



1

Utilizing Public Data for Data Enhancement and Analysis of Federal Acquisition Data

NPS 2018, Panel 19

Ningning Wu, Richard Wang, Mihail Tudoreanu

Monterey, California

Thursday, May 10, 2019





Bottom Line Up Front (BLUF)

- Small contractors tend to be less robust and easier to fail when facing disasters.
- More than a hundred of subcontractors were located nearby an epicenter of 7.0-magnitude quake.
- Hidden subcontractors could be identified through federal funding cut and higher unemployment







May 10, 2018 8:12 am ET



Goals of Our Study

- Data profiling and quality assessment of sample acquisition database
- Leverage web data for
 - quality improvement of acquisition data
 - crystalizing the needed information to enhance data analytics capabilities
- Apply data analytics on acquisition data to discover patterns that help identify the room for improvement in acquisition management



Data Sources

- Usaspending.gov
 - Contains spending information of all US Departments between the year 2000 to present for a selected state or all states.
 - The data can be downloaded in different formats, such as CSV, TSV, and XML.
 - Spending data are further categorized under prime award and sub-award. The types of spending include contracts, grants, loans, and other financial assistance.
 - Our downloaded data contains 47GB data in total, covers the DoD budget between 2000-2017 including each type of spending data for both prime award and sub-award.





USASpending Data

• Data is organized into four tables

Table Name	RecCnt	ColCnt	CompleteCols/ SingleValCols	EmtpyCols	IncompleteCols
PrimeContracts	23,677,787	212	50/1	0	162
SubContracts	395,569	101	41/0	3	57
PrimGrants	202,166	67	32/5	2	33
SubAGrants	11,115	101	29/4	25	47





Research Methodology

- Assess the quality of the sample acquisition database in terms of accuracy, field length consistency, and completeness.
- Identify the fields for quality improvement.
 - The primary objective is to investigate the feasibility of leveraging the information online for both quality assessment and improvement.
- Apply appropriate data analytics methods to discover useful patterns from the data.
- Utilize online data to complement the information of the sample database.
 - The primary objective is to research appropriate text mining methods to retrieve the information for the purpose of advanced data analytics.





Data Enhancement and Analytics System Framework







Quality Assessment

- We classify attributes into two categories
 - Identity attributes: attributes that provide identity information for a contract/contractor and agency including
 - Project identifier, contractor identifier such as DUNS number, Business name, Address information, phone, fax
 - Non-identity attributes: attributes that do not provide identity informaiton
- Quality assessment focuses on three dimensions: accuracy, completeness, and field length consistency.

Our preliminary study focuses on identity attributes only





Quality Assessment --Completeness

• PrimeContracts Table







Quality Assessment –Completeness (cont.)

• SubContracts Table



Completeness





Quality Assessment --Length Distribution

PrimeContracts Table







Quality Assessment --Length Distribution (cont.)



17, 19)



Accuracy

- Data is accurate if it is free of error and conforms to gold standards of data.
 - Accessing accuracy requires the knowledge of correct data and needs to compare each data item with the known correct value
- For proof of concept, the accuracy assessment of this study focuses on DUNS numbers first.
 - The first phase focuses on Duns numbers other than 9 digit long.
 - Duns & Bradstreet database provides free service that allows user to search a business' DUNS number based on its name and address. However, the search result is delivered by email.
 - Address information is *critical* as each branch of a business has their own DUNS.





Improving Accuracy on DUNS Number of PrimeContracts

There are 26 records in *PrimeContracts* table that are ten-digit long.

• A closer study on those DUNS numbers revealed that most of them are actually the phone numbers; most records have misplaced addresses and business names.

ID	dunsnumber	multipleorsingleawardidc	vendorname	vendorenabled	vendorlocationdisablef	streetaddr	streetaddress2	streetaddress3	city	state	zipcode	vendoraltern
					lag	ess						atesitecode
1	2022611902	CB&I FEDERAL SERVICES LLC			2370 TOWNE CENTER B	LVD	BATON ROUGE	A	708068172	UNITED ST	LA	386491765
2	2082332929	SUNDANCE-TLI			275 S 5TH AVE_STE 21	5	POCATELLO	D	832016429	UNITED ST	ID	967391967
3	2164515588	LUCAS PRECISION_LIMITED	LUCAS PRECISION_LIN	IITED	13020 SAINT CLAIR AVE		CLEVELAND	он	441082033	UNITED ST	ОН	622006591
4	2536803243	U.S. OIL TRADING LLC			3001 MARSHALL AVE		TACOMA	WA	984213116	UNITED ST	WA	784187226
5	3053925669	WORLD FUEL SERVICES FL		9800 NW 41ST ST STE	400	MIAMI	FL	331782980	USA: UNITED	FL	25	3054288000
6	3256738838	AVFUEL CORPORATION			47 W ELLSWORTH RD		ANN ARBOR	мі	481082206	UNITED ST	MI	131423808
7	4103792800	MABEY BRIDGE & SHORE_ IN	С		6770 DORSEY ROAD		ELKRIDGE	MD	210756205	UNITED ST	MD	296358146
8	4106942749			1580A W NURSERY RE		LINTHICUM	MD	210902202	USA: UNITED	MD	2	8004439219
9	4155675899	UNITED BLOOD SERVICES		6210 E OAK ST		SCOTTSDALE	AZ	852571101	USA: UNITED	AZ	9	4159010740
10	4794524727	GREENSCAPES TOTAL LAWNO	AR		3118 S 64TH CIR		FORT SMITH	AR	729034971	UNITED ST	AR	799080705
11	4806000407			2423 W MINERAL RD		PHOENIX	AZ	850 <mark>419559</mark>	USA: UNITED	AZ	7	4806000407
12	6034311331	HECKLER & KOCH DEFENSE IN	1		19980 HIGHLAND VIST	DR S	ASHBURN	VA	201474189	UNITED ST	VA	166031588
13	6103858200	STV INCORPORATED			205 W WELSH DR		DOUGLASSVILLE	PA	195188713	UNITED ST	PA	106768252
14	6192388341			3589 DALBERGIA ST		SAN DIEGO	CA	921132123	USA: UNITED	CA	51	6192388338
15	7038491000			6200 GUARDIAN GAT	EWAY DR	ABERDEEN F	MD	210051327	USA: UNITED	MD	2	7036413735
16	7578734959	AH-BC NAVY JV_ A JOINT VE			804 OMNI BLVD STE 20	1	NEWPORT NEWS	VA	236064422	UNITED ST	VA	830647272
17	7736371666	DEHLER MANUFACTURING C	D		5801 W DICKENS AVE		CHICAGO	L	606394030	UNITED ST	IL	5069661
18	7818717449	NOBLE SUPPLY AND LOGISTIC	-	302 WEYMOUTH ST		ROCKLAND	МА	23701171	USA: UNITED	MA	9	7818711911
19	8044847840			12650 E ARAPAHOE R	D BLDG	ENGLEWOO	со	801123901	USA: UNITED	СО	6	8044847839
20	8082451911	SENTER PETROLEUM INC			3011 AUKELE ST STE C		LIHUE	н	967661465	UNITED ST	HI	606679520
21	8085229712	ALOHA PETROLEUM LTD	ALOHA PETROLEUM L	TD	1132 BISHOP STREET ST	E 17	HONOLULU	н	968132807	UNITED ST	HI	81909046
22	8088335825			707 KAKOI ST		HONOLULU	н	968192017	USA: UNITED	HI	1	8088361957
23	8177636775	LOCKHEED MARTIN CORPORA	LOCKHEED MARTIN CO	ORPORATI	1 LOCKHEED BLVD		FORT WORTH	тх	761083619	UNITED ST	ТХ	834951691
24	9036650030	TIGER LAWN AND LANDSCAP	E		1389 MCR 3222		EFFERSON	ТХ	756575878	UNITED ST	ТХ	969073886
25	9184266191			209 SW 7TH ST		ANTLERS	ок	745233834	USA: UNITED	OK	2	9184262871
26	9519409686	PERRIS WIND TUNNEL		2093 GOETZ RD		PERRIS	ĊĂ	925709315	USA: UNITED	CA	41	9518057728





Approach

 The research focuses on fixing incorrect DUNS number based on the online Duns & Bradstreet database, which contains DUNS numbers of 285 million commercial entities and 100 million associated contacts.

Dun & Bradstreet D-U-N-S® Number Lookup			S	Advanced Search Search by one of the three options below:				
Look up a partner	's company or find your	company's D-U-N-S Number.	Search By Company Name	Search By Telephone Number	Search By D-U-N-S Number			
Search For	thereen		Company Name*:					
	ther company		Address:					
Business N	lame	Business Phone	City:					
Business Name*	Name		State*: State		*			
Street & Suite	Street	Suite	Zip Code: Country*: United States of America		v			
City & Zip Code	City	Zip Code		Search Now				
State & Country*	State	✓ U.S.	*Required Field					





Extracting the Missing DUNS Number – Example 1

Dunsnumber	Vendorname	Zip	State	
<mark>2164515588</mark>	LUCAS PRECISION_LIMITED	441082033	он	

Query Duns&Bradstreet Database for DUNS number by Company Name

ı k	un & bradstreet
	03/25/2018
	DUNS number: 079699253
	credit building solution.



Dunsnumber	Vendorname	Zip	State	
079699253	LUCAS PRECISION_LIMITED	441082033	он	

• The figure shows an example of using *Duns&Bradstreet* database to retrieve the DUNS number for the business *LUCAS PRECISION LIMITED,* which only has a *single* duns number.





Extracting the Missing DUNS Number – Example 2

	Dunsnumber	Vendorname	Street	Zip	City	State	PhoneNo]	
	8085229712	ALOHA PETROLEUM LTD	ALOHA PETROLEUM LTD	1132 BISHOP STREET STE 1700	968132807	Honolulu	н	8085229712]
	8177636775	LOCKHEED MARTIN CORPORATI	1 LOCKHEED BLVD	761083619	FORT WORTH	тх	8177636775]	
]	
			Ĺ		tify the be ins & Brad	st mat Istreet	ch record i database	n		
ALOHA PETROLEUM LLC 1112 Bishop St Ste 1201 More	ALL NE 04813-3930	Rea.	nda Artica	6253 Trail Lake	Dr. Fort Worth, T	x 76133-34	5	Br	anch	
(808) 522-9789	nen, ra 70012-2020			(301) 897-6000						
ALOHA PETROLEUM LTD 2314 N School St, Honolulu, H	1 96819-2540	Bra	nch Active	LOCKHEED N 1029 Lindstrom (817) 763-3035	AARTIN CORPO	RATION DX 76131-53	12	в	andh	
ALOHA PETROLEUM LTD										
4339 Waialae Ave, Honolulu, H	196816-5710	Bra	nch Active	LOCKHEED N	ARTIN CORPO	RATION				
(808) 734-5450				1 Lockheed Bh (817) 777-2000	vd Bidg 10, Fort V	Jorth, TX 76	108-3619	Br	andh	

Query the database for DUNS number

Dunsnumber	Vendorname	Street	Zip	City	State	PhoneNo	
080385182	ALOHA PETROLEUM LTD	1132 BISHOP STREET STE 1700	968132807	Honolulu	н	8085229712	
114945988	LOCKHEED MARTIN CORPORATI	1 LOCKHEED BLVD	761083619	FORT WORTH	тх	8177636775	

- The figure shows an example of extracting the DUNS numbers of companies that have multiple DUNS numbers.
 - The vendor name, address, zip code, and telephone numbers are used to identify the highlighted best match.
 - Then the corresponding DUNS number is retrieved.





Filling the Missing DUNS Number

ID	dunsnumber(old)	dunsnumber(new)	vendorname	Street	city	state	zip	tel
1	2022611902	079811713	CB*I Mahan LLC	2370 Towne Center Blvd	Baton Rouge	LA	70806-8172	(225)932-6500
2	2082332929	967391967*	sundance-tli	305 N 3rd Ave Ste B	pocatello	ID	83201-6306	(208) 233-2929
3	2164515588	079699253	LUCAS PRECISION_LIMITED	13020 Saint Clair Ave	Cleveland	OH	44108-2033	(216) 451-5588
4	2536803243	784187226	U.S. OIL TRADING LLC	3001 Marshall Ave	Tacoma	WA	98421-3116	(253) 383-1651
5	3053925669	131504342	WORLD FUEL SERVICES FL	9800 Nw 41st St Ste 400	Miami	FL	33178-2980	(305) 428-8000
6	3256738838	020829396	AVFUEL CORPORATION	47 W Ellsworth Rd	Ann Arbor	MI	48108-2206	(734) 663-6466
7	4103792800	171902265	MABEY INC	6770 Dorsey Rd	Elkridge	MD	MD	(410) 379-2800
8	4106942749	005128988*	Northrop Grumman Systems Corporation	1580a W Nursery Rd	Linthicum Heights	MD	21090-2202	(410) 765-1000
9	4155675899	006902498	BLOOD SYSTEMS INC	6210 E Oak St	Scottsdale	AZ	85257-1101	(480) 946-4201
10	4794524727	799080705	GREENSCAPES TOTAL LAWNCARE& LANDSCAPING SE	3118 S 64th Cir	Fort Smith	AR	72903-4971	(479) 452-4727
11	4806000407	363821294*	ORBIT INDUSTRIAL SERVICE & MAINTENANCE	5316 W Missouri Ave	Glendale	AZ	85301-6006	(480) 704-4849
12	6034311331	134466999	HECKLER & KOCH DEFENSE INC.	19980 Highland Vista Dr Ste 190	Ashburn	VA	20147-4189	(703) 450-1900
13	6103858200	059946819	STV INCORPORATED	205 W Welsh Dr	Douglassville	PA	19518-8713	(610) 385-8200
14	6192388341	080090672	SOUTHBAY SANDBLASTING & TANK CLEANING INC	3589 Dalbergia St	San Diego	CA	92113-3810	(619) 238-8338
15	7038491000	827488979	ABERDEEN PROVING GROUND US ARM	6200 Guardian Gtwy	Aberdeen Proving Ground	MD	21005-1327	(410) 273-2640
16	7578734959	830647272*	AH/BC NAVY JV LLC	11837 Rock Landing Dr Ste 300	Newport News	VA	23606-4493	(757) 873-4959
17	7736371666	059530805	DEHLER MANUFACTURING COM	5801 W Dickens Ave	Chicago	IL	60639-4030	(773) 637-0615
18	7818717449	107910259	NOBLE SALES CO. INC	302 Weymouth St	Rockland	MA	02370-1171	(781) 871-1911
19	8044847840	079427300	PERFORMANCE FOOD GROUP	12650 E Arapahoe Rd	Englewood	CO	80112-3901	(303) 662-7100
20	8082451911	069891588	SENTER PETROLEUM INC	3011 Aukele St Ste C	Lihue	HI	96766-1430	(808) 245-1911
21	8085229712	080385182	ALOHA PETROLEUM LTD	1132 Bishop St Ste 1700	Honolulu	HI	96813-2820	(808) 522-9789
22	8088335825	007050586	GARLOW PETROLEUM	707 Kakoi St	Honolulu	HI	96819-2017	(808) 836-1957
23	8177636775	008016958	Lockheed Martin Corporation	1 Lockheed Blvd Bldg 10	Fort Worth	TX	76108-3619	(817) 777-2000
24	9036650030	969073886	CHRIS GIBBONS	1389 Mcr 3222	Jefferson	TX	75657	(903) 665-8190
25	9184266191	619338010	CHOCTAW MANUFACTURING DEFENSE CONTRACTO	209 Sw 7th St	Antlers	OK	74523-3834	(580) 298-2203
26	9519409686	141883517	PERRIS SKYVENTURE	2093 Goetz Rd	Perris	CA	92570-9315	(951) 940-4290





Filling the Missing DUNS Number (cont.)

ID	dunsnumber (old)	dunsnumber (new)	vendorname	Street	city	state	zip	tel
2	2082332929	607564825	Sundance Consulting inc	305 N 3rd Ave Ste B	Pocatello	ID	83201-6306	(208) 233-2929
			SUNDANCE-TLI	420 S 4th Avenue	Pocatello	ID	83201	(208)233-2929
16	7578734959	830647272	AH/BC NAVY JV LLC	11837 Rock Landing Dr Ste 300	Newport News	VA	23606-4493	(757) 873-4959
				804 Omni Blvd Ste 201	Newport News	VA	236064422	(757) 873-4959



Data Mining

- Data mining is the process of examining large data sets to uncover hidden but interesting patterns
 - unknown correlations, market trends, customer preferences and other useful business information
- The analytical findings can shed significant insights to help add perspective to use the data and to lead to more effective decision makings.
- Some major data mining techniques include association discovery, classification, clustering, regression, sequence or path analysis, and structure and network analysis.





Data Mining

- Association discovery aims to find frequent patterns that represent the inherent regularities in the datasets.
 - Association, correlation, and causality analysis, basket data analysis, cross-marketing, etc.
- Classification, also called supervised learning, is the task of inferring a function from labeled training dataset. The function can then be used to classify new data instances.
 - Decision tree, Bayesian networks, support vector machine, and neural networks are some of the commonly used models for classification.
- Clustering, also called non-supervised learning, group a collection of data objects into groups according a predefined distance function.
 - Clustering can be employed as a stand-alone tool to get insights about data or as a preprocessing tool for other algorithms.
- Sequence analysis discovers patterns among sequences of ordered events or elements. Application of sequence patterns include customer shopping sequence, DNA sequences and gene structures, sequences of stock market changes, etc.
- Graph and network analysis aims to discover frequent subgraphs, trees, or substructures. It has been used for social networks analysis and web mining.



Cluster Analysis

 Network analysis is performed on prime contractors and their subcontractors of the sample database







Cluster Analysis (cont.)









Cluster Analysis (cont.)





(a) Relationship between contractors and subcontractors by State

(b) Relationship of contractors of different business types





Pattern Discovery – Finding 1

- Small contractors are usually less robust and easier to fail compared to large contractors when facing natural or man-made disasters.
 - PTB is a small, single-location company with less than 200 employees.
 - It was involved in 6 different projects led by some key primary contractors including *Boeing, Lockheed Martin,* and *L-3 Communications*.
 - The average award amount is about \$5400.
 - Company website: <u>http://www.precision-tube-bending.com/index.html</u>





PRECISION TUBE BENDING Website Screen Capture



About PTB

⊐ Home ❑ Customers-Approvals ⇒ Precision Tube Bending ⊐ Aircraft Tube Bending ⇒ Aerospace Tube Bending □ CNC Machining ⇒ Precision Tube & Duct Welding Tube & Duct Cleaning □ ISO 9001:2015 - AS 9100D Tube Bending Engineering Tooling & Assembly Fixtures ⊐ Materials Experience Transport - Packaging - Kitting





PRECISION TUBE BENDING

Welcome! The Precision Tube Bending Team of experienced and dedicated employees is ready to serve you for all of your tube bending and tube fabrication needs. Let us demonstrate what we can do to earn vour business

Precision Tube Bending is well known throughout the industry for excelling in "tough" tube bending applications such as:





PRECISION TUBE BENDING

Contact Us



PTB Capabilties Request For Quotation

⊐ Home □ Customers-Approvals ⊐ Precision Tube Bending ⇒ Aircraft Tube Bending ⇒ Aerospace Tube Bending ⊐ CNC Machining ⊐ Precision Tube & Duct Welding ☐ Tube & Duct Cleaning □ ISO 9001:2015 - AS 9100D ⊐ Tube Bending Engineering ⊐ Tooling & Assembly Fixtures ⊃ Materials Experience Transport - Packaging - Kitting





Welcome! The Precision Tube Bending Team of experienced and dedicated employees is ready to serve you for all of your tube bending and tube fabrication needs. Let us demonstrate what we can do to earn vour business

Precision Tube Bending is well known throughout the industry for excelling in "tough" tube bending applications such as:



PRECISION TUBE BENDING

Request For Quotation

Contact Us



Welcome! The Precision Tube Bending Team of experienced and dedicated employees is ready to serve you for all of your tube bending and tube fabrication needs. Let us demonstrate what we can do to earn

Precision Tube Bending is well known throughout the industry for excelling in "tough" tube bending applications such as:





Pattern Discovery – Finding 2

- We retrieved locations of 7.0-magnitude quake epicenters in US from U.S. Geological Survey website <u>www.usgs.gov</u>, and identified 118 subcontractors in *SubContracts* table located nearby an epicenter.
 - 984 awards have at least 1 subcontractor located in the high risk areas.
 - 41 of them have at least 2 subcontractors located in the high risk areas.
 - Table below shows the top 5 contracts with the most number of subcontractors in high risk earth quake areas.

Project ID	#Subcontractors
1	15
2	15
3	11
4	10
5	8



Pattern Discovery -- Finding 3

- <u>Small privately owned company</u> making RF and microwave filters, multiplexers, and sub-assemblies for the military market in the United States and internationally
- Subcontractor on 8 projects for 2 different contractors: *Boeing* and *Lockheed Martin*.
- Single location with less than 50 employees
- Company website: <u>http://www.rsmicro.com/compa</u> <u>ny-information/about</u>



Everyone in the company is certified to ESD standards. Our manufacturing system is LEAN. We perform environmenta



Inferring Undisclosed Subcontractors from Public Contract Records and Employment Data

- We did a preliminary exploration of whether a correlation exists between procurement data and employment information from *Bureau of Labor Statistics (BLS) Employment Data*.
- The result shows 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.

Sample Bureau of Labor Statistics (BLS) Employment



Sample industry bundle for Essex, MA, fourth quarter 2014

INSTITUTE FOR CHIEF DATA OFFICERS





Sample Instances of Contract Modification

Name of Vendor / Cut Date	Percent change in employment	Percent change in employment
	level (the three months before	location quotient (the three
	the cut compared the three	months before the cut compared
	months after)	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%

Percentage Change in Employment Level and Location Quotient Before and After a Reduction



ATA OFFICERS





Thank You