SYM-AM-22-088

Excerpt from the Proceedings of the Nineteenth Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 11–12, 2022

Published: May 2, 2022

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

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

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

Making Federal Financial Data More Reliable With Emerging Tech

David I. Gill—Procurement Data Scientist, Analytics, Research and Technology Division, Office of the Chief Procurement Officer, Internal Revenue Service. [david.gill@irs.gov]

Avram Ibrahim, MBA—Program Manager, ASI Government. [aibrahim@ASIGovt.com]

Chaudry Umer, Esq.—Attorney, Cooley LLP. [uchaudhry@cooley.com]

Sonia Jolly—Accountant, Office of the Chief Financial Officer, Internal Revenue Service. [sonia.jolly@irs.gov]

Alicia M. Miller—Director, Analytics, Research and Technology Division, Office of the Chief Procurement Officer, Internal Revenue Service. [alicia.m.miller@irs.gov]

Abstract

Federal agencies are stewards of billions in taxpayer funds. Given the scale of federal financial transactions, maintaining reliable, high-quality financial data can be challenging. The use of emerging technologies such as robotic process automation (RPA) and natural language processing can reduce manual work for agency employees and improve the consistency of financial data. These technologies are key to success on financial audits and maintaining public confidence in the reliability of procurement and nonprocurement financial information.

Legislative Goals and Historical Perspective on Federal Contract Spending

Federal government spending has not always been open to public scrutiny. In fact, public access to government spending is a recent development. Prior to World War II, the key federal law controlling disclosure of government information was the Housekeeping Statute of 1789 (now codified at 5 U.S.C. § 301). This statute gave the federal government the "general authority to operate their agencies" and withhold records from the public. Information restrictions continued to prevail in the United States during the first half of the 20th century, as federal agencies claimed exemptions from disclosing such data due to security risks associated with pre- and post-wartime activities (Yu & Robinson, 2012). As such, the public remained unaware of how the government spent federal dollars.

The passage of the Freedom of Information Act (FOIA) in 1966 was a breakthrough for advocates of a more open and transparent government. Through a FOIA request, any person now has a right, enforceable in court, to obtain access to federal agency records. The history of the FOIA is important because the act continues to serve as the foundation for all transparency-related initiatives. The FOIA continues to evolve and respond to the changes in technology as it fulfills its objective of providing access to government information through electronic media.

In 1980, a report from the U.S. Comptroller General described the newly created Federal Procurement Data System (FPDS; General Accounting Office, 1980). The system was created to collect government-wide information on what is bought, who bought it, when, where, and how the contract was awarded. FPDS was created to serve several purposes:

- Enable Congress to make informed public policy decisions relating to procurement programs.
- Provide the executive branch with information necessary for managing the procurement process.
- Support interagency acquisition activities.
- Furnish suppliers with information on federal agency needs and enhance competition. (General Accounting Office, 1980)

The report noted some challenges with collecting data from agencies in a timely manner and ensuring the accuracy/completeness of reporting. Nonetheless, establishing FPDS was a significant milestone in collecting standardized, electronic data on federal procurement purchases.

On April 6, 2006, Senator Tom Coburn and Senator Barack Obama introduced S. 2590 in the Senate (Federal Funding Accountability and Transparency Act, 2006). The bill proposed a more proactive approach to transparency, where the federal government would connect disparate data sets to provide a more comprehensive picture of federal contracts and grants spending. Anyone with access to the internet would be able to download transaction-level details related to federal grants and contracts. This accessibility would eliminate the various inefficiencies and hurdles resulting from formal FOIA requests for such data.

Congress has enacted a number of statutes regarding contracting and other financial information. The Federal Funding Accountability and Transparency Act (FFATA) of 2006 requires "disclosing direct Federal agency expenditures and linking Federal contract, loan, and grant spending information to programs of Federal agencies to enable taxpayers and policy makers to track Federal spending more effectively." Further, the data verification and validation burden on federal employees can be eased thru RPA. The FFATA mandates "consistent" and "reliable" contract spending data. At the bill's signing, President Bush noted that the government issues more than $400 billion in grants and more than $300 billion in contracts annually. "Taxpayers have a right to know where that money is going, and you have a right to know whether or not you're getting value for your money," the president said. "By allowing Americans to Google their tax dollars, this new law will help taxpayers demand greater fiscal discipline" (Government Contractor, 2006).

The Digital Accountability and Transparency Act of 2014 (DATA Act) was enacted to

- Expand the Federal Funding Accountability and Transparency Act of 2006 (31 U.S.C. 6101 note) by disclosing direct Federal agency expenditures and linking Federal contract, loan, and grant spending information to programs of Federal agencies to enable taxpayers and policy makers to track Federal spending more effectively;
- Establish Government-wide data standards for financial data and provide consistent, reliable, and searchable Governmentwide spending data that is displayed accurately for taxpayers and policy makers on USASpending.gov … ;
- Improve the quality of data submitted to USASpending.gov by holding Federal agencies accountable for the completeness and accuracy of the data submitted. (Digital Accountability and Transparency Act, 2014)

The U.S. government is a global leader in spending transparency. A 2017 study identified the United States as one of several "advanced jurisdictions" with respect to open data (Maurushat et al., 2017). Foreign governments such as the United Kingdom, Netherlands, and Australia have been inspired by U.S. open spending data policies (Huijboom, 2011). Another article cited the U.S. DATA Act as an inspiration for Dubai's open data law (Rizvi, 2016).

Extramural Acquisition Research and Journalism

A literature review was conducted identifying research and news stories based on open federal spending data. A high level of data quality (timeliness, accuracy, completeness) is important to research integrity and public confidence. Ready availability and improving data quality and completeness has spurred extramural research involving federal spending. Data portals such as USASpending.gov have enabled research studies in a variety of professional fields. A paper by university researchers explains, "The opening of datasets in machine readable linked data is of particular importance to university and private industry researchers as

it opens hundreds of thousands of previously private datasets to be used for new research" (Maurushat et al., 2017). Data from USASpending.gov is available in a granular, disaggregated formation with detailed information each grant, contract, modification to a contract, and so forth.

Table 1. Examples of Extramural Research with Open Federal Spending Data

Professional Field	Organization	Study
Global Development	Columbia University	U.S. Spending in Haiti
Public Affairs	University of Missouri	Federal Contracting Trends in Missouri
Business Innovation	Hoover Institution	Supporting Advanced Manufacturing in Alabama
Disaster Medicine and Public Health Preparedness	Cambridge University Press	U.S. Governmental Spending for Disaster-Related Research
Sustainability	Environmental Research: Infrastructure and Sustainability	United States Federal Contracting and Pollution Prevention
Nonprofit Sector	Syracuse University	What Big Data Can Tell Us About Government Awards to the Nonprofit Sector
Education	Heritage Foundation	Pandemic Education Spending
Public Health	British Medical Journal Medicine	Use of Private Management Consultants in Public Health
Political Science	Cambridge University Press	Implementing Presidential Particularism: Bureaucracy and the Distribution of Federal Grants

Journalists have used open spending data to inform the public. News stories on contract and grant spending serve a variety of purposes and foster civil society. Local news stories highlight federal dollars and jobs spent in communities. News stories can spur debate about government spending priorities.

Table 2. Sample of News Stories Based on Open Data

Money Matters: Who Were USAID's Top Grantees in 2021? - *Devex*
Wyoming Left Out of Federal Coal Community Assistance Program - *WyoFile*
Nearly a Third of All Pentagon Contracts Have Gone to 5 Major Weapon Contractors - *The Boston Globe*
Government Says Contract for Covid-19 Database Was Competitively Bid - *The New York Times*

Architecture of Federal Award Reporting Systems

The data for USASpending.gov come from three primary sources. The Federal Procurement Data System–Next Generation (FPDS–NG) provides the federal prime contracts data; the Federal Assistance Awards Data System (FAADS) provides the federal assistance data, which is submitted to the Data Submission and Validation Tool (DSVT) hosted by the General Services Administration (GSA); and the FFATA Subaward Reporting System (FSRS) provides first-tier subaward data. Individual agencies report data on prime contracts and federal assistance. However, the prime grantees are responsible for reporting their subgrants data to FSRS. In addition to the primary sources of data for USASpending.gov, the website utilizes

specific data sets from the CFDA and vendor/grantee registration information from the System for Award Management (SAM).

It is no longer difficult to release millions of federal award records in machine-readable formats—the technical constraints are limited. However, such advances in technology pose increased risk of mixture of good and bad quality data leading to unintended consequences. In such cases, agencies may feel that they have fulfilled their obligation of providing access to the data. In reality, an incomplete and inconsistent data set provides little added value. In fact, it may even deter third parties from expending their resources and energy on data that are incorrect, depriving the public of valuable insights.

Data Quality Challenges and Audits

Agencies use a variety of different contract writing systems and financial systems. These systems capture contract numbers, dates, dollar amounts, and other information in neatly organized databases.

Achieving high-quality, reliable data can be challenging for agencies. Ten years ago, it was estimated that 66% of data on USASpending.gov were inaccurate (Sheridan, 2011).

Agency inspector generals perform an annual validation and verification audits of procurement data (Council of Inspectors General on Integrity and Efficiency, 2020). Audits examine whether procurement and nonprocurement data element reporting is current, accurate, and timely. Internal agency data is compared against USASpending.gov reporting to verify complete and timely reporting. Accuracy is checked by comparing with supporting documentation such as official contract files (FAR 4.8, 2021). For example, the Treasury Inspector General for Tax Administration's (TIGTA's) Fiscal Year (FY) 2020 DATA Act Audit identified IRS procurement data elements with a relatively high error rate (Treasury Inspector General for Tax Administration [TIGTA], 2021). Table 3 provides a TIGTA Analysis of IRS DATA Act procurement and grant statistical sample transactions.

Table 3. Comparison of FY2020 and FY2019 Statistical Sample Testing Results of Reported Data Elements With Error Rates Over 20% (TIGTA, 2021)

Data Element Name	*FY 2020*	*FY 2019*	*Change*
Primary Place of Performance Address	**44%**	52%	- 8%
Potential Total Value of Award	**29%**	35%	- 6%
Primary Place of Performance Congressional District	**26%**	21%	5%
Action Date	**25%**	28%	- 3%
Legal Entity Address	**23%**	19%	4%
Current Total Value of Award	**23%**	35%	- 12%
Period of Performance Current End Date	**22%**	24%	- 2%
Period of Performance Potential End Date	**21%**	28%	- 7%
Ultimate Parent Legal Entity Name	**21%**	52%	- 31%
Ultimate Parent Unique Identifier	**21%**	23%	- 2%

Automating DATA Act Validations

Given the scale of federal financial transactions—maintaining reliable, high-quality financial data can be challenging. The use of emerging technologies such as RPA and natural language processing can reduce manual work for agency employees and improve the consistency of financial data. RPA bots emulate humans in performing computer tasks (e.g., bot clicks with mouse and types information in systems). Further intelligent automation adds complex, artificial intelligence (AI) reasoning capabilities (e.g., locate specific information in varying locations within contract documents). These technologies are key to success on financial audits and maintaining public confidence in the reliability of contract and noncontract financial information.

The FPDS has a number of built-in data validation rules. When contract specialist users input invalid information that conflicts with an observable rule, the contract action record will be held in a draft status and prevented from publishing. Additionally, when mandatory data elements are left blank, the system will also prevent publication of contract action requires with missing data.

Table 4. Data Validation Rules Pulled from the FPDS User Interface on March 7, 2022

Sample Business Rule Validations	Sample Missing Data Validations
1. If the transaction is an initial award, then positions 7 and 8 of the PIID must be equal to the fiscal year of the date signed. 2. If the "Date Signed" on the action is on or later than February 3, 2017, only the values "Consolidated Requirements," "Consolidated Requirements with Written Determination," "Consolidated Requirements Under FAR 7.107-1(b) Exception," "Not Consolidated" can be selected for the Data Element "Consolidated Contract." 3. "Place of Manufacture" can only be "Mfg in U.S.," "Mfg outside U.S. - Use outside the U.S.," "Mfg outside U.S. - Resale," "Mfg outside U.S. - Trade Agreements," "Mfg outside U.S. - Commercial information technology," "Mfg outside U.S - Public interest determination," "Mfg outside U.S. - Domestic non-availability," "Mfg outside U.S. - Unreasonable cost," or "Mfg outside U.S. - Qualifying country (DOD only)" when the "PSC Code" is "1000-9999," except the values: 5510, 87**, 88**, 89**, 9410, 9430, 9440, 9610, 9620 or 9630.	1. Mandatory element: "nationalInterestActionCode" is missing for the award. 2. Mandatory element: "Emergency Acquisition" is missing for the award. 3. Mandatory element: "signedDate" is missing for the award.

A growing marketplace of vendors and technology solutions is helping to improve federal contract and financial data quality.[1] For example, FedDataCheck (www.feddatacheck.com) is an

[1] References to brand names and vendors are provided to assist agencies in finding products suitable for meeting agency needs. No endorsement is implied. This paper describes the salient functional

automation solution that runs additional validation checks and emails contract specialists requesting correction of data that appear to be erroneous. Having an automated data validation tool assists agencies in conducting FPDS validation and verification activities.

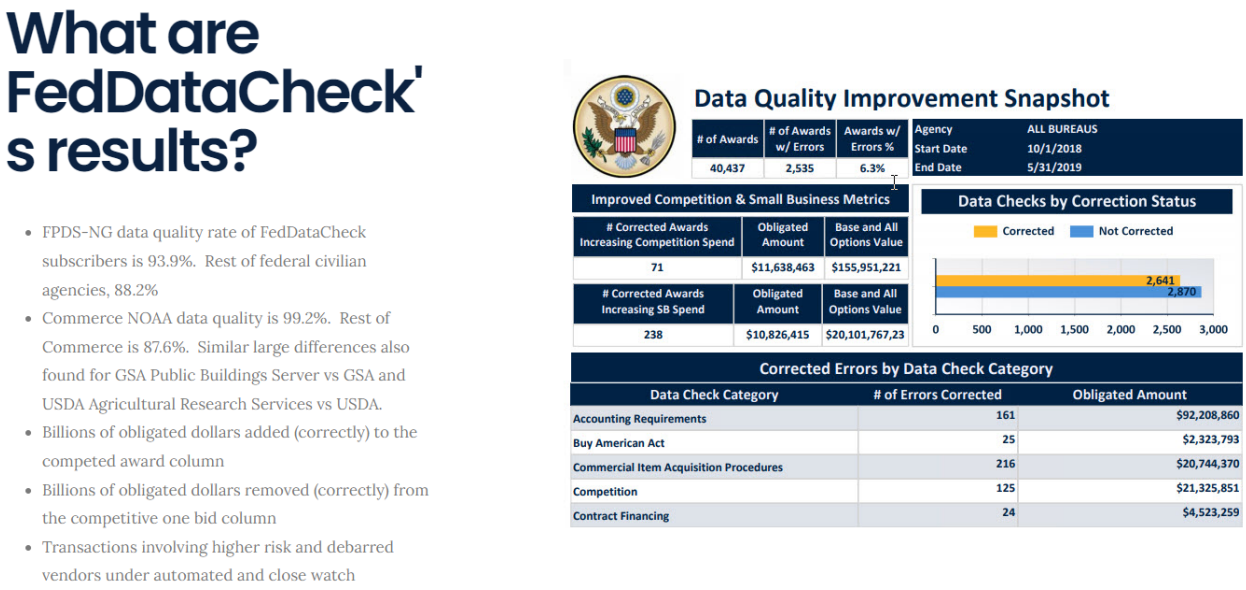


Figure 1. What Are FedDataCheck's Results?

The IRS held a competition seeking solutions to further improve data quality. Goals of the acquisition were to "achieve incremental improvement in IRS data" and "limit the amount of manual work required by government personnel" (G2Xchange FedCiv, 2019). Five vendors were selected to participate in an IRS Data Act Pilot. Significant progress had been made in improving the quality of FPDS data, but audits indicated that database entries did not always match information in signed contract award and modification documents.

Intelligent automation can locate specific pieces of information in contract documents. Data elements extracted from contract documents, such as dollar amounts, dates, and addresses, can then be compared with corresponding database records for consistency. The below screenshot shows the DATA Act bot in action. A PDF format contract modification was downloaded from a contract writing information system by the bot. DATA Act information was extracted from the contract modification document and compared for consistency with the corresponding FPDS contract action report.

performance characteristics of data validation software products to support acquisition planning in accordance with FAR 10.001(a)(3)(ii) and FAR 11.104.

clarifi Upload Retrieve Generate SF-30 Save Draft in FPDS | FPDS-NG user: Mmurugesan

Upload contract document: Upload ☑ Get Place of performance Document: **18 SF30.pdf**

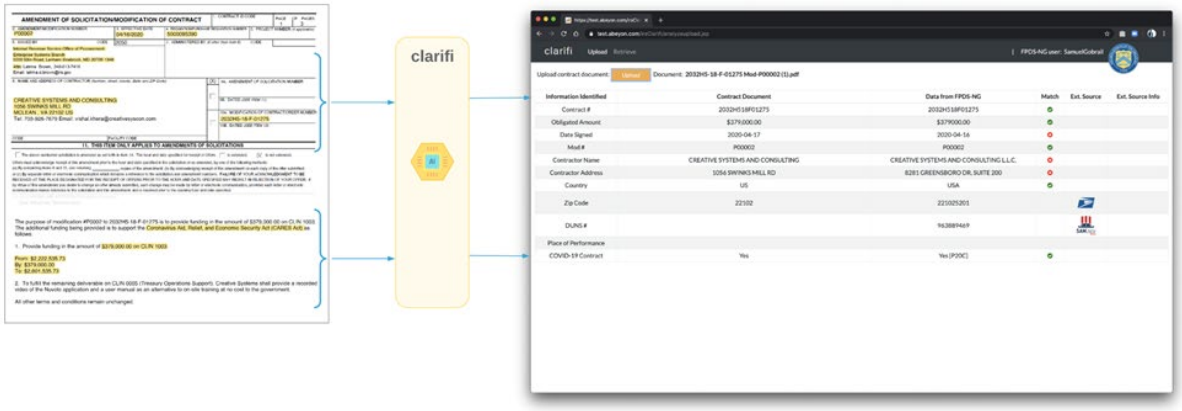
Information Identified	Contract Document	Data from FPDS-NG	Match	Ext. Source	Ext. Source Info
Contract #	205AE918P00215	205AE918P00215			
IDV Number					
Agency ID	2050	2050			
Office ID	205AE9	205AE9			
Obligated Amount	$-327.0	$-327.00			
Date Signed	2020-04-29	2020-04-29			
Mod #	P00001	P00001			
Contractor Name	TOBY FELDMAN, INC.	FELDMAN, TOBY INC			
NAICS Code		561492			
Contractor Address	3 COLUMBUS CIRCLE	450 FASHION AVENUE SUITE 502			
Country	US	USA			
Zip Code	10019	101230592			10019-8760

The DATA Act bot utilizes AI as well as RPA. These technologies combined allow for the automation of the process in four major stages. The first stage is to properly identify the relevant contract document that will contain the valid source of data elements, amongst nonrelevant procurement documents within a procurement folder management system. The second stage is to dive into the identified document and identify structured and unstructured data elements that are aligned to the FPDS–NG record. The solution then compares the extracted data to the data present in FPDS–NG and identifies all discrepancies in a report. Finally, through the report interface, the human administrator can choose to have the bot resolve all or specific discrepancies identified within FPDS–NG.

DATA Act Bot Stage 1: Extraction of the Correct Documents

The DATA Act bot success relies heavily on the identification of the correct document, replicating the ability of a human administrator to sieve through many documents and utilize historical knowledge to select the document that has the highest likelihood of containing the information pursued, based on training of the AI models implemented. Within the IRS, for example, procurement documents are filed electronically in accordance with a specific contract filing checklist, which varies depending on the stage of the procurement and type of procurement that the documents are associated with. The automated solution must determine the appropriate location for each document according to the filing path of each checklist and upload the document into the solution interface for Stage 2 utilizing RPA.

DATA Act Bot Stage 2: Data Elements Extraction



Stage 2 of the solution involves the extraction of data elements from the documents selected in Stage 1, across structured and unstructured fields. The solution is designed to extract 32 data elements ranging from **Date Signed**, an important field capturing the data of the procurement action's execution, to more complex elements such as **Period of Performance**. The complexity of each data element depends heavily on two factors: (1) data element availability and (2) the predictability of the location of the data element, as in, the frequency of the data element appearing in the same location(s) of the document every time. Data element availability refers to presence of this data element across a variety of different documents. A good example of this is **Contract Number** or **Vendor Name**, where the data can be found in documents ranging from official procurement forms (SF30s, SF1449, etc.) to Statement of Work documents, whereas data elements such as **NAICS Code** and **Obligated Amount** are limited to specific documents. AI models for complex data elements would have to properly identify the proper document where they would reside as well as the exact location of such data element within the document, considering human variability of data input.

Accuracy Assessment

As previously mentioned, to automate the validation of FPDS–NG records, the automation must be able to detect data elements from source data with a high degree of accuracy, comparable to that of a human administrator. Consider Acceptable Quality Levels (AQLs) typically included in contract Quality Assurance Surveillance Plans. An iterative, incremental process was used to improve model performance on each FPDS data element. Data validation models are given additional data, or changes are made to machine learning algorithm hyperparameters until an AQL is reached. F1 scores are calculated for each data variable (i.e., data field in FPDS) to determine accuracy and reliability of the model.

DATA Act Bot Stage 3: Reporting Discrepancies

Once the data elements are properly identified from the source procurement document, they are extracted and compared to the FPDS–NG data elements to identify discrepancies. The results are then downloaded and presented to a human administrator with discrepancies highlighted for validation. The administrator can select within the report which discrepancy should be corrected by the automation. This is done through a simple prompt in the report.

DATA Act Bot Stage 4: Resolving FPDS–NG Discrepancies

The marked discrepancies triggered by the administrator would prompt the automation to log onto FPDS–NG and resolve the discrepancies by modifying the data element in FPDS–NG to the data found in the procurement documents. The modified record could be saved in either draft mode or final mode depending on user and system input.

DATA Act Bot: Results

The DATA Act bot has brought demonstratable improvement to agency performance on DATA Act audits. The bot automates the tedious work of verifying the consistency of contract dates, dollar amounts, addresses, and other information. The TIGTA's FY2020 DATA Act Audit states that the "IRS received a score of 97.7 based on our sample and therefore has an overall quality rating of 'Excellent'" (TIGTA, 2021). TIGTA recommended that the IRS continue automated quality review efforts.

Conclusion

The United States has a robust program for spending transparency that has already brought significant benefits. That said, opportunities for improvement exist. Greater use of

automated data validation with business rules, natural language processing, and machine learning will increase data quality and transparency. A 2021 Government Accountability Office (GAO) report found that USASpending.gov fosters greater transparency, informs decision-making, and helps identify fraud (Government Accountability Office, 2021). Further, GAO noted that the website employs user-friendly, human-centered design. GAO recommended increasing training for targeted user groups to obtain more benefits.

References

Council of Inspectors General on Integrity and Efficiency. (2020, December 4). *Inspectors General guide to compliance under the DATA Act*. https://oig.treasury.gov/sites/oig/files/2021-01/OIG-CA-21-008.pdf

Digital Accountability and Transparency Act, Pub. L. No. 113-101 (2014). https://www.congress.gov/113/plaws/publ101/PLAW-113publ101.pdf

FAR 4.8, Government Contract Files (2021). https://www.acquisition.gov/far/part-4#FAR_Subpart_4_8

Government Accountability Office. (2021, December). *Federal spending transparency: Opportunities exist for Treasury to further improve USAspending.gov's use and usefulness* (GAO-22-104127). https://www.gao.gov/products/gao-22-104127

General Accounting Office. (1980, April). *The Federal Procurement Data System: Making it work better*. HeinOnline.

Government Contractor. (2006). President signs act creating web site of contract and grant recipients. *Government Contractor*, *48*(345).

Federal Housekeeping Statute, 5 U.S.C. § 301 (1789).

Huijboom, N. (2011). Open data: An international comparison of strategies. *European Journal of ePractice*. https://research.vu.nl/en/publications/open-data-an-international-comparison-of-strategies

Maurushat, A., Trzaskowski, J., & Al-Alosi, H. (2017). Open data: Turning data into information, and information into insights that allow for evidence-based policy. *Newcastle Law Review*, *12*, 104–147.

G2Xchange FedCiv. (2019, October 1). *IRS awards Data Act Improvement Pilot to five companies*. MileMarker10. https://etc.g2xchange.com/statics/irs-awards-data-act-improvement-pilot-to-five-companies

Rizvi, Z. (2016). Dubai data law. *Court Uncourt*, *3*(3), 2–4.

Federal Funding Accountability and Transparency Act of 2006, Pub. L. No. 109-282 (2006). https://www.congress.gov/congressional-report/109th-congress/senate-report/329/1

Sheridan, B. (2011). *DATA Act: The federal case for XBLR*. Maryland Association of Certified Public Accountants.

Treasury Inspector General for Tax Administration. (2021, October). *Fiscal Year 2020 Digital Accountability and Transparency Act reporting compliance* (Report No. 2022-10-003). https://www.treasury.gov/tigta/auditreports/2022reports/202210003fr.pdf

Yu, H., & Robinson, D. G. (2012). The new ambiguity of "open government." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2012489

Acquisition Research Program
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net