaloniki, Greece.
- Houseman, S., & Bernhardt, A. (2017). Memorandum: Data Needs for Research on Domestic Outsourcing in the United States. Retrieved from https://research.upjohn.org/staff_publications/1.
- Krzysko, M. & Baney B., (2012) "The Need for "Acquisition Visibility"." *Journal of Software Technology* 15-1, p. 4-9.
- Llorens, L., & Rodriguez, J.M., & Vafopoulos, M. (2015). Enabling policy making processes by unifying and reconciling corporate names in public procurement data. The CORFU technique (Vol. 41, pp. 28-38). Madrid, Spain. *Computer Standards & Interfaces*.
- Location Quotient, LQ (2008). *Bureau of Economic Analysis*. Retrieved from https://www.bea.gov/faq/index.cfm?faq_id=478.
- Miller, A., Risch, M., & Wilking, E. (2017). Independent contractor or employee? The changing relationship between firms and their workforce and potential consequences for the U.S. income tax.
- NAICS (2017). North American Industry Classification System. *US Office of Management and Budget*. Retrieved from 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. *In Proceedings of the 15th Annual Acquisition Research Symposium*, Vol. 2, p. 491-499, Monterey, CA.

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). Monterey, CA: Naval Postgraduate School. September.





ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

WWW.ACQUISITIONRESEARCH.NET