

Educational Infrastructure for Data-Driven Performance: Chief Data Officers in the Big Data Era

Lawrence Whitman, Richard Wang, Mihail Tudoreanu, Ningning Wu
University of Arkansas at Little Rock
Little Rock, AR

Abstract

High-quality data and data analytics are two inter-related areas in the era of big data. Leading organizations are rapidly appointing Chief Data/Digital Officers (CDO) and Chief Analytic Officers (CAO) for competitive advantages and strategic necessities. Gartner Inc. reported that more than one thousand CDO/CAO positions have been established as of 2017, a tenfold increase from 2013; and the trend is growing steadily in both public and private sectors. As such, it is critical to provide training and education, which are based on rigorous and relevant research. The Institute for Chief Data Officers (iCDO) established by the State Department of Higher Education in 2016, anchored in the University of Arkansas at Little Rock, was conceived to address these critical issues. In this paper, we report our progress in the following areas: CDO training programs and on research using publicly available federal data to improve defense acquisition policies, contract, and procurement processes.

Keywords

Data Quality, Chief Data Officer, Data Analytics

1. Introduction

In this article, we present the role of the Chief Data Officer (CDO) and how a university can develop the body of knowledge and conduct research in the public and private sectors. To gain competitive advantage, organizations are increasingly deploying data analytics encompassing a wide range of techniques such as artificial intelligence and machine learning (AI/ML), statistical learning, and information visualization. To perform these tasks however, organizations must meet the challenge of big data, with its focus on volume, velocity, variety, value and veracity.

The Institute for Chief Data Officers (iCDO) seeks to bring together knowledge from three sources: big data practitioners, research on data analytics and quality, and higher education methods to address the following key areas of activity:

- CDO Certificate Program and Data-Driven Performance Management (DDPM) Training Program
- Seminal research such as the one highlighted in this article that employs publicly available federal data such as those from usaspending.gov and gsa.gov to improve defense acquisition policies, contract, and procurement processes

2. CDO's Roles and Responsibilities

Organizations need to determine who should manage big data. Data scientists and data engineers have emerged to capitalize on the analytical opportunities of big data, but placing these specialists in operational business units without leadership at the corporate level are necessary, but not sufficient, to harness the full value of big data. To address the challenges and opportunities of big data, leading organizations have established a new breed of executive, the CDO.

Wikipedia describes the CDO role as including: "... defining strategic priorities for the company in the area of data systems, identifying new business opportunities pertaining to data, optimizing revenue generation through data, and generally representing data as a strategic business asset at the executive table." In reality, although some CDOs

strive to exploit big data for business strategy, others focus solely on data preparation for external reports, overseeing compliance and establishing data governance.

Leading organizations have learned an important lesson—that seemingly tedious data problems are critical for data analytics. Data problems can reflect weaknesses in business strategy and operations. Traditionally, organizations have addressed data problems by assigning a small group within the IT department to clean up data. As it has become evident that data problems, particularly business problems rooted in data problems, cannot be solved by the IT group alone, organizations have appointed data managers with a variety of titles, such as data quality managers, data quality analysts and data stewards. Data-governance mechanisms, committees, councils and workgroups have also been developed to identify and solve data-related problems and resolve conflicts. Finally, enterprise architecture and data architecture have also been employed to align data, IT, and business processes and strategies. Despite these efforts, organizations have continued to face data issues, and their ongoing concerns have led a growing number to establish an enterprise-level, executive-rank CDO. Some might argue that traditional data-related managers and data-governance mechanisms can deliver the same results as a CDO. However, there are critical differences between the efforts of low-level data managers and those of executive-rank CDOs. The key contrast lies in organizationally sanctioned leadership and the accountability given to executive-level CDOs [1]. In the article published in the *MIS Quarterly for Executives*, Lee, et al. further indicated that a CDO is different than a data manager and leads the charge in establishing capabilities for the extended enterprise. A key issue is determining who owns and has authority for the data itself (who is the data owner?). Data managers are severely restricted in this regard, and the CDO is generally granted this authority. Therefore, a CDO can be held responsible for resolving data issues. Governance issues between and across heterogeneous organizations can be managed by a CDO, who therefore also can assume responsibility. However, a CDO cannot replace a data manager; rather, a CDO should lead data managers in exploiting data in a manner that provides the most value to the overall organization and not just a single sub-component of the enterprise. The primary intent is for data itself to provide value, not simply an output of the enterprise which must be managed [2].

3. Training Programs

Workforce development is an important focus and addresses the need to train professionals in our state (Arkansas) in the areas of data management and data science. The goal is to provide training at higher managerial levels for employees who supervise teams or projects, and who must understand how to extract the most value from the data available to an enterprise. Expanding on the Institute's for Chief Data Officer's expertise on providing certification for executives of a company, this work adapts CDO training to the equivalent mid-level management for a corporation. Such training can also benefit governmental mid-level management and academic principal investigators working on a typical research project. In cooperation with local companies, the training program also provides additional certification in technical and basic aspects of data science and data analytics. The development of the certificate is iterative and consists of the following four phases:

- Curriculum development for a 3-4 day certificate program. Also, adaptation of real-world problems for 60-90 day apprenticeship both at local enterprises and University.
- Delivery of the certification, and supervision of the apprentices. To support a heterogeneous class, iCDO seeks to include both start-up ventures, established corporations, government and academic students in each class.
- Assessment of the program by collecting data about the training and seeking input for trainees and their supervisors.
- Integration of best practices and theories by having at least one data expert from outside state of Arkansas provide input and participate in the certification sessions, by encouraging continuous professional development of the instructors to stay current with the latest research, and by maintaining regular contact with the training divisions of local enterprises.

4. Defense Acquisition Research

Defense acquisition provides the perfect sandbox for research into the various roles the CDO must play in an organization. Defense acquisition is one of the most complex enterprises taking place today, with hundreds of thousands of individual contracts and grants that are performed by tens of thousands of contractors and service providers. Contracts and awards from the federal government, which is almost \$700 billion in 2019, involving both public and private companies, from multi-billion dollar contractors to individuals, to other federal agencies. While some contracts involve only two parties, others are performed jointly by many providers, and yet others involve a

hierarchical structure of subcontractors performing part of the work without the explicit knowledge of the DoD or other contractors.

Defense acquisition can be approached from multiple angles involving performance management, and this team considered research into exploiting data external to the enterprise (i.e., external to DoD acquisition) to better perform the mission of the enterprise. Our approach is to investigate appropriate data science techniques for improving the quality of federal acquisition data as well as discovering useful patterns that can further acquisition management and planning research. The research examines the feasibility of leveraging the information on the Internet for verification and validation of acquisition data. Utilizing online information faces several challenges. One of the key challenges is how to find the information that is credible and accurate from often an enormous number of unstructured documents returned by a search. In addition, our research investigates how public data from government documents and reports, webpages, online forums, and social medium, when connected together, can be utilized for inferring potential sensitive information on acquisition projects. The key outcome is to identify areas of improvement for defense acquisition policies and procurement process. Some findings are highlighted as follows:

- Small contractors tend to be less robust and more likely to fail when facing disasters.
- More than a hundred subcontractors were located near an epicenter of 7.0-magnitude quake. Some projects have more than a dozen subcontractors in such high-risk areas.
- Hidden subcontractors could be identified through federal funding cuts and higher unemployment.

This effort [3-12] is based on Federal Acquisition databases at USAspending.gov, which contains spending information of all US Departments between the years 2000 to 2018. 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. We set up a custom-made database system to host the data.

Figure 1 shows the framework of the Data Enhancement and Analytics System. The system has four major components, namely Quality Assessment engine (QA), Data Cleaning (DC) Engine, Data Enhancement and Analytics engine (DEA), and Text Retrieval and Analysis engine (TRA). The key component is Text Retrieval and Analysis engine as it supports the functionalities of the remaining three components. TRA is responsible for four tasks: 1) performing searches on the Internet, 2) identifying the websites that contain the most reliable data, 3) extracting the desired information from the text; and 4) fusing together information by collectively integrating data from multiple sources. When information for quality assessment and data cleaning is not available, TRA will search and extract the needed information online. Data Analytics and Enhancement engine aims to enhance our knowledge about data by discovering hidden and interesting patterns in the data as well as complementing the internal data with the information that is not found in the database, but is potentially useful for advanced data analytics.

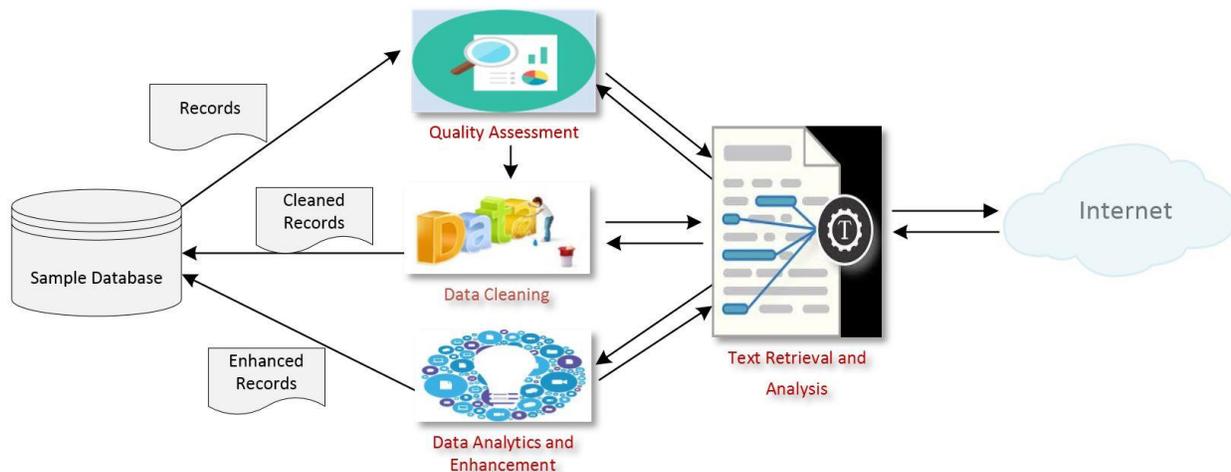


Figure 1: Framework of Data Enhancement and Analytics System

4. Research Method

The research has the following tasks:

- Assess the quality of the sample acquisition database in terms of accuracy, field length consistency, and completeness.
- Identify the fields for quality improvement.
- Apply appropriate data analytics methods to discover useful patterns that can help acquisition research community to design strategies for effective project planning and management.
- Research appropriate text mining methods to retrieve the information for the purpose of advanced data analytics.

Our preliminary findings are summarized as follows:

- It is feasible to leverage the information online for both quality assessment and improvement of acquisition database. Of a sample table, we are able to fill 86% of missing DUNS numbers, the key identifier for business, with high confidence.
- Some problem areas are found in past projects. For example, a project had two subcontracts located in the same building. Small contractors lack not only financial and personnel resources but also proper training and awareness to protect themselves from emergencies, disasters, and cyber-attacks. In addition, their company webpages tend to have a lot of information, some of which might be sensitive.
- Publicly accessible data on government spending and contractor information can be easily utilized by bad actors to infer sensitive information, to disrupt the supply chain, and to stage attacks to delay or even destroy a mission critical project.
- Hidden subcontractors might be discovered by the correlation between large reductions in federal contracts and drops in employment in a given region and industry.

5. Conclusions

A key challenge in today's big-data era is the tremendous volume, velocity, and variety of data being produced; much of the data lacks veracity and value. Organizations must adapt to the rapidly changing data environment, delivering high-quality data products to support the agility necessary to produce business value. Importantly, emerging technologies such as AI/ML, 5G, block chain, and quantum computing are reshaping the landscape of data science research and practice. Equally, if not more important, research impact and efficacy must be demonstrated in data-driven applications with benefits received by organizations.

Under the aegis of the Institute for Chief Data Officers (iCDO) based at the University of Arkansas at Little Rock, we are developing the next generation of data-driven capabilities, with emphases on data analytics, quality, and science research and workforce development. Our team approaches data quality as a critical aspect of any enterprise. We are conducting research foundational to the development of new paradigms, tools, techniques, and solution approaches to create state-of-the-art, data-driven capabilities. We recognize the need to collaborate with organizations in the public and private sectors to achieve our mission in education and research. Our research encompasses data governance, data architecture, data integration, and data quality, producing powerful user interfaces and data analytics tools, and leveraging techniques ranging from statistical analysis to machine learning algorithms. Furthermore, we plan to investigate how emerging technologies such as artificial intelligence (AI), machine learning (ML), blockchain, 5G networks, and cloud computing can be exploited for advanced, data-driven capabilities.

References

1. Lee, Y., Madnick, S.E., Wang, R.Y., Wang, F. and Zhang, H., 2014. "A cubic framework for the chief data officer: Succeeding in a world of big data," MIS Quarterly. March 2014.
2. Lee, Y. W., Pipino, L. L., Wang, R. Y and Funk, J. D., 2006, Journey to Data Quality, MIT Press.
3. Krzysko, M. & Baney B., 2012, "The Need for Acquisition Visibility," Journal of Software Technology, Feb.: 4-9.
4. DAU, 2016, "DAU Center For Defense Acquisition Research Agenda 2016-2017," http://dau.dodlive.mil/files/2016/01/ARJ-76_ONLINE-FULL.pdf.
5. Undersecretary of Defense, 2007, "DoDI 5000.01 – Operation of the Defense Acquisition System"
6. Undersecretary of Defense, 2015, "DoDI 5000.02 – Operation of the Defense Acquisition System"

7. 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," Proc. of the 15th Annual Acquisition Research Symposium, Vol. 2, 491-499, Monterey, CA.
8. Wang, R., Wu, N., 2017, "On Data Capabilities for Acquisition Management," Proc. of the 14th Annual Acquisition Research Symposium, p. 376-390, Monterey, CA.
9. Wu, N., Wang, R., Tudoreanu, M. E., 2018, "Utilizing Public Data for Data Enhancement and Analysis of Federal Acquisition Data," Proc. of the 15th Annual Acquisition Research Symposium, Vol. 2, p. 94-113, Monterey, CA.
10. Metzger, M.J. & Flanagan, A.J., 2013, "Credibility and Trust of Information in Online Environments: The Use of Cognitive Heuristics," *Journal of Pragmatics*, 59, 210-220.
11. McKernan, M., Moore, N.Y., Connor, K., Chenoweth, M.E., Drezner, J. A., Dryden, J., Grammich, C.A., Mele, J.D., Nelson, W., Orrie, R., Shontz, D., & Szafran, A., 2016, "Issues with Access to Acquisition and Information in the Department of Defense," Technical report, Rand Corporation.
12. 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). Monterey, CA: Naval Postgraduate School.