Dialing for Dollars: Improving Cost Estimating Accuracy through Data Mining

Brittany Holmes, James Glenn, Dr. Christian Smart
Missile Defense Agency
June 2016

Abstract

This paper shows the benefits of leveraging obligation and expenditure data from DoD Enterprise Resource Planning (ERP) Systems to supplement cost data from CCDRs or IPMRs. One benefit is that cost trends and funding patterns can be analyzed with more precision because the ERP data is updated real-time vice monthly or annually. Another benefit is that more comprehensive CERs can be created because ERP systems capture all cost vice only contracts with cost reporting CDRLs.

Keywords: Budgeting, Data Collection, Government, Methods, Microsoft Excel, Parametrics, Regression, Data Mining, Big Data, Data Science, Analytics

Table of Contents

List of Figures

Introduction

The purpose of this paper is to present another data source for cost analysts to add to their repertoire and some examples of ways this data might be utilized in creating cost estimates. While Cost and Software Data Reports (CSDRs) provide detailed costs for programs, these reports may not be provided for an entire contract, small contracts less than $50 million, or for other funding vehicles such as Military Interdepartmental Purchase Requests (MIPRs). Integrated Program Management Reports (IPMRs) are also very useful for cost estimators, but are not available on Firm Fixed Price (FFP) contracts and contracts less than $20 million. CSDRs may only be delivered annually or at specific milestone events and IPMRs are typically delivered monthly. Unlike CSDRs and IPMRs, the data contained within an Enterprise Resource Planning (ERP) database is dynamic and provides updates as transactions are made.

Because ERP systems collect data across all funding vehicles, they can be used to comprehensively report and analyze cost for high level activities. Examples include how much money was spent on a specific flight test, or how much money was spent on a specific program in the past three years. ERP data can also be used to create parametric cost relationships to support future cost estimates. This paper provides several examples of how ERP data can support cost estimating.

Using ERP systems in cost estimating has the potential to increase the accuracy of cost estimates, since the leveraging of this information can facilitate the development of estimates that are timelier, and more detailed. This use is one example of the application of data mining techniques in cost estimating, and reflects a growing trend across multiple industries to use the larger amount of data now available. These techniques using large amounts of data are often referred to as "big data" or "data science" and are already prevalent in many other industries, as discussed by Eric Druker in his training session on Big Data at the 2015 ICEAA Conference [1].

Following a brief overview of an ERP system and the current DoD ERP financial systems in use, a description of how the data within an ERP financial system needs to be transformed is provided.   Several examples are then provided for how the decoded data can be used to
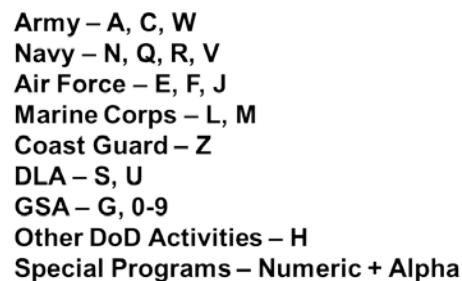
supplement other cost sources. Cost applications are then examined using expenditure data: Property Accountability, cost growth, obligation profiles, and cost estimating relationships (CERs).

ERP System Background

ERP tools provide the ability to deliver an integrated suite of business applications across a broad range of operational processes: finance, HR, distribution, manufacturing, service and supply chain [2]. ERP systems allow for IT cost savings, business process efficiencies, and business innovations across an organization [2].   In the Department of Defense (DoD), there are several different ERP systems in use that focus on financial, asset, and logistics management. The sample DoD ERP systems listed below focus on financial management:

- General Fund Enterprise Business Systems (GFEBS) is utilized by the Army and was initially deployed in 2008 [3]. GFEBS is a web-based COTS ERP solution, based on a SAP background, that manages financial, asset, and accounting across the active Army, the Army National Guard (ARNG), and the US Army Reserves (USAR) [4].

- DAI was initially implemented in 2008 [3] and there are 28 agencies that are planned to adopt this ERP system [5]. This ERP system is also COTS, but based on Oracle Federal Financial [5].

- The Navy Enterprise Resource Planning, Navy ERP, was initially implemented in 2007 [3] and is an integrated financial, acquisition, and logistics information technology system that provides financial and budgetary management for all Navy system commands. Navy ERP is based on a COTS SAP business suite [6].

- Defense Enterprise Accounting and Management System (DEAMS) is a COTS-based financial management initiative implemented in 2007 [3] for US Transportation Command, Defense Financial and Accounting Service (DFAS), the US Air Force, and other DoD agencies [7].

Decoding the Data

Besides obligation and expenditure amounts, there are a number of fields contained in a financial ERP system that can prove useful to a cost estimator. The fields in Figure 1 are often contained in financial ERP systems and can be very helpful in understanding the context of the expenditure records.

Field Title	Short Description	Field Title	Short Description
Expenditure Organization	Name of the Program Office	Program Element	PE submitted in budget requests
Vendor	Contractor/OGA performing the work	Award Creation Date	Contract award date
ACRN	Accounting Classification Reference Number	Last Update Date	Last date changes were made to the document
Net Obligation	Amount obligated minus any cancelled obligations	Item Description	Description of the work performed on the invoice record
Product Service Code	Defines type of product or service	Project/Task Numbers	Program Office and Agency specific tracking codes
Funding Year	Year document was funded	CLIN	Contract Line Item Number
Award Number	Contract or purchase request number	Last Payment Date	Last date a payment was made on the document
Distribution Creation Date	Date of the original funding document	Currency Type	Type of currency
Invoiced/Delivered Amount	Amount invoiced	Fund	Funding agency, color of money, funding years

Figure 1. ERP Financial System Fields

In order to utilize the data within an ERP system decoding all of the fields may be required as different agencies utilize different fields and code systems. Often data is combined to reduce the number of fields in the database and therefore reports. At one DoD agency, the "Fund Code" in Figure 2 provides some very helpful information for a cost

Figure 2. Fund Code Decoded

estimator. From one field, the cost estimator can understand the appropriation that was utilized for the expenditure record and the start and end date of the funding. In the example in Figure 2, "9704001415D" is provided in one field in the database.

Another useful field is the "Award Number", the contract number or the purchase request number (Figure 4 and Figure 5). Understanding the parts of this field can also help describe some of the expenditure data. The first six digits of the award number refer to the DoD Activity Address (DoDAAC) code. This code uniquely identifies a DoD organization that has the authority to fund or pay bills or requisition, contract for, receive, have custody of, issue, or ship DoD assets. The Defense Logistics Agency (DLA) has a tool to decipher any DoDAAC code called DAASINQ [8]. The first digit of the DoDAAC code, as

Figure 3. DoDAAC Code First Digits

seen in Figure 3, are helpful in determining which branch of DoD the contract belongs to and depending on the service, each of the following digits might also have a specific meaning [9].

The following digits differ for contracts and purchase requests. For contracts, the next two digits are the last two digits of the year the contract was issued. The next letter provides the type of contract: "A" for a blanket purchase agreement, "C" for non-delivery order contracts, "D" for delivery order

Figure 4. Contract Number Decoded

contracts, "F" for UNICOR purchases, and "P" for purchase orders. The last 4 digits are the PIID Code which are just sequentially numbered based on when the contract was awarded. Some agencies have specific sequences of numbers reserved for the type of business that the contract is with, for example, Small Business Innovation Research (SBIR) or Small Business Technology Transfer (STTR) business contracts.

For purchase requests the digit following the DoDAAC is the last digit of the year of the issue year of the purchase request. The following 5 digits are the requisition numbers which are sequentially issued based on the order that the purchase requests are received that year.

Figure 5. Purchase Request Number Decoded

Using Excel functions or Visual Basic for Applications (VBA), the different parts of the fields, like the "Fund Code", can be broken out into separate columns. Then by utilizing lookup functions or by creating relationships in an Access Database, the codes can be translated into words to make it easy for any analyst to utilize, search, or create summary Pivot tables.   For example, instead of providing a Product Service Code in an Excel column the output can read "R&D Engineering Development Services", enabling an analyst to search records on a product or service type.

Once the data has been decoded, other databases can be joined to the ERP financial database. Using Access, contract databases and other purchase request databases can be joined to the financial data based on the contract, delivery order, CLIN numbers, or any other unique field. This will provide added details to the expenditure records such as the total award value, a description of what is being funded on specific CLINs, detail about the vendors, and any other information that is linked to the contract or purchase request number if that data is not already incorporated in the ERP financial database.

Cost Estimating Applications

As seen in Figure 6, one DoD ERP database contains 65,470 expenditure records from 2011 to 2015 for contracts, purchase requests, travel documents, and other documents.

Type of Document	Documents Between 2011 and 2015
Travel	54,205
MIPR	7,100
Contracts	1,803
Other Documents	2,362
Total	65,470

Figure 6. Types of Documents in Database

While cost reporting CSDRs provide a large amount of detailed data, these cost reports are not provided for all of an agency's contracts. Cost reports are required on all contracts greater than $50 million, but can be requested for other contracts. IPMR reports are only required on contracts greater than $20 million that are not Firm Fixed Price (FFP) type contracts.

Small Contracts and Other Funding Vehicles

For smaller contracts and other funding vehicles, cost data is often limited to proposal data, initial contract documents, contract modifications, or a Statement of Work (SOW). Within the expenditure data, actual cost can be extracted for small contracts that do not qualify for cost

reporting. The execution years / period of performance (PoP), the total expenditure amount, and a brief description of what is included in the contract is shown Figure 7 for a small contract. If more research was performed for this contract, this data could be incorporated into a cost estimate for polymer parts.

Contract Number	Execution Years/PoP	Total Expenditures	Description
AB0123-12-C-6013	FY12-FY13	$836,998.92	Polymer Parts Optimized for an Application

Figure 7. Small Contract Information from Database

The expenditure data can also provide insight into the actual cost of MIPRs or other types of purchase requests. As the funding organization does not have direct access to cost reporting for these types of purchase requests, SOWs and Independent Government Estimates (IGEs) are typically the only sources for information. As long as a contract number or other distinguishing factor about the work is known, the expenditure data can be searched to find out exactly how much was expended for the purchase request. In the example in Figure 8, there was a $2.24M difference between the SOW and the actual expended amount.

Figure 8. MIPR SOW Comparison

Utilizing data from an ERP system can provide a cross check resource to cost reports, especially if cost reports are provided annually or after significant milestones. For one contract the Actual Cost of Work Performed (ACWP) for one Task Order from an Earned Value Management (EVM) database can be compared to the total expenditures for that Task Order from the ERP database, as seen in Figure 9. The EVM database is dated as of June 2015, but the ERP database has total expenditures for that Task Order as of today's date.   An additional $6.35M has been expended on this Task Order since June.

Database	Total Cost	Expenditure Period
EVM Data	$7.78M	08/14 – 06/15
Expenditure Records	$14.13M	08/14 – 02/16

Figure 9. EVM Database vs Expenditure Records

Property Accountability

As part of DoD's property accountability task, expenditure data has been utilized to estimate the original value of several DoD assets. As an example in Figure 10, contract research found a contract document mentioning Asset B. As Asset B was the only item listed on this specific CLIN,

ITEM NO 0029	SUPPLIES/SERVICES	QUANTITY	UNIT Monetary Value	UNIT PRICE	AMOUNT $14,156,413.00
	Asset B CPFF FOB: Destination				
			ESTIMATED COST		$13,191,938.00
			FIXED FEE		$964,475.00
			TOTAL EST COST + FEE		$14,156,413.00

Figure 10. Contract Document Listing Asset B

the Contract and CLIN number were used to extract the ERP expenditure data. For this specific CLIN, the total expenditures were $13,525,309 which is significantly lower than the amount provided on the contract document $14,156,413.

Understanding Trends

Besides understanding total costs for contracts or purchase requests, the expenditure data can be used to determine annual trends or cost growth. As an example, understanding annual cost increases for the utilities of a facility. First a search was performed on the item description for utilities, since specific contract numbers were unknown. Second that list was paired down based on the facility name in the item description. This list was then summarized based on the item description and the funding year. Assuming that the facility has not increased in size or capacity, an average percent increase can be determined by inflating to a common base year, Figure 11. This can be used to predict the total cost for utilities for this facility for the following year.

Year	Cost (BY$M)	Cost (BY16$M)	% Increase
2011	3.40	3.67	
2012	3.50	3.71	1.1%
2013	3.70	3.87	4.2%

2014	3.80	3.91	1.2%
Average	**3.60**	**3.79**	**2.2%**

Figure 11. Utility Annual Cost Growth

Mapping Expenditure Records to a Test WBS

As part of an ongoing study to understand the total cost of tests, expenditure data was utilized to understand variable test costs. For this project, it was assumed that any expenditure record naming a specific test event was a variable cost; otherwise, test costs were considered fixed.

Expenditure records for the past several years were first categorized based on budget type that was obtained from the "Program Element" code in the data. As there is a specific budget project code for "Test" these records were flagged for examination. The database and several other databases were merged via Access to obtain as much descriptive information about the expenditure records as possible.

Starting first with the "Test" budget records, expenditures were mapped to the lowest level possible to a Test WBS. The test event name was also documented. As an example, if the expenditure record stated "Range 14 for Flight Test #45", the record was documented to WBS 1.2.45.3, Flight Test #45 / Test Range. As the expenditure records already have the program, expenditure year, contract number, and other fields to research, these fields were also incorporated into the report. In some cases, it was not possible to map expenditure records for a single test to a single WBS element. In those cases, two major rules were developed and followed:

1. If more than one test, and up to three tests, were mentioned in the record description, the costs were assigned a percentage of the total expenditure based on the date of the historical test event:

$$y_i = \frac{\frac{x_i}{4}}{\frac{x_1}{4} + \frac{x_2}{4} + \frac{x_3}{4}} \times y_T$$

- y_i is the cost assigned to the ith test event
- x_i is the FY quarter in which that test occurred
- x_1, x_2, and x_3 are the quarters that each test occurred
- y_T is the total cost for the record

2. If an expenditure record fell into more than one, but less than 3 WBS categories, the expenditure record was split equally to each WBS category unless specific information was provided.

After all historical test expenditure records were mapped to the Test WBS, summaries were created showing the total cost belonging to each test event in each WBS category.

Test Obligation Profile and Effects of Delayed Tests

Utilizing the expenditure records for one historical flight test, a notional flight test obligation profile was created. The "Distribution Date" from the expenditure records was summarized into a cumulative percentage by quarter. Quarters were then categorized relative to the Best Estimated Test Date (BETD). As seen in Figure 12, 90% of the total flight test obligations have accrued 1 quarter before the flight test, meaning that 10% of the test costs are obligated during the execution of the test event or post event.

To visualize how a delayed test affects the total obligation profile of a flight test, two different scenarios were examined: a flight test delay in the qu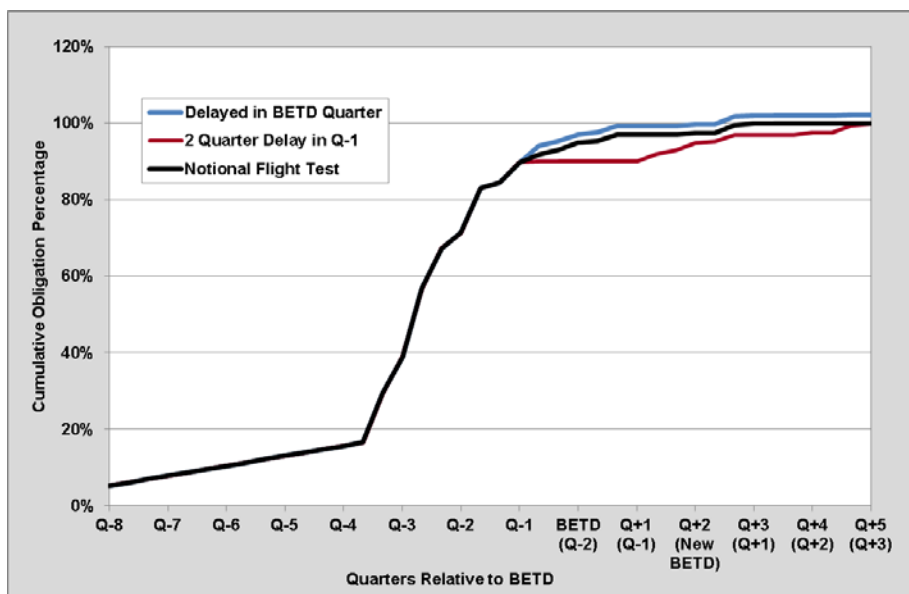arter of the BETD and a 2 quarter test delay with 1 quarter lead time. A delay during the quarter of the BETD assumed an additional cost would be accrued

Figure 12. Notional Test Obligation Profile and Effects of Delayed Tests

because assets and personnel are already deployed, which increases the total obligations for the flight test. If the test was shifted prior to the quarter of the BETD, this would extend the test, but not cause a significant increase in the total obligations. A similar graphic and analysis can be completed for expenditures in addition to obligations.

Creating a Total Cost CER for Ground Tests

Once all of the historical test records were mapped to specific test events, data was collected about the historical tests in order to obtain independent variables for CERs. For ground tests, configuration diagrams were utilized to count the total number of assets involved. If multiple test architectures existed for a ground test, only unique assets were counted. For tests with multiple ground test phases, the assets were summed across all of the phases. In progress ground tests and

tests that occurred prior to 2011 were removed from the dataset to limit capturing partial test costs.

Actuals for specific ground tests were plotted versus the total number of assets and a positive correlation was observed ($y = 1.6x - 8.44$, $R^2 = 0.87$, $F = 65.43$) where costs are in BY15\$M. As can be seen above, the ERP expenditure

Figure 13. Total Ground Test Cost CER

data contains a greater proportion of the historical cost than using only CCDR & IPMR reports allowing for a more comprehensive CER to be created.
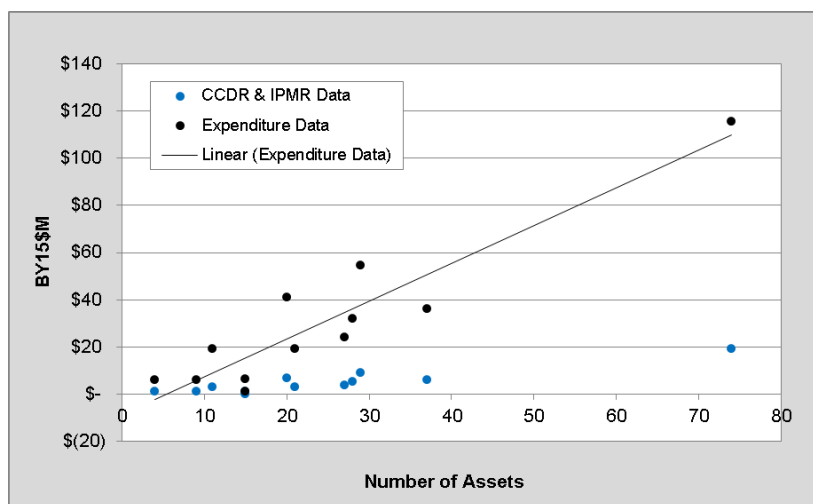
Cautions about Utilizing ERP Expenditure Data

While the data within the ERP expenditure data can be very useful, there are several issues in solely utilizing this data for creating a cost estimate. First, the data is dependent upon the analyst that inputs the information. If for example, in the expenditure description, the analyst only provides a high level overview such as "For FY13 Contract #XYZ, CLIN 0004", capturing data via the

description might reduce the overall amount of data that can be captured. If this one record was related to a specific test event, this record would have been missed in the test cost categorizations.

The data that is held within the ERP expenditure data is also dependent on the contract structure. In the test categorization example, some contractors provide all of their flight test information in one CLIN. In order to accurately represent their participation in specific flight tests, the expenditures must be allocated based on the date, size of flight test, or some other allocation strategy. If contractors provide the cost of individual events as a specific CLIN, then the data collection is much easier and the data is more accurate.

While anyone can look through the data for keywords, this does not mean that you capture all of the data that is needed or that the correct data was captured to create an estimate. If for example, in the utilities example, there were multiple facilities in FY2013 and only one facility captured in the expenditure line in FY2014, then the percentages are erroneously representing cost growth, where in reality utility costs may have decreased between those two years. It is highly advised before utilizing expenditure data, the analyst should contact a budget analyst or cost analyst who is acquainted with the program to make sure data is being captured accurately.

Summary

This paper presented another data source for cost analysts to add to their tool box and a few examples of methods to utilize the wealth of data contained within ERP financial systems. Once an analyst decodes the ERP financial system data, this data can be utilized for supplementing other cost data sources for understanding cost trends, creating cross checks, updating estimates, or creating CERs. While the expenditure data can be very useful, it should be used with a similar caution as any other data source, in that before utilizing the data in an estimate the analyst should research the background of the program or contract.

References

[1] Druker, E., "Data Science & Cost Estimating: The Future of Analytics," presented at the 2015 International Cost Estimating and Analysis Association Conference, San Diego, June, 2015.

[2] "Enterprise Resource Planning (ERP)", (2016). Gartner Research Official Site. Accessed February 15. http://www.gartner.com/it-glossary/enterprise-resource-planning-erp/.

[3] Department of Defense Inspector General. (2012). "Enterprise Resource Planning Systems Schedle Delays and Reengineering Weaknesses Increase Risks to DoD's Auditability Goals (DODIG-2012-111). Alexandria, VA: Office of the Deputy Inspector General for Auditing. https://defensesystems.com/articles/2012/07/17/~/media/GIG/ Defense%20Systems/PDF/DODIG%202012%20111%201.ashx.

[4] "General Fund Enterprise Business Systems (GFEBS)", (2016). US Army PEO EIS (US Army PEO Enterprise Information Systems) Official Site. Accessed February 15. http://www.eis.army.mil/programs/gfebs.

[5] Gansler, J.S. and W. Lucyshyn. "Defense Business Transformation." (2009). The Center for Technology and National Security Policy, National Defense University. Accessed: February 15, 2016. http://ctnsp.dodlive.mil/files/2014/10/Defense-Business-Transformation-Report.pdf

[6] Office of the Secretary of Defense. (2013). "Navy Enterprise Resource Planning (ERP)". The Office of the Director, Operational Test and Evaluation. http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved =0ahUKEwiXsNrG1frKAhUGPz4KHevxDxQQFggvMAM&url=http%3A%2F%2Fwww.dote.os d.mil%2Fpub%2Freports%2FFY2013%2Fpdf%2Fnavy%2F2013nerp.pdf&usg=AFQjCNGsE e5wry7gPwcH3luw2_4Ood75Yw

[7] "Defense Enterprise Accounting and Management System (DEAMS) Public Website", (2005). Defense Acquisition University Official Website. Accessed February 15 2016. https://acc.dau.mil/CommunityBrowser.aspx?id=32306

[8] "DAASINQ" (2016). "Defense Logistics Agency Transaction Services Official site. Accessed March 3. https://www.transactionservices.dla.mil/daasinq

[9] Department of Logistics and Materiel Readiness. (2013). Defense Logistics Management System Volume 6, Logistics Systems Interoperability Support Services Change 2 (DLM 4000.25, Volume 6) Washington, DC: Office of the Department of Logistics and Materiel Readiness. http://www2.dla.mil/j-6/dlmso/elibrary/manuals/ dlm/dlm_pubs.asp