NPS-LM-14-010

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Visualization of Big Data Through Ship Maintenance Metrics Analysis for Fleet Maintenance and Revitalization

27 February 2014

LT Isaac J. Donaldson, USN

Dr. Thomas J. Housel, Professor

Johnathan Mun, Professor

Sandra Hom, Research Associate

Trent Silkey, Research Assistant

Graduate School of Operational & Information Sciences

Naval Postgraduate School

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 Business & Public Policy at the Naval Postgraduate School.

Abstract

There are between 150 and 200 parameters for measuring the performance of ship maintenance processes in the U.S. Navy. Despite this level of detail, budgets and timelines for performing maintenance on the Navy's fleet appear to be problematic. Making sense of what these parameters mean in terms of the overall performance of ship maintenance processes is clearly a big data problem.

The current process for presenting data on the more than 150 parameters measuring ship maintenance performance costs and processes, containing billions of data points, is still done by static, cumbersome spreadsheets. The central goal of a recent research project was to provide a means to aggregate voluminous maintenance data in such a way that the causal factors contributing to cost and schedule overruns can be better understood by ship maintenance leadership.

Big data visualization software was examined to determine if visualization tools could improve the understanding of U.S. Navy ship maintenance by its leaders. Our research concludes that the visualization of big data supports decision making by enabling leaders to quickly identify trends, develop a better understanding of the problem space, establish defensible baselines for monitoring activities, perform forecasting, and evaluate metrics for use.

Keywords: Big Data, Big Data Visualization, Visualization Software, 3D Printing, 3D Laser Scanning Technology, Collaborative Product Lifecycle Management.

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

The extraordinary demand placed on U.S. armed forces requires that the highest levels of readiness be maintained. The pressure to reduce costs, while maintaining the highest levels of readiness, compels each of our military services to periodically review internal processes to ensure responsible use of our nation's resources. One such process currently in review involves Department of Defense maintenance programs. In FY2011, the U.S. Navy spent $682 million maintaining its destroyers, representing only 22% of the 286 ships currently in the fleet. According to a 2012 Government Accountability Office report on ship readiness, by 2019, the U.S. Navy expects to have grown its fleet by another 14 ships to a total of 300. The size of the U.S. Navy's ship maintenance budget makes it a prime candidate for review.

Reviewing ship maintenance programs is a complex task. There are between 150 and 200 parameters for measuring the performance of ship maintenance processes in the U.S. Navy. Despite this level of detail, budgets and timelines for performing maintenance on the Navy's fleet appear to be problematic. Making sense of what these parameters mean is clearly a big data problem. Fortunately, the value of big data analysis has become evident and many analysis solutions exist. Big data visualization was selected for closer examination and a sample of U.S. Navy ship maintenance availabilities were used to explore the technique.

Big data visualization software was examined to determine if visualization tools could improve the understanding of U.S. Navy ship maintenance by its leaders. This report concludes that the visualization of big data supports decision making by enabling leaders to quickly identify trends, develop a better understanding of the problem space, establish defensible baselines for monitoring activities, perform forecasting, and evaluate metrics for use. For U.S. Navy ship maintenance decision makers desiring ways to improve the speed and accuracy of their decisions, they should consider the use of visualization software in their industry. To optimize the use of big data visualization, the authors recommend the continued and expanded collection of data, identification of performance accounting software for tracking, and

the use of forecasting once accurate ship maintenance performance baselines are established.

Acknowledgments

Much thanks and appreciation to John Kornitsky for the copious amounts of his personal time he sacrificed to ensure we understood the outputs of the visualization software enough to write about it. Also, his review and suggestions for improvement were beneficial parts of the revision process.

Thanks to David J. Furey, an employee of Norfolk Ship Support Activity, for the many hours he spent on the phone , helping us to understand the importance and impact of new maintenance, deferred maintenance, and modernizations.

THIS PAGE INTENTIONALLY LEFT BLANK

About the Authors

Isaac J. Donaldson is an Information Professional officer in the U.S. Navy and is completing a master's degree in Network Operations at Naval Postgraduate School. For the past 16 years, Lieutenant Donaldson has served tours of duty onboard the aircraft carriers USS *Carl Vinson*, USS *John F. Kennedy*, and the USS *Harry S. Truman*, both at sea and during maintenance in the Norfolk Naval Shipyard. He has attended many academic and service training facilities focused on information systems education including Joint Forces Staff College in Norfolk, Virginia and the Center for Information Dominance in Pensacola, Florida. He has earned multiple professional qualifications required of Information Professional officers in the U.S. Navy including the Information Dominance Warfare Officer designation.

Isaac J. Donaldson
Lieutenant, United States Navy
Monterey, CA 93940
Tel: (386)383-5584
E-mail: ijdonald@nps.edu

Sandra Hom is a research associate at the Naval Postgraduate School and specializes in market structures, industry benchmarking research, and knowledge value added analysis.

Sandra Hom
Graduate School of Operational & Information Sciences
Monterey, CA 93943-5000
Tel: (831) 656-7657
Fax: (831) 656-3407
E-mail: schom@nps.edu
Naval Postgraduate School

Thomas Housel specializes in valuing intellectual capital, knowledge management, telecommunications, information technology, value-based business process reengineering, and knowledge value measurement in profit and non-profit organizations. He is currently a tenured full professor for the Information Sciences (Systems) Department. He has conducted over 80 knowledge value added (KVA) projects within the non-profit, Department of Defense (DoD) sector for the Army, Navy, and Marines. He also completed over 100 KVA projects in the private sector. The results of these projects provided substantial performance improvement strategies and tactics for core processes throughout DoD organizations and private sector companies. He has managed a $3 million+ portfolio of field studies, educational initiatives, and industry relationships. His current research focuses on the use of KVA and “real options” models in identifying, valuing, maintaining, and exercising options in military decision-making.

Thomas Housel
Graduate School of Operational & Information Sciences
Monterey, CA 93943-5000
Tel: (831) 656-7657
Fax: (831) 656-3407
E-mail: tjhousel@nps.edu
Naval Postgraduate School

Johnathan Mun is a research professor at the U.S. Naval Postgraduate School (Monterey, CA) and teaches executive seminars in quantitative risk analysis, decision sciences, real options, simulation, portfolio optimization, and other related concepts. He has also researched and consulted on many Department of Defense and Department of Navy projects and is considered a leading world expert on risk analysis and real options analysis. He has authored 12 books. He is also the founder and CEO of Real Options Valuation Inc., a consulting, training, and software development firm specializing in strategic real options, financial valuation, Monte Carlo simulation, stochastic forecasting, optimization, and risk analysis located in northern California.

Johnathan Mun
Graduate School of Operational & Information Sciences
Naval Postgraduate School
Monterey, CA 93943-5000
Tel: (925) 271-4438
E-mail: jcmun@nps.edu
E-mail: jcmun@realoptionsvaluation.com

Trent Silkey is a former research assistant at the Naval Postgraduate School and specializes in knowledge value added methodology, market comparables valuation, and the interpretation of intellectual capital valuation analysis.

Trent Silkey
Graduate School of Operational & Information Sciences
Naval Postgraduate School
Monterey, CA 93943-5000
E-mail: trsilkey@nps.edu
E-mail: jcmun@realoptionsvaluation.com

NPS-LM-14-010

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Visualization of Big Data Through Ship Maintenance Metrics Analysis for Fleet Maintenance and Revitalization

27 February 2014

LT Isaac J. Donaldson, USN

Dr. Thomas J. Housel, Professor

Johnathan Mun, Professor

Sandra Hom, Research Associate

Trent Silkey, Research Assistant

Graduate School of Operational & Information Sciences

Naval Postgraduate School

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Visualization of Big Data Through Ship Maintenance Metrics Analysis for Fleet Maintenance and Revitalization

Introduction

Overview

There are between 150 and 200 parameters for measuring the performance of ship maintenance processes in the Navy. Despite this level of detail, budgets and timelines for performing maintenance on the Navy's fleet appear to be problematic. Making sense of what these parameters mean in terms of the overall performance of ship maintenance processes is clearly a "big data" problem.

A team from the Naval Postgraduate School (NPS) was requested by Program Executive Office (PEO) Ships to work with naval ship maintenance metrics groups to provide additional options regarding how large datasets could be optimized. The current process for presenting data on the more than 150 parameters measuring ship performance maintenance costs and processes, containing billions of data points, is still done by static, cumbersome spreadsheets. The central goal of this project was to provide a means to aggregate voluminous maintenance data in such a way that the causal factors contributing to cost and schedule overruns can be better understood by ship maintenance leadership. By providing this kind of information in an intuitively visual form, leadership could be assisted in budget and scheduling decision making.

The results of the project are presented in several sections in this report. In the Literature Review section, we review the big data world by looking at the $11 billion dollar industry in 2012. We examine the issues, components, technologies, and tools surrounding big data. In the section titled, "Governement Spending on Big Data," the focus is on big data and the federal government, which spent approximately $5 billion in 2012 on national security and military applications. Included in this section are public- sector big data projects, case studies, and lessons learned. "Ship Maintenance Vignettes" are presented next to provide a framework for understanding ship maintenance activities in the U.S. Navy.

The following section, "Ship Maintenance Simulations," illustrates the power of big data visualization software, with data provided by naval ship maintenance metrics groups. It provides examples of how large datasets could be optimized with alternative presentation methods showing a ship's maintenance status, including all operational costs and schedule deviations from planned maintenance. It shows how visualization tools can dig deeper into numbers to improve how key information is summarized and ultimately used in making critical maintenance allocation decisions. Data were collected on 19 U.S. Navy guided missile destroyers (DDG) including 21 maintenance availabilities for those 19 DDGs. Information that was collected included definitized estimates prepared by subject matter experts (SMEs) in the planning process, along with the actual cost and availability data on three maintenance categories. Two simulations were run testing the potential impact of incorporating select technologies on ship maintenance processes. Conclusions and recommendations are presented in the final section.

Literature Review

Big Data

The world is exploding in digital data. IDC Corporation predicts that from 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 zettabytes. Moreover, the digital universe will about double every two years from now to 2020, a 50-fold growth in 10 years as seen in Figure 1 (Gantz & Reinsel, 2012).

More than five billion people are calling, texting, tweeting, and browsing on mobile phones worldwide and 350 million tweets are sent per day (Gantz & Reinsel, 2012). Companies around the world are capturing trillions of bytes of information on customers, suppliers, and operations. The McKinsey Global Institute (MGI; Manyika et al., 2011) estimates that global enterprises stored more than seven exabytes of new data on disk drives in 2010, while consumers stored more than six exabytes of new data on devices such as PCs and notebooks. The U.S. government produced 848 petabytes of data in 2009. Data collected by the U.S. Library of Congress as of April 2011 totals 235 terabytes.

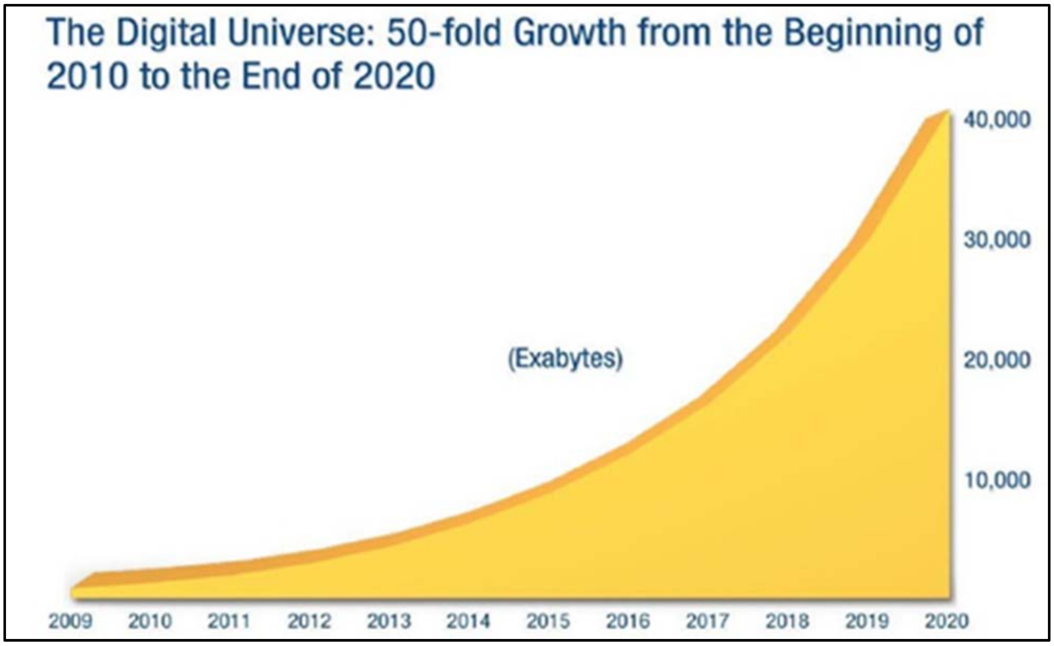


Figure 1. The Digital Universe
(Gantz & Reinsel, 2012)

For the purposes of our research, we will use MGI's definition of big data as "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011). There are many challenges with big data, including the ability to capture, store, curate, search, transfer, share, analyze, and visualize the data. This section focuses on the big data eco structure. It begins with a discussion of the market size, then discusses some of the tools and technologies used in big data analysis, and looks at federal government initiatives involving big data.

The total big data market reached $11.59 billion in 2012 and was estimated to grow at an annual growth rate of 61% to $18.1 billion in 2013, according to Wikibon (Kelly, Floyer, Vellante, & Miniman, 2013). Figure 2 shows revenue by type while Figure 3 gives a breakdown by component. Big data consists of software, hardware, services, and storage.

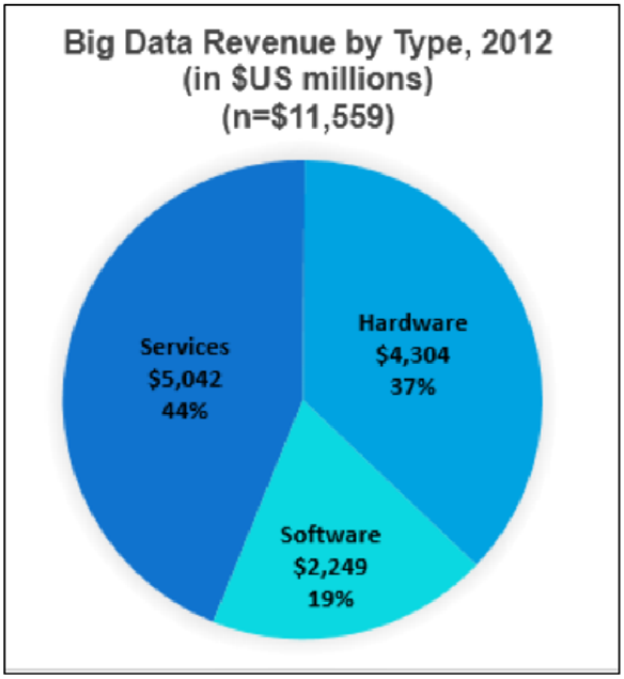


Figure 2. Big Data Revenue by Type
(Kelly et al., 2013)

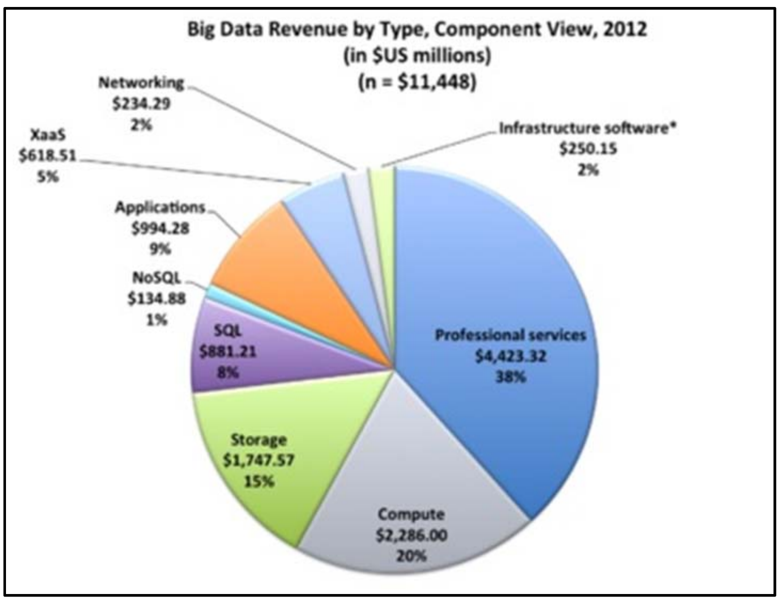
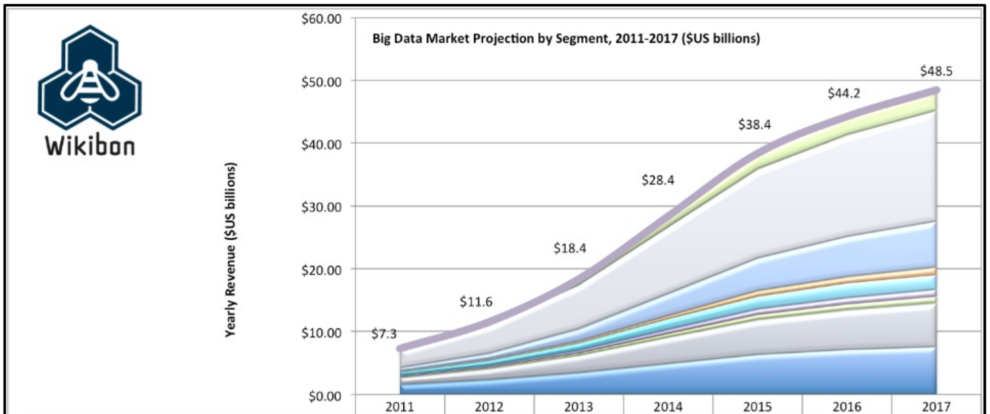


Figure 3. Big Data Revenue by Component
(Kelly et al., 2013)

In addition, Wikibon (Kelly et al., 2013) predicts the big data market to exceed $47 billion by 2017, growing at a 31% compound annual growth rate over the five-year period from 2012 to 2017, as seen in Figure 4.



	2011	2012	2013	2014	2015	2016	2017
Big Data XaaS Revenue	$0.36	$0.62	$1.07	$1.78	$2.52	$2.97	$3.31
Big Data Professional Services Revenue	$2.80	$4.42	$6.98	$10.62	$14.15	$16.17	$17.59
Big Data Application (Analytic and Transactional) Revenue	$0.52	$0.99	$1.90	$3.47	$5.29	$6.48	$7.38
Big Data NoSQL Database Revenue	$0.07	$0.13	$0.27	$0.50	$0.79	$0.98	$1.12
Big Data SQL Database Revenue	$0.62	$0.88	$1.25	$1.72	$2.14	$2.36	$2.51
Big Data Infrastructure Software Revenue	$0.14	$0.24	$0.40	$0.64	$0.88	$1.03	$1.14
Big Data Networking Revenue	$0.15	$0.23	$0.37	$0.56	$0.75	$0.86	$0.93
Big Data Storage Revenue	$1.10	$1.75	$2.76	$4.20	$5.59	$6.39	$6.95
Big Data Compute Revenue	$1.53	$2.29	$3.40	$4.89	$6.26	$7.01	$7.53
Total Big Data Revenue	$7.3	$11.6	$18.4	$28.4	$38.4	$44.2	$48.5

Figure 4. Big Data Market Projection by Segment
(Kelly et al., 2013)

The Big Data Ecosystem

Fueling the growth in big data sales are several factors:

- increased awareness of the benefits of big data as applied to industries beyond the web, most notably financial services, pharmaceuticals, and retail;
- maturation of big data software such as Hadoop, NoSQL (not only structured query language), data stores, in-memory analytic engines, and massively parallel processing analytic databases;
- increasingly sophisticated professional services practices that assist enterprises in practically applying big data hardware and software to business use cases;
- increased investment in big data infrastructure by massive Web properties—most notable Google, Facebook, and Amazon—and government agencies for intelligence and counter-terrorism purposes. (Kelly et al., 2013, Growth Drivers and Adoption Barriers, para. 3))

Wikibon has been tracking the market size, following more than 60 vendors that include both big data pure-plays and others for whom big data is part of multiple revenue sources. Table 1 is a current list of the vendors.

Table 1. 2012 Worldwide Big Data Revenue by Vendor ($US millions)
(Kelly et al., 2013)

Vendor	Big Data Revenue	Total Revenue	Big Data Revenue as % of Total Revenue	% Big Data Hardware Revenue	% Big Data Software Revenue	% Big Data Services Revenue
IBM	$1,306	$103,930	1%	19%	31%	50%
HP	$664	$119,895	1%	34%	29%	38%
Teradata	$435	$2,665	16%	31%	28%	41%
Dell	$425	$59,878	1%	83%	0%	17%
Oracle	$415	$39,463	1%	25%	34%	41%
SAP	$368	$21,707	2%	0%	67%	33%
EMC	$336	$23,570	1%	24%	36%	39%
Cisco Systems	$214	$47,983	0%	58%	0%	42%
PwC	$199	$31,500	1%	0%	0%	100%
Microsoft	$196	$$71,474	0%	0%	67%	33%
Accenture	$194	$29,770	1%	0%	0%	100%
Palantir	$191	$191	100%	0%	36%	64%
Fusion-io	$190	$439	43%	71%	0%	29%
SAS Institute	$187	$2,954	6%	0%	59%	41%
Splunk	$186	$186	100%	0%	71%	29%
Deloitte	$183	$31,300	1%	0%	0%	100%
NetApp	$138	$6,454	2%	77%	0%	23%
Hitachi	$130	$112,318	0%	0%	0%	100%
Opera Solutions	$118	$118	100%	0%	0%	100%
CSC	$114	$15,825	1%	0%	0%	100%
Mu Sigma	$114	$114	100%	0%	0%	100%
Booz Allen Hamilton	$88	$5,802	1%	0%	0%	100%
Amazon	$85	$56,825	0%	0%	0%	100%
TCS	$82	$10,170	1%	0%	0%	100%
Intel	$76	$53,341	0%	83%	0%	17%
Capgemini	$72	$14,020	0%	0%	0%	100%
MarkLogic	$69	$78	88%	0%	63%	38%
Cloudera	$56	$56	100%	0%	47%	53%
Actian	$46	$46	100%	0%	50%	50%

2012 Worldwide Big Data Revenue by Vendor ($US millions)						
Vendor	**Big Data Revenue**	**Total Revenue**	**Big Data Revenue as % of Total Revenue**	**% Big Data Hardware Revenue**	**% Big Data Software Revenue**	**% Big Data Services Revenue**
SGI	$43	$769	6%	83%	0%	17%
GoodData	$38	$38	100%	0%	0%	100%
1010data	$37	$37	100%	0%	0%	100%
10gen	$36	$36	100%	0%	42%	58%
Google	$36	$50,175	0%	0%	0%	100%
Alteryx	$36	$36	100%	0%	55%	45%
Guavus	$35	$35	100%	0%	57%	43%
VMware	$32	$3,676	1%	0%	71%	29%
ParAccel	$24	$24	100%	0%	44%	56%
TIBCO Software	$24	$1,024	2%	0%	53%	47%
Informatica	$24	$812	2%	0%	63%	37%
MapR	$23	$23	100%	0%	51%	49%
Pervasive Software	$22	$51	37%	0%	41%	59%
Attivio	$21	$26	80%	0%	62%	38%
Fractal Analytics	$20	$20	100%	0%	0%	100%
Hortonworks	$18	$18	100%	0%	50%	50%
Rackspace	$18	$1,300	1%	0%	0%	100%
QlikTech	$16	$321	5%	0%	74%	26%
DataStax	$15	$15	100%	0%	59%	41%
Basho	$14	$14	100%	0%	63%	38%
Microstrategy	$13	$595	2%	0%	59%	41%
Tableau Software	$13	$130	10%	0%	59%	41%
Kognitio	$13	$12	100%	0%	47%	53%
Couchbase	$12	$12	$100%	0%	64%	36%
Datameer	$10	$10	100%	0%	80%	20%
LucidWorks	$9	$9	100%	0%	60%	40%
Digital Reasoning	$10	$10	100%	0%	51%	49%

2012 Worldwide Big Data Revenue by Vendor ($US millions)						
Vendor	**Big Data Revenue**	**Total Revenue**	**Big Data Revenue as % of Total Revenue**	**% Big Data Hardware Revenue**	**% Big Data Software Revenue**	**% Big Data Services Revenue**
Aerospike	$9	$9	100%	0%	80%	20%
Neo Technology	$9	$9	100%	0%	62%	38%
Think Big Analytics	$8	$8	100%	0%	0%	100%
Calpont	$8	$8	100%	0%	60%	40%
RainStor	$8	$8	100%	0%	67%	33%
SiSense	$7	$7	100%	0%	40%	60%
Revolution Analytics	$7	$13	56%	0%	55%	45%
Talend	$6	$51	12%	0%	80%	20%
Jaspersoft	$6	$31	20%	0%	62%	38%
Juniper Networks	$6	$4,365	0%	70%	0%	30%
Pentaho	$6	$31	19%	0%	62%	38%
DDN	$4	$278	2%	63%	0%	38%
Actuate	$5	$137	3%	0%	63%	37%
Original Device Manufacturers	$2,375	$100,000	2%	100%	0%	0%
Other	$1,613	$197,170	1%	17%	13%	70%
Total	**$11,565**	**$1,244,602**	**1%**	**37%**	**19%**	**44%**

Big data is generated by a variety of sources. Big data originates from sources including industry specific transactions, machine/sensor indications, web applications, and text (Ferguson, 2013). Industry specific transactions can include call records and geographic location data. Machines generate extremely large volumes of information every day and can range in complexity from simple temperature readings to the performance parameters of a gas-turbine engine. Big data on the web also ranges in format from machine language to customer comments on social networks and is also produced in considerably sizeable portions. Text sources can include archived documents, external reports, or customer account information (Ferguson, 2013).

Because big data comes from a variety of sources, it also possesses characteristics that distinguish it from data in the traditional context. Common terms used to define the qualities of big data include volume, variety, velocity, and value (Dijcks, 2013). From this listing of sources, one can understand that the *volume* of data generated on a daily basis is enormous. For example, Dijcks (2013) stated that just a single jet engine produces 10 terabytes of data in 30 minutes. Extrapolate that example to include all the aircraft currently airborne, and then include all the factory infrastructure around the globe collecting data on production, service life, and maintenance requirements, and the enormity of big data volumes begins to emerge. Another characteristic of big data, *variety*, can be directly translated from the various sources into the variety of data formats. Various data formats require additional consideration to ensure the ability of all systems to share data. *Velocity*, which is related to volume, is the frequency with which big data is created. To illustrate velocity, consider the relative size of a single Twitter feed (140 characters) to the large number of feeds generated in a given time period (Dijcks, 2013). Finally, *value* is the feature of big data that is important to any enterprise. Refer to Appendix A for a paper regarding the implications of big data on enterprise architecture (EA), the information technology infrastructure of an organization.

Big Data Technologies and Tools

Many techniques can be used to analyze datasets. These techniques, which often draw upon statistics and computer science, can be applied to big data to generate insights into large and diverse datasets, as well as smaller, diverse datasets. Table 2 summarizes some techniques.

Table 2. Big Data Analysis Techniques
(Manyika et al., 2011)

A/B testing	• Technique in which a control group is compared with a variety of test groups in order to determine what treatments (i.e., changes) will improve a given objective variable. • Big data enables huge numbers of tests to be executed and analyzed, ensuring that groups are of sufficient size to detect meaningful (i.e., statistically significant) differences between the control and treatment groups.
Association rule learning	• Set of techniques for discovering interesting relationships (i.e., “association rules”) among variables in large databases. • These techniques consist of a variety of algorithms to generate and test possible rules. • An application is market basket analysis, in which a retailer can determine which products are frequently bought together and use this information for marketing (a commonly cited example is the discovery that many supermarket shoppers who buy diapers also tend to buy beer). • Used for data mining.

Classification	• Set of techniques to identify the categories in which new data points belong, based on a training set containing data points that have already been categorized. • One application is the prediction of segment-specific customer behavior (e.g., buying decisions, churn rate, consumption rate) where there is a clear hypothesis or objective outcome. • These techniques are often described as supervised learning because of the existence of a training set; they stand in contrast to cluster analysis, a type of unsupervised learning. • Used for data mining.
Cluster analysis	• Statistical method for classifying objects that splits a diverse group into smaller groups of similar objects, whose characteristics of similarity are not known in advance. • An example of cluster analysis is segmenting consumers into self-similar groups for targeted marketing. • This is a type of unsupervised learning because training data are not used. • Used for data mining.
Crowdsourcing	• Technique for collecting data submitted by a large group of people or community (i.e., the "crowd") through an open call, usually through networked media such as the Web. • This is a type of mass collaboration and an instance of using Web 2.0.
Data fusion and data integration	• Set of techniques that integrate and analyze data from multiple sources in order to develop insights in ways that are more efficient and potentially more accurate than if they were developed by analyzing a single source of data. • Signal processing techniques can be used to implement some types of data fusion. • One example of an application is sensor data from the *Internet of Things* being combined to develop an integrated perspective on the performance of a complex distributed system such as an oil refinery. • Data from social media, analyzed by natural language processing, can be combined with real-time sales data, in order to determine what effect a marketing campaign is having on customer sentiment and purchasing behavior.
Data mining	• Set of techniques to extract patterns from large datasets by combining methods from statistics and machine learning with database management. • These techniques include association rule learning, cluster analysis, classification, and regression. • Applications include mining customer data to determine segments most likely to respond to an offer, mining human resources data to identify characteristics of most successful employees, or market basket analysis to model the purchase behavior of customers.
Ensemble learning	• Using multiple predictive models (each developed using statistics and/or machine learning) to obtain better predictive performance than could be obtained from any of the constituent models. • This is a type of supervised learning.
Genetic algorithms	• Technique used for optimization that is inspired by the process of natural evolution or "survival of the fittest." • Potential solutions are encoded as "chromosomes" that can combine and mutate. • These individual chromosomes are selected for survival within a modeled "environment" that determines the fitness or performance of each individual in the population. • Often described as a type of "evolutionary algorithm," these algorithms are well-suited for solving nonlinear problems. • Examples of applications include improving job scheduling in manufacturing and optimizing the performance of an investment portfolio.
Machine learning	• Subspecialty of computer science (within a field historically called "artificial intelligence") concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data. • A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. Natural language processing is an example of machine learning.

Natural language processing (NLP)	• Set of techniques from a subspecialty of computer science (within a field historically called “artificial intelligence”) and linguistics that uses computer algorithms to analyze human (natural) language. • Many NLP techniques are types of machine learning. • One application of NLP is using sentiment analysis on social media to determine how prospective customers are reacting to a branding campaign.
Neural networks	• Computational models, inspired by the structure and workings of biological neural networks (i.e., the cells and connections within a brain), that find patterns in data. • Neural networks are well-suited for finding nonlinear patterns. • Can be used for pattern recognition and optimization. Some neural network applications involve supervised learning and others involve unsupervised learning. • Examples of applications include identifying high-value customers that are at risk of leaving a particular company and identifying fraudulent insurance claims.
Network analysis	• Set of techniques used to characterize relationships among discrete nodes in a graph or a network. • In social network analysis, connections between individuals in a community or organization are analyzed (e.g., how information travels) or who has the most influence over whom. • Examples of applications include identifying key opinion leaders to target for marketing and identifying bottlenecks in enterprise information flows.
Optimization	• Portfolio of numerical techniques used to redesign complex systems and processes to improve their performance according to one or more objective measures (e.g., cost, speed, or reliability). • Examples of applications include improving operational processes such as scheduling, routing, and floor layout, and making strategic decisions such as product range strategy, linked investment analysis, and R&D portfolio strategy. • Genetic algorithms are an example of an optimization technique.
Pattern recognition	• Set of machine learning techniques that assign some sort of output value (or label) to a given input value (or instance) according to a specific algorithm. • Classification techniques are an example.
Predictive modeling	• A set of techniques in which a mathematical model is created or chosen to best predict the probability of an outcome. • Example of an application in customer relationship management is the use of predictive models to estimate the likelihood that a customer will “churn” (i.e., change providers) or the likelihood that a customer can be cross-sold another product. • Regression is one example of the many predictive modeling techniques.
Regression	• Set of statistical techniques to determine how the value of the dependent variable changes when one or more independent variables is modified. • Often used for forecasting or prediction. • Examples of applications include forecasting sales volumes based on various market and economic variables or determining what measurable manufacturing parameters most influence customer satisfaction. • Used for data mining.
Sentiment analysis	• Application of natural language processing and other analytic techniques to identify and extract subjective information from source text material. • Key aspects of these analyses include identifying the feature, aspect, or product about which a sentiment is being expressed, and determining the type, “polarity” (i.e., positive, negative, or neutral), and the degree and strength of the sentiment. • Examples of applications include companies applying sentiment analysis to analyze social media (e.g., blogs, microblogs, and social networks) to determine how different customer segments and stakeholders are reacting to their products and actions.
Signal processing	• Set of techniques from electrical engineering and applied mathematics originally developed to analyze discrete and continuous signals (i.e., representations of analog physical quantities [even if represented digitally] such as radio signals, sounds, and images). • This category includes techniques from signal detection theory, which quantifies the ability to discern between signal and noise. • Sample applications include modeling for time series analysis or implementing data fusion to determine a more precise reading by combining data from a set of less precise data sources (i.e., extracting the signal from the noise).

Spatial analysis	• Set of techniques, some applied from statistics, which analyze the topological, geometric, or geographic properties encoded in a dataset. • Often the data for spatial analysis come from geographic information systems (GIS) that capture data including location information (e.g., addresses or latitude/longitude coordinates). • Examples of applications include the incorporation of spatial data into spatial regressions (e.g., how is consumer willingness to purchase a product correlated with location?) or simulations (e.g., how would a manufacturing supply chain network perform with sites in different locations?).
Statistics	• Science of the collection, organization, and interpretation of data, including the design of surveys and experiments. • Statistical techniques are often used to make judgments about what relationships between variables could have occurred by chance (the “null hypothesis”) and what relationships between variables likely result from some kind of underlying causal relationship (i.e., that are “statistically significant”). • Statistical techniques are also used to reduce the likelihood of Type I errors (“false positives”) and Type II errors (“false negatives”). • Example of an application is A/B testing to determine what types of marketing material will most increase revenue.
Supervised learning	• Set of machine learning techniques that infer a function or relationship from a set of training data. • Examples include classification and support vector machines.
Simulation	• Modeling the behavior of complex systems, often used for forecasting, predicting, and scenario planning. Monte Carlo simulations, for example, are a class of algorithms that rely on repeated random sampling (i.e., running thousands of simulations, each based on different assumptions). • Result is a histogram that gives a probability distribution of outcomes. • One application is assessing the likelihood of meeting financial targets given uncertainties about the success of various initiatives.
Time series analysis	• Set of techniques from both statistics and signal processing for analyzing sequences of data points, representing values at successive times, to extract meaningful characteristics from the data. • Examples of time series analysis include the hourly value of a stock market index or the number of patients diagnosed with a given condition every day. • Time series forecasting is the use of a model to predict future values of a time series based on known past values of the same or other series. • Some of these techniques (e.g., structural modeling) decompose a series into trend, seasonal, and residual components, which can be useful for identifying cyclical patterns in the data. • Examples of applications include forecasting sales figures, or predicting the number of people who will be diagnosed with an infectious disease.
Unsupervised learning	• Set of machine learning techniques that finds hidden structure in unlabeled data. • Cluster analysis is an example of unsupervised learning (in contrast to supervised learning).
Visualization	• Techniques used for creating images, diagrams, or animations to communicate, understand, and improve the results of big data analyses.

There are a growing number of technologies used to aggregate, manipulate, manage, and analyze big data. Some of the more widely used technologies used to aggregate, manage and analyze big data are found in Table 3.

Table 3. Big Data Analysis Technologies
(Manyika et al, 2011)

TECHNOLOGY	COMMENTS
Big Table	• Proprietary distributed database system built on the Google File System. • Inspiration for HBase.
Business Intelligence	• A type of application software designed to report, analyze, and present data. • Often used to read data previously stored in a data warehouse or data mart. • Also used to create standard reports that are generated on a periodic basis, or to display information on real-time management dashboards (i.e., integrated displays of metrics that measure the performance of a system).
Cassandra	• An open source (free) database management system designed to handle huge amounts of data on a distributed system. • System was originally developed at Facebook and is now managed as a project of the Apache Software foundation.
Cloud Computing	• A computing paradigm in which highly scalable computing resources, often configured as a distributed system, are provided as a service through a network.
Data Mart	• Subset of a data warehouse, used to provide data to users usually through business intelligence tools.
Data Warehouse	• Specialized database optimized for reporting, often used for storing large amounts of structured data. • Data uploaded using ETL (extract, transform, and load) tools from operational data stores, and reports are often generated using business intelligence tools.
Distributed System	• Multiple computers, communicating through a network, used to solve a common computational problem. • Problem is divided into multiple tasks, each of which is solved by one or more computers working in parallel. • Benefits of distributed systems include higher performance at a lower cost (i.e., because a cluster of lower end computers can be less expensive than a single higher end computer), higher reliability (i.e., because of a lack of a single point of failure), and more scalability (i.e., because increasing the power of a distributed system can be accomplished by simply adding more nodes rather than completely replacing a central computer).
Dynamo	• Proprietary distributed data storage system developed by Amazon.
Extract, Transform, and Load (ETL)	• Software tools used to extract data from outside sources, transform them to fit operational needs, and load them into a database or data warehouse.
Google File System	• Proprietary distributed file system developed by Google; part of the inspiration for Hadoop.31
Hadoop	• Open source software framework for processing huge datasets on certain kinds of problems on a distributed system. Its development was inspired by Google's MapReduce and Google File System. It was originally developed at Yahoo! and is now managed as a project of the Apache Software Foundation.
HBase	• Open source, distributed, non-relational database modeled on Google's Big Table. • Originally developed by Powerset and is now managed as a project of the Apache Software foundation as part of the Hadoop.
MapReduce	• Software framework introduced by Google for processing huge datasets on certain kinds of problems on a distributed system. • Also implemented in Hadoop.

Mashup	• Application that uses and combines data presentation or functionality from two or more sources to create new services. • Applications are often made available on the Web, and frequently use data accessed through open application programming interfaces or from open data sources.
Metadata	• Data that describes the content and context of data files (e.g., means of creation, purpose, time and date of creation, and author).
Non-Relational Database	• A database that does not store data in tables (rows and columns).
R	• Open source (free) programming language and software environment for statistical computing and graphics. • R language has become a de facto standard among statisticians for developing statistical software and is widely used for statistical software development and data analysis.
Relational Database	• Database made up of a collection of tables (relations; i.e., data are stored in rows and columns). • Relational database management systems (RDBMS) store a type of structured data. • SQL is the most widely used language for managing relational databases.
Semi-Structured Data	• Data that do not conform to fixed fields but contain tags and other markers to separate data elements. • Examples include XML- or HTML-tagged text.
SQL	• Originally an acronym for structured query language, SQL is a computer language designed for managing data in relational databases. • Technique includes the ability to insert, query, update, and delete data, as well as manage data schema (database structures) and control access to data in the database.
Stream Processing	• Technologies designed to process large real-time streams of event data. • Enables applications such as algorithmic trading in financial services, RFID event processing applications, fraud detection, process monitoring, and location-based services in telecommunications.
Structured Data	• Data that reside in fixed fields. • Examples include relational databases or data in spreadsheets.
Unstructured Data	• Data that do not reside in fixed fields. • Examples include free-form text (e.g., books, articles, body of e-mail messages) and untagged audio, image and video data.
Visualization	• Technologies used for creating images, diagrams, or animations to communicate a message that are often used to synthesize the results of big data analyses.

In working with massive amounts of data, displaying summary data and using visualization is critical to finding connections and relevance among millions of parameters and variables to convey linkages, hypotheses, metrics, and project future outcomes. Taken one level further, Interactive Visualization moves visualization from static spreadsheets and graphics to images capable of drilling down for more details, and immediately changing how data are presented and processed.

Examples of visualization methods include the following:

- Bar charts are commonly used for comparing the quantities of different categories or groups. An example is shown in Figure 5.

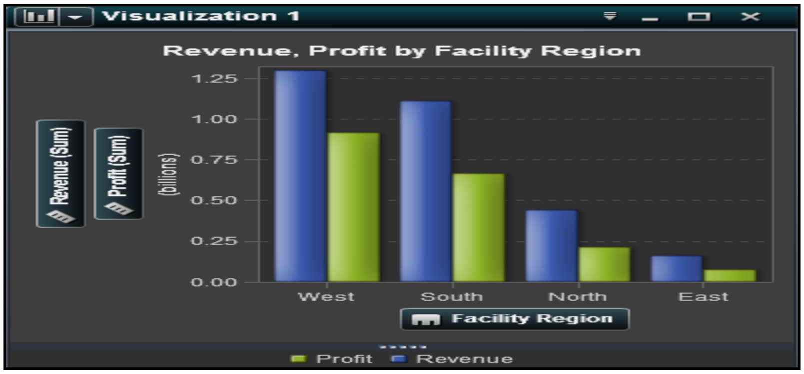


Figure 5. Bar Chart
(Choy, Chawla, & Whitman, 2012)

- Box plots (see Figure 6) represent a distribution of data values. They display five statistics of minimum, lower quartile, median, upper quartile, and the maximum values that summarize the distribution of a set of data. Extreme values are represented by whiskers extending from the edges of the box.

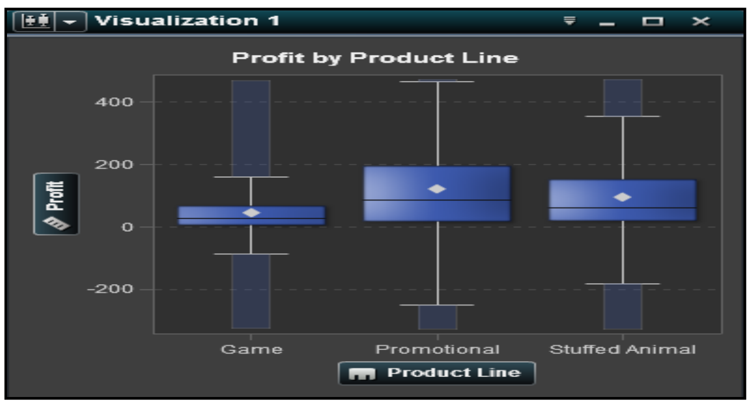


Figure 6. Box Plot
(Choy et al., 2012)

- Bubble plots (see Figure 7) are variations of a scatter plot in which the data markers are replaced with bubbles, with each bubble representing an observation (or group of observations). They are useful for datasets with many values or when values differ by orders of magnitude.

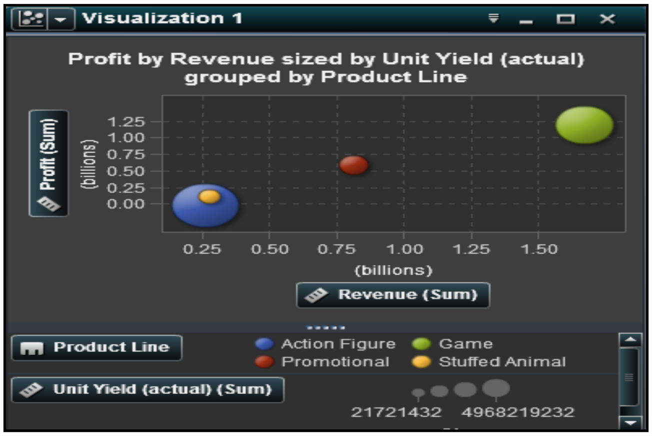


Figure 7. Bubble Plot
(Choy et al., 2012)

- Correlation matrices (see Figure 7) combine big data with fast response times to identify quickly which variables among millions/billions are related. They also show the relationship strength between variables.

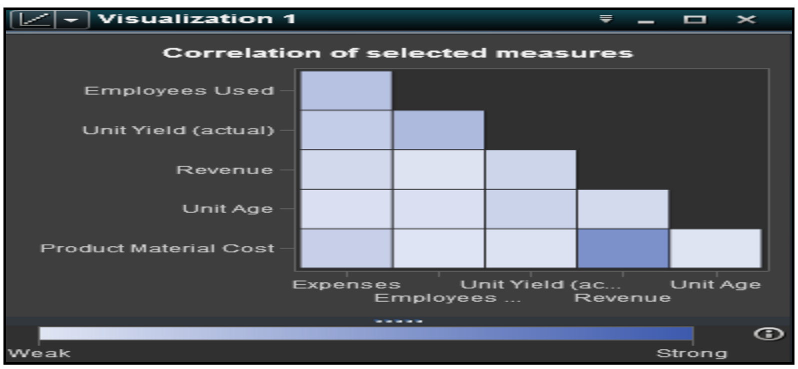


Figure 8. Correlation Matrix
(Choy et al., 2012)

- Cross-tabulation charts (see Figure 9) show frequency distributions or other aggregate statistics for the intersections of two or more category data items. Crosstabs enable examination of data for intersections of hierarchy nodes or category values.

Type	DriveTrain	Asia Horsepower	Asia MPG (Highway)	Europe Horsepower
Hybrid	Front	92	56	
SUV	All	218	22	26
	Front	208	22	
Sedan	All	192	27	24
	Front	171	31	20
	Rear	273	24	26
Sports	All	264	26	26
	Front	209	28	31
	Rear	226	26	32
Truck	All	209	20	
	Rear	158	25	
Wagon	All	200	26	25
	Front			

Figure 9. Cross-Tabulation Chart
(Choy et al., 2012)

- Clustergrams (see Figure 10) display how individual members of a dataset are assigned to clusters as the number of members increases.

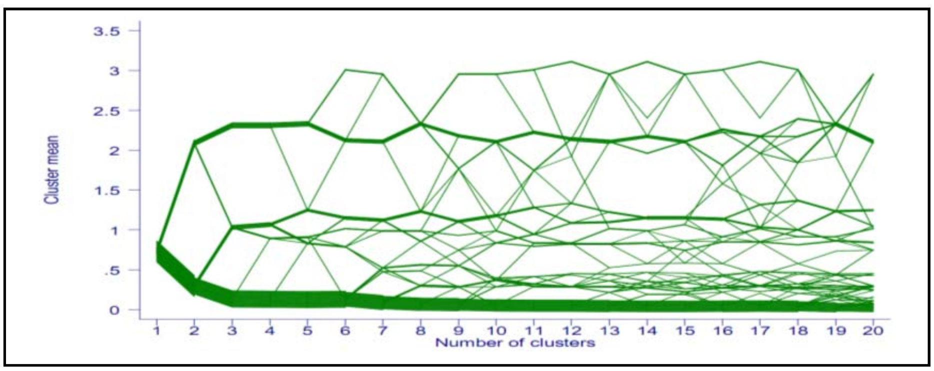


Figure 10. Clustergram
(Manyika et al., 2011)

- Geo maps (see Figure 11) display data as a bubble plot overlaid on a geographic map. Each bubble is located either at the center of a geographic region or at location coordinates.

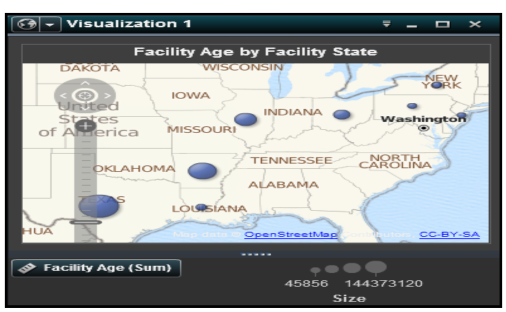


Figure 11. Geo Map
(Choy et al., 2012)

- Heat maps (see Figure 12) display a distribution of values for two data items using a table with colored cells. Colors are used to communicate relationships between data values.

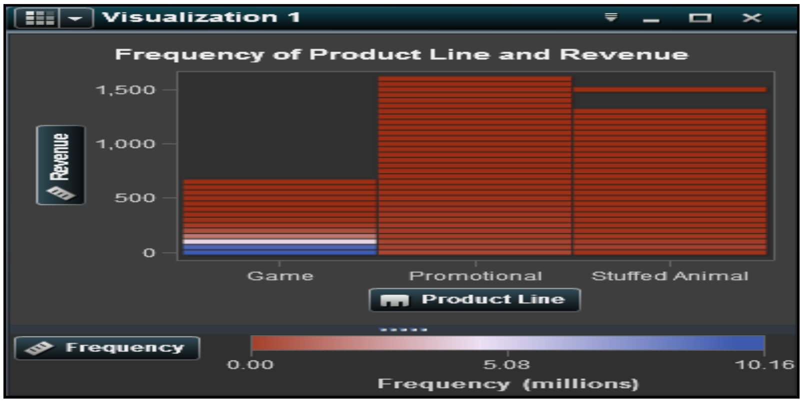


Figure 12. Heat Map
(Choy et al., 2012)

- Histograms (see Figure 13) are variations of bar charts using rectangles to show the frequency of data items in successive numerical intervals of equal size. They are often used to quickly show distribution of values in large datasets.

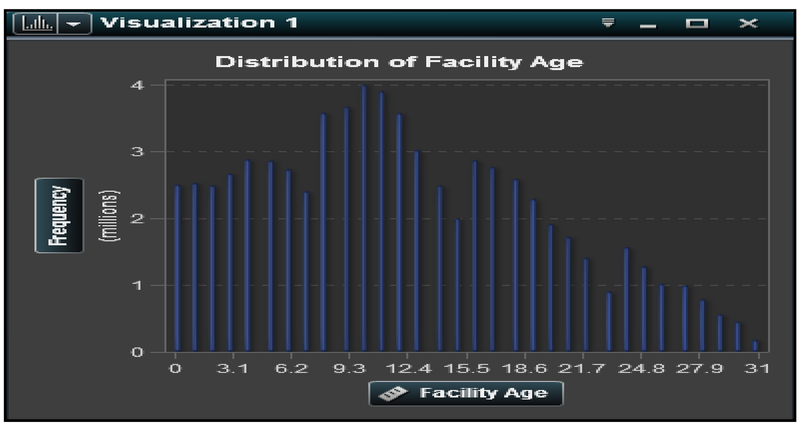


Figure 13. Histogram
(Choy et al., 2012)

- History flow charts (see Figure 14) show the evolution of a document edited by multiple contributing authors. Time appears on the horizontal axis, while contributions to the text are on the vertical axis; each author has a different color code and the vertical length of a bar indicates the amount of text written by each author.

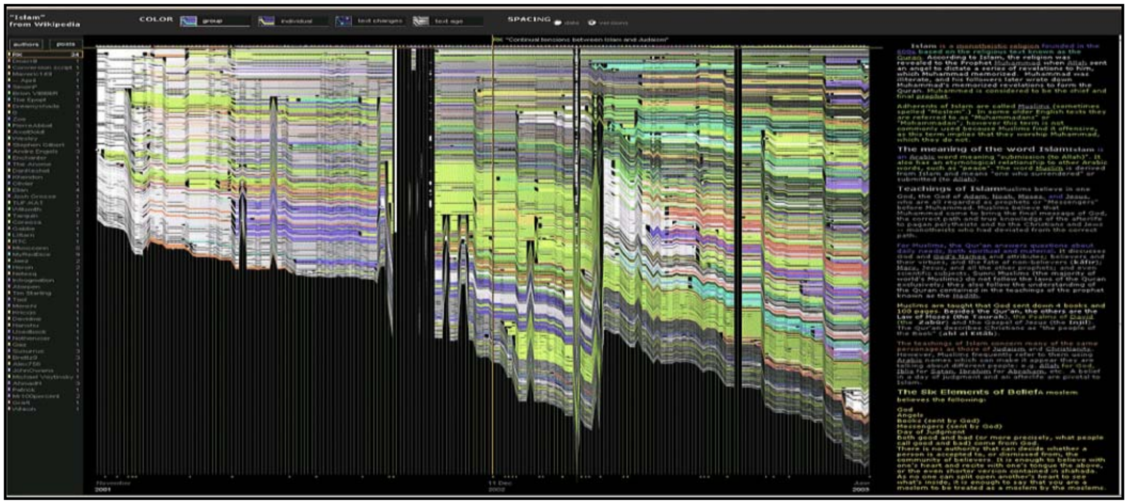


Figure 14. History Flow
(Manyika et al., 2011)

- Line charts (see Figure 15) show the relationship of one variable to another by using a line that connects the data values. They are most often used to track changes or trends over time.

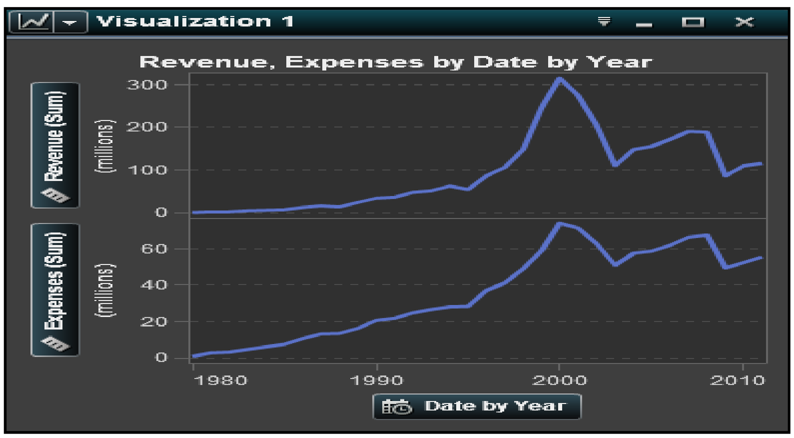


Figure 15. Line Chart
(Choy et al., 2012)

- Pareto charts (see Figure 16) are a specialized type of vertical bar chart where values of the dependent variables are plotted in decreasing order of frequency from left to right. They are used to quickly identify when certain issues need attention.

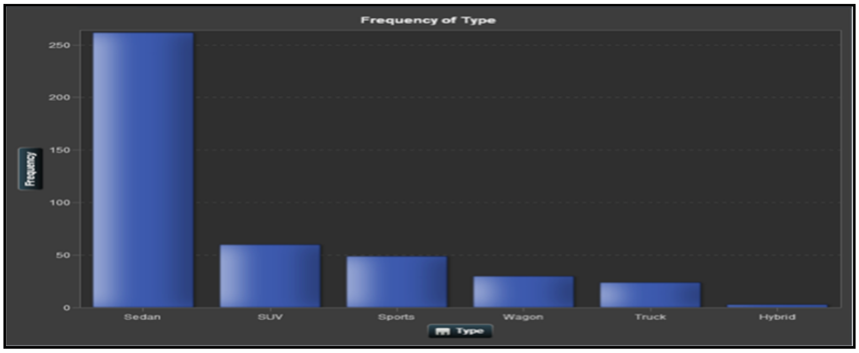


Figure 16. Pareto Chart
(Choy et al., 2012)

- Scatter plots (see Figure 17) are two-dimensional plots showing joint variation of two (or three) variables from a group of table rows. They are useful for examining the relationships, or correlations, between numeric data items.

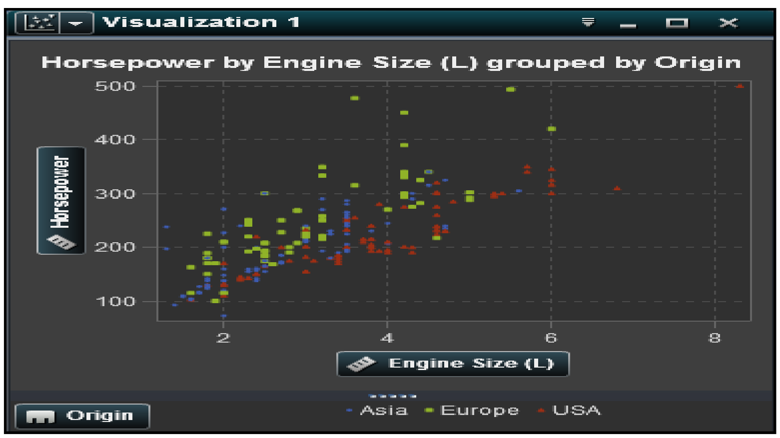


Figure 17. Scatter Plot
(Choy et al., 2012)

- Tag clouds (see Figure 18) are weighted visual lists in which words appearing most frequently are larger and words appearing less frequently are smaller.

Figure 18. Tag Cloud
(Manyika et al., 2011)

- Tree maps (see Figure 19) are a variation of heat maps using rectangles (tiles) to represent data components. The largest rectangle represents the dominant division of the data and smaller rectangles represent subdivisions.

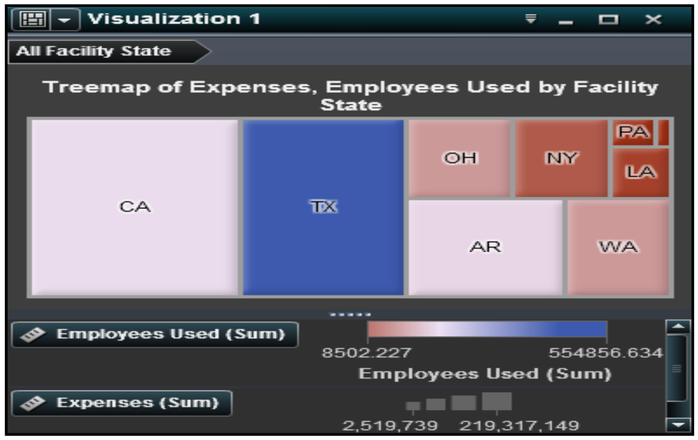


Figure 19. Tree Map
(Choy et al., 2012)

Government Spending on Big Data

The federal government is fueling the growth of big data spending on national security and military applications. According to the Biometrics Research Group (King, 2013), federal agencies spent approximately US$5 billion on big data resources in fiscal year (FY) 2012, and they estimate annual spending will grow to US$6 billion in 2014. By 2017, that figure will reach US$8 billion, growing at a compound annual growth rate of 10% as shown in Figure 20.

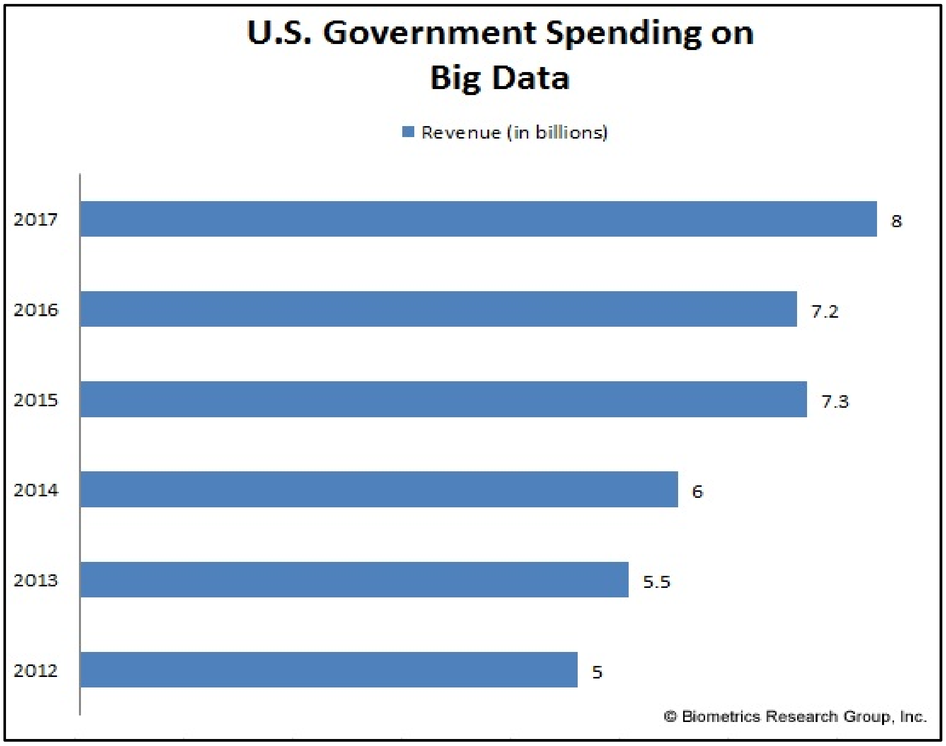


Figure 20. U.S. Government Spending on Big Data
(King, 2013)

During the near to midterm, Biometrics Research Group (King, 2013) predicts that most of the spending will be on military applications of the U.S. government with federal agencies pursuing more than 150 big data projects (grants, procurements, or related activities). The agency leading big data research is the U.S. DoD, with more than 30 projects and, in particular, the Defense Advanced Research Projects Agency (DARPA) with nine major projects (King, 2013).

In a recent study sponsored by EMC (King, 2013) that surveyed 150 U.S. government information technology (IT) executives, 70% of respondents stated that big data will be critical to all government operations within five years. Big data, according to the survey, has the potential to save nearly $500 billion, or 14%, of agency budgets across the federal government by increasing efficiency, enabling smarter decisions, and deepening insight. However, only 31% of respondents said their agency has an adequate big data strategy (King, 2013).

Government agencies are seeking to make big data a greater part of their mission. The Department of Homeland Security (DHS) posted a solicitation July 24, 2013 (DHS, 2013), requesting additional information from industry in order to identify transformational opportunities to improve mission and operational efficiencies and lower costs through advanced analytic automation for the DHS and Homeland Security Enterprise (HSE). The request for information (RFI) read as follows:

> The purpose of this RFI is to ascertain available sources to provide widely used big data infrastructure, computing, storage, analytics, and visualization capabilities that are based on open source or commonly available commercial technologies and represent technology options of high value to the future of homeland security.

Big Data Projects in Government

In 2012, the Obama administration announced the *Big Data Research and Development Initiative* to help solve challenges by improving the ability to extract knowledge and insights from large and complex collections of digital data (Office of Science and Technology Policy, 2012). The initiative's objective is to analyze big data and achieve advances in several sectors, such as healthcare, security, the environment, education, and the sciences. Six federal departments and agencies launched the initiative with more than $200 million in commitments that promise to greatly improve the tools and techniques needed to access, organize, and glean discoveries from huge volumes of digital data.

The Big Data Research and Development Initiative was created to achieve the following:

- Advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyze, and share huge quantities of data;
- Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning; and
- Expand the workforce needed to develop and use big data technologies (Office of Science and Technology Policy, 2012, p. 1).

The DoD announced plans to invest approximately $250 million annually across the military departments in a series of programs that will

- Harness and utilize massive data in new ways and bring together sensing, perception, and decision support to make truly autonomous systems that can maneuver and make decisions on their own.
- Improve situational awareness to help warfighters and analysts and provide increased support to operations. The Department is seeking a 100-fold increase in the ability of analysts to extract information from texts in any language, and a similar increase in the

number of objects, activities, and events that an analyst can observe (Office of Science and Technology Policy, 2012, pp. 2-3).

According to King (2013), DoD big data programs include XDATA, Cyber-Insider Threat (CINDER), Anomaly Detection at Multiple Scales (ADAMS), Insight, Mind's Eye, Machine Reading, Mission-Oriented Resilient Clouds, Programming Computation on Encrypted Data (PROCEED), and Video and Image Retrieval and Analysis Tool (VIRAT).

XDATA is a four-year, $25 million per-year program to develop computational techniques and software tools for analyzing large volumes of data, both semi-structured (e.g., tabular, relational, categorical, meta-data) and unstructured (e.g., text documents, message traffic). Some core challenges include developing scalable algorithms for processing imperfect data in distributed data stores and effective human-computer interaction tools that are rapidly customizable to facilitate visual reasoning for diverse missions. XDATA envisions open source software toolkits for flexible software development, enabling processing of large volumes of data for use in targeted defense applications (King, 2013, para. 13).

The CINDER program seeks to develop innovative approaches to detect activities consistent with cyber espionage in military computer networks. CINDER will apply various models of adversary missions to *normal* activity on internal networks as a method to expose hidden operations. The program also intends to increase the accuracy, rate, and speed with which cyber threats are detected (King, 2013, para. 6).

The ADAMS program addresses the issue of anomaly detection and characterization in massive datasets. Data anomalies are intended to cue the collection of additional, actionable information in a wide variety of real-world contexts. Initially, ADAMS will focus on insider threat detection, in which anomalous actions by an individual are detected against a background of routine network activity (King, 2013, para. 5).

The Insight program addresses key shortfalls in current intelligence, surveillance, and reconnaissance systems. Automation and integrated human-machine reasoning enable operators to analyze greater numbers of potential threats ahead of time-sensitive situations. This program seeks to develop a resource management system that automatically identifies threat networks and irregular warfare operations by the analysis of information from imaging and non-imaging sensors and other sources (King, 2013, para. 7).

The Mind's Eye program seeks to develop a capability for *visual intelligence* in machines. Unlike the traditional study of machine vision where progress has been made in recognizing a wide range of objects and their properties or the nouns in the description of a scene, Mind's Eye seeks to add the perceptual and cognitive underpinnings needed for recognizing and reasoning about the verbs in those scenes. Collectively, these technologies could enable a more complete visual narrative (King, 2013, para. 9).

The Machine Reading program seeks to realize artificial intelligence applications by developing learning systems that process natural text and insert the resulting semantic representation into a knowledge base rather than relying on expensive and time-consuming current processes for knowledge representation that require expert and associated knowledge engineers to hand craft information (King, 2013, para. 8).

The Mission-Oriented Resilient Clouds program aims to address security challenges inherent in cloud computing by developing technologies to detect, diagnose, and respond to attacks (King, 2013, para. 10).

The PROCEED research effort targets a major challenge for information security in cloud-computing environments by developing practical methods and associated modern programming languages for computation on data that remains encrypted the entire time it is in use. Interception by an adversary would be more difficult if users had the ability to manipulate encrypted data without first decrypting (King, 2013, para. 11).

The VIRAT program aims to develop a system to provide military imagery analysts with the capability to exploit the vast amount of overhead video content being collected. If it is successful, VIRAT will enable analysts to establish alerts for activities and events of interest as they occur. Tools will also be developed to enable analysts to rapidly retrieve, with high precision and recall, video content from extremely large video libraries (King, 2013, para. 12).

Government Big Data Case Studies

Government agencies have implemented big data projects to transform agencies' processes and procedures. The U.S. Army, for example, is already leveraging big data technologies in conjunction with cloud computing (Conway, 2012). Started in April 2009, the U.S. Army's Big Data Cloud program extends to forward operating bases, which can double as local nodes that collect data from various sources. The private cloud, which went live in March 2011, conveys the latest intelligence information to U.S. troops in Afghanistan in real or near-real time (Conway, 2012).

The National Archive and Records Administration (NARA) challenge is to digitize a huge volume of unstructured data to provide quick access while maintaining the data in both classified and unclassified environments

(TechAmerica Foundation, 2012. NARA is charged with providing the Electronic Records Archive (ERA) and online, public access systems for U.S. records and documentary heritage. In January 2012, NARA managed approximately 142 terabytes of information, consisting of more than seven billion objects and incorporating records from across the federal agencies, Congress, and several presidential libraries. There are more than 350 million annual hits on its website. In addition to managing the ERA, NARA must digitize more than four million cubic feet of traditional archival holdings, including about 400 million pages of classified information scheduled for declassification, pending review with the intelligence community (TechAmerica Foundation, 2012).

NARA used big data tools to address those challenges. In conjunction with traditional data capture, digitizing, and storage capabilities, advanced big data capabilities were used for search, retrieval, and presentation, all while supporting strict security guidelines. Faster result ingestion and categorization of documents, improved end user experience, and dramatically reduced storage costs were the results (TechAmerica Foundation, 2012). Other big data cases involving government agencies are summarized in Table 4.

Table 4. **High-Level Summary of Case Studies**
(TechAmerica Foundation, 2012)

Agency/Org/Co. Big Data Project Name	Underpinning Technologies	Big Data Metrics	Initial Big Data Entry Point	Public/ User Benefits
NARA ERA	Metadata, Submission, Access, Repository, Search, and Taxonomy applications for storage and archival systems	Petabytes, Terabytes/sec, Semi-structured	• Warehouse Optimization • Distributed Info Mgt	Provides ERA and Online Public Access systems for U.S. records & documentary heritage
National Aeronautics and Space Administration (NASA) Human Space Flight Imagery	Metadata, Archival, Search, and Taxonomy applications for tape library systems, government off-the-shelf (GOTS)	Petabytes, Terabytes/sec, Semi-structured	Warehouse Optimization	Provide industry and the public with iconic and historic human spaceflight imagery for scientific discovery, education, and entertainment
National Oceanic and Atmospheric Administration (NOAA) National Weather Service	HPC modeling; data from satellites, ships, aircraft, and deployed sensors	Petabytes, Terabytes/sec, Semi-structured, ExaFLOPS, PetaFLOPS	Streaming Data & Analytics, Warehouse Optimization, Distributed Info Mgt	Provide weather, water, and climate data, and forecasts and warnings for the protection of life and property and enhancement of the national economy.
Internal Revenue Service (IRS) Compliance Data Warehouse	Columnar database architecture; multiple analytics applications; descriptive, exploratory, and predictive analysis	Petabytes	Streaming Data & Analytics, Warehouse Optimization, Distributed Info Mgt	Provide taxpayers top quality service by helping them to understand and meet their tax responsibilities and enforce the law with integrity and fairness.
Centers for Medicare & Medicaid Services (CMS) Medical Records Analytics	Columnar and NoSQL databases, Hadoop being looked at, EHR on the front end, with legacy structured database systems (including DB2 and COBOL)	Petabytes, Terabytes/day	Streaming Data & Analytics, Warehouse Optimization, Distributed Info Mgt	Protect the health of all Americans and ensure compliant processing of insurance claims

Lessons Learned

It is useful to better understand big data, this somewhat ambiguous concept, by taking advantage of lessons learned by other organizations dealing with similar problems. The TechAmerica Foundation Big Data Commission released a study in October 2012 on how big data can move beyond the tidal wave of data and transform government. The Commission's mandate was to demystify the term *big data* by defining its characteristics, describing the key business outcomes it will serve, and providing a framework for policy discussion. Its goal was to provide guidance to federal government's senior policy- and decision-makers.

The Commission identified a number of lessons learned from early government big data initiatives (TechAmerica Foundation, 2012):

- The path towards becoming big data "capable" will be iterative and cyclical.
- Successful big data initiatives seem to begin with a burning business or mission requirement that government leaders are unable to address with traditional approaches.
- Successful big data initiatives commonly start with a specific and narrowly defined business or mission requirement, and not a plan to deploy a new and universal technical platform to support perceived future requirements.
- Successful initiatives seek to address the initial set of use cases by augmenting current IT investments, but do so with an eye to leveraging these investments for inevitable expansion to support far wider use cases in subsequent phases of deployment.
- Once an initial set of business requirements has been identified and defined, the leaders of successful initiatives then assess the technical requirements, identify gaps in their current capabilities, and plan the investments to close those gaps.
- Successful initiatives tend to follow three patterns of deployment underpinned by the selection of one big data "entry point" that corresponds to one of the key characteristics of big data—volume, variety, and velocity.
- After completing their initial deployments, government leaders typically expand to adjacent use cases, building out a more robust and unified set of core technical capabilities. These capabilities include the ability to analyze streaming data in real time; the use of Hadoop or Hadoop-like technologies to tap huge, distributed data sources; and the adoption of advanced data warehousing and data mining software (TechAmerica Foundation, 2012, p. 7).

The Commission made the following recommendations for government agency leaders to adopt when implementing big data solutions:

- Understand the “Art of the Possible” by reviewing case studies of prior implementations to understand practical examples.
- Identify two to four key business or mission requirements that big data can address for the government agency, and define and develop underpinning use cases that would create value for both the agency and the public.
- Take inventory of the “data assets.” Explore the data available both within the agency enterprise and across the government ecosystem within the context of the business requirements and the use cases.
- Assess current capabilities and architecture against what is required to support goals, and select the deployment entry point that best fits your big data challenge, whether it is volume, variety, or velocity.
- Explore which data assets can be made open and available to the public to help spur innovation outside the agency (TechAmerica Foundation, 2012, p. 8).

Big Data in the U.S. Navy

The U.S. Naval Air Systems Command (NAVAIR) has optimized its resources with big data. NAVAIR implemented the Decision Knowledge Programming for Logistics Analysis and Technical Evaluation (DECKPLATE) system to centralize and streamline management of aircraft fleet and aircraft carriers deployed around the world (Sverdlik, 2012). DECKPLATE is used to manage fleet resources during both military and humanitarian missions. When the Fukushima Daiichi nuclear power plant was leaking radiation, DECKPLATE was used to determine readiness of the fleet operating in the area. It also provided real-time data on the danger of radiation exposure by the Navy’s assets during this time (Sverdlik, 2012).

DECKPLATE provides the following:

- Enterprise-Wide Visibility. DECKPLATE uses about 23 years of trend analysis of aircraft readiness, checking data on areas such as aircraft maintenance, flight usage and inventory, configuration baseline management, engine total asset visibility, technical directives, and supply cost.

- Daily Reporting. Daily readiness reporting is provided with messages going out every day from an aircraft carrier deployed at sea concerning aircraft status. In 2004, these reports would be correlated on a monthly basis, put on a DVD, and sent to commanders with the readiness status.
- Constant Process Optimization. DECKPLATE provides on-going improvements of its processes. It can provide data to address problems pro-actively, before they occur, which the traditional reporting process did not allow for.
- Changing Logistics Philosophy. Historically, the military wanted 100% of its assets up 100% of the time and that required expenditures to fix things that weren't really necessary. With DECKPLATE, an initiative was created to optimize the logistics process to have the right assets with the right configuration in the right place at the right time (Sverdlik, 2012, Enterprise-Wide Visibility, para. 6).

The next phase for DECKPLATE is *binning* in which data would be evaluated on a more granular level (Sverdlik, 2012). In the binning project, a history of some 200 million maintenance actions would be broken down into the individual maintenance actions required. The historical maintenance actions would then be further broken down into every 15 minutes. This process will answer the question, "Was the aircraft awaiting maintenance during that time or was it awaiting supply?" The final objective of identifying exactly how and where time was spent on the aircraft during the maintenance period requires a massive amount of data to be collected and analyzed over a five-year period on approximately 5,000 aircraft (Sverdlik, 2012).

Ship Maintenance Vignettes

Introduction

Maintenance is crucial to the Navy's fleet readiness and ensures that the fleet reaches its expected service life. This section uses vignettes to show three aspects of ship maintenance that provide a framework for understanding these types of activities within the Navy. It begins with a general discussion on maintenance and modernization budgets, and then provides specific ship case examples.

Maintenance and Modernization Spending

Maintenance and modernization are essential to derive full benefits of DoD assets and, more importantly, they enable the U.S to respond quickly to security challenges and offer humanitarian assistance around the world. In FY2010, the DoD spent approximately $83.7 billion to maintain strategic material readiness for 13,900 aircraft, 800 strategic missiles, 350,000 ground combat and tactical vehicles, 283 ships, and myriad other DoD weapon systems (Office of the Assistant Secretary of Defense for Logistics and Material Readiness [OASD(L&MR)], 2011). Figure 21 shows the systems supported by the DoD. Maintenance was provided through the efforts of approximately 657,000 military and civilian maintainers and thousands of commercial firms.

Figure 21. Systems Supported by DoD Maintenance
(OASD[L&MR], 2011)

Performed at several levels, DoD material maintenance ranges in complexity from daily system inspections, to rapid removal and replacement of components, to complete overhauls or rebuilds of a weapon system. The three levels of maintenance are as follows: depot-level maintenance for the most complex and extensive work; intermediate-level maintenance for less complex maintenance activities performed by operating unit back-shops, base-wide activities, or consolidated regional facilities; and field-level maintenance, a combination of organizational depot and intermediate levels (OASD[L&MR], 2011).

In early 2011, the DoD operated 17 major depot activities and expended more than 98 million direct labor hours (DLHs) annually (Avdellas, Berry, Disano, Oaks, & Wingrove, 2011). DoD depots' property, plants, and equipment were valued at more than $48 billion with an infrastructure consisting of more than 5,600 buildings and structures (Avdellas et al., 2011).

To maintain readiness and ensure that the fleet reaches its expected service life, the Navy spent $8.5 billion on ship maintenance in FY2011. Figure 22 shows the Navy's maintenance budget.

(Dollars in Millions)	FY2011	FY2012	FY2013
Active Forces			
Ship Maintenance	$4,726	$4,533	$5,090
Depot Operations Support	$1,326	$1,296	$1,315
Baseline Ship Maintenance (O&M,N)	**$6,052**	**$5,829**	**$6,405**
Overseas Contingency Operations	$2,484	$1,493	$1,310
Total Ship Maintenance (O&M,N)	**$8,536**	**$7,322**	**$7,715**
Percentage of Projection Funded	100%	97%	100%
Annual Deferred Maintenance	$0	$217	$0
CVN Refueling Overhauls (SCN)	1,664	530	1,683
% of SCN Estimates Funded	100%	100%	100%

Figure 22. U.S. Navy Ship Maintenance Costs
(Department of the Navy [DoN], 2012a)

Maintenance Vignettes

Each of the three vignettes describes an aspect of ship maintenance work: new work (NW), deferred maintenance (DM), and modernizations. Although there is another category, *original work*, maintenance for which planning has been completed and is included in the maintenance package before the evolution (an *availability*, or *avail*), this section focuses on NW, DM, and modernizations.

NW is maintenance added to a specific ship's availability after planning has been completed (i.e., not part of the original maintenance package). NW can result from discrepancies that have not yet been discovered or from work which was not added to the availability work package until after planning was complete. DM refers to the status of maintenance rather than the time of its inclusion in the maintenance package and may be either original work or NW. DM is work that is rescheduled to be completed later in the current availability or as part of a future maintenance period. Modernizations (or *mods*) are system upgrades. A modernization can range in scope from a short-term software upgrade to a long-term ship infrastructure remodeling. Generally, the planning for all the modernization work is completed before the availability begins and is therefore classified as original work. However, in the modernization vignette in this section, two cases demonstrate that situations can arise that require modernization work to become NW. Figure 23 shows the relationships among the different categories.

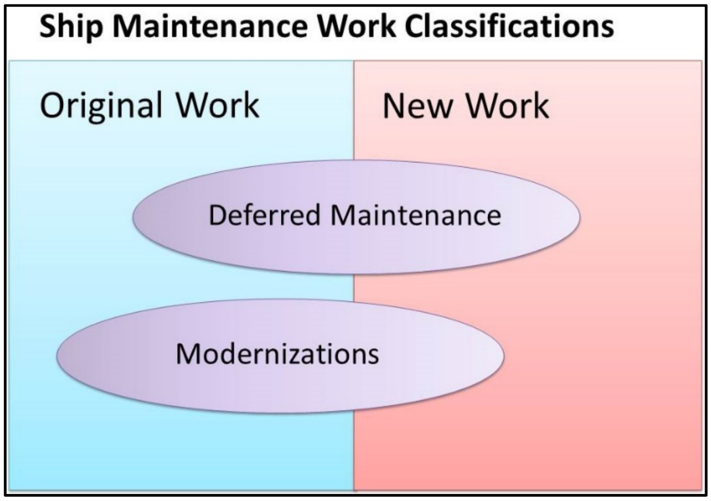


Figure 23. Ship Maintenance Work Classifications

The three ships used in the vignettes, under the cognizance of Norfolk Ship Support Activity (NSSA), are the United States Ship (USS) *Wasp* (LHD-1), the USS *Bataan* (LHD-5), and the USS *Iwo Jima* (LHD-7). First, LHD-7 will be a case study to describe NW. Second, to depict DM, both LHD-5 and LHD-7 will be examples. Finally, LHD-1 and LHD-7 are used to illustrate modernizations, as shown in Figure 24.

Ships	Vignette: New Work	Deferred Maintenance	Modernizations
USS *Wasp* (LHD-1)			X
USS *Bataan* (LHD-5)		X	
USS *Iwo Jima* (LHD-7)	X	X	X

Figure 24. Vignette Overview

The three vignettes that follow were derived from two phone conversations with David J. Furey, a civilian employee of the NSSA, on September 9, and September 11, 2013.

New Work Vignette: USS *Iwo Jima*

The USS *Iwo Jima* is an example requiring NW, DM, and modernization. In addition, this case examines NW and how complications from NW can impact schedules. In this vignette, the focus is on the rudder and the bilge. The rudder, a critical portion of the ship's steering system, caused a schedule extension due to degradation that was not readily apparent. All appropriate assessments, checks, and leakage tests were conducted by maintenance technicians, and the results indicated the rudder was in good condition. All the tests associated with the rudder were within specified parameters, and the rudder passed the preliminary inspection. Unfortunately, *bearing clearance testing*, tests which analyze rudder performance over the entire range of operation (full left to full right), exposed inconsistencies prior testing did not reveal. Results from the test were irregular and upon examination of the rudder bearings, metal debris and rust were discovered. The NSSA ultimately made the decision to remove and replace the rudder, which resulted in the availability schedule being extended by 14 days.

NW was also required on the bilge of the USS *Iwo Jima*. As part of the entire availability, high pressure washing was required in the bilge. While performing this evolution, fuel piping was damaged and a leak developed. *Ship's force*, a term which describes the active duty sailors onboard the ship, repaired the damage by using a *soft patch*. A soft patch is a temporary repair method for low pressure piping. However, the NSSA was constrained by more restrictive requirements and was required to replace the faulty piping. To determine the extent of the damage, *ultrasonic testing* (UT) was used, which uses sound wave properties to determine the amount of pipe wall thickness remaining. If less than 50% of the pipe wall remains, the NSSA is required to replace the pipe. UT was performed and revealed 40 ft. of fuel piping, and an additional 20 ft. of oily waste piping, that required replacement. The availability schedule was extended by 40 days to replace the identified piping.

Deferred Maintenance Vignette: USS *Bataan* and USS *Iwo Jima*

In these vignettes, the USS *Bataan* example relates to cost-cutting while the USS *Iwo Jima* example relates to prioritization. The overall magnitude of work to be accomplished during the USS *Bataan* availability made it a target of cost-cutting during shrinking fiscal budgets in 2012. A common item to be deferred is paintwork and the USS *Bataan* was not an exception. Much of the tank paintwork was deferred from the 2012 availability to the 2015 availability as a result of fiscal cutbacks.

The USS *Iwo Jima* also experienced DM, but the maintenance was deferred because higher priorities required the ship to be waterborne. Specifically, the 7-K-O-W tank, the forward feed tank for the ship's ballast system, was due for preservation and required the ship to remain in drydock. The tank had not been opened since commissioning as this was the ship's first drydock availability. Inspection revealed the tank to be in Tank Condition 4, which means that a profound failure had been discovered. UT showed that no more than 17% surface wastage had occurred and, therefore, the tank had become a candidate for deferral. Higher priority maintenance necessitated that the ship be waterborne, so the drydock was flooded and the 7-K-O-W tank preservation was deferred.

While the effect on a ship's availability schedule of the addition of NW can be directly measured, the consequence of deferring maintenance is a matter of risk. The USS *Iwo Jima* added NW to its availability and incurred schedule delays, or *lost operating days* (LODs); 14 days were attributed to work on the rudder and 40 days to the replacement of pipe. In both cases, the impact can be easily measured.

As for DM, the impact can range from minimal to substantial. For instance, the tank paintwork for the USS *Bataan* was deferred until the next planned availability in 2015. The paintwork would have cost a certain dollar amount in 2012 and would have provided the tank a level of preservation protection. In 2015, the paintwork will cost more not only because of inflation and the degradation of the paint associated with time, but also because corrosion will have developed at a higher rate than it would have with a fresh application of paint. The difference between the cost of paintwork in 2015 versus the cost in 2012 (including corrosion correction) is the impact of this DM example and would be comparatively minimal. However, the possibility of a larger effect exists. Perhaps the development and growth of corrosion on the 7-K-O-W tank is underestimated. If the corrosion progresses significantly faster, then the likelihood of structural failure increases. Should the structural failure occur outside the maintenance environment of the shipyard, then the impact would be far greater and the costs associated with unscheduled maintenance much higher. The decision to defer the preservation of the tank must consider both the likelihood and severity of all the possible outcomes. In other words, the decision-maker must consider all the associated risks before deferring maintenance.

Modernizations Vignettes: USS *Iwo Jima* and USS *Wasp*

Modernizations have the most potential to impact schedule of the three classifications of shipyard maintenance examined in this section. In the cases of the USS *Iwo Jima* and USS *Wasp*, modernizations may affect the timetable because not all the required drawings had been completed prior to the start of work. For the USS *Iwo Jima*, a single modernization will be presented, whereas the USS *Wasp* serves as a more general example. However, a brief overview of the shipyard planning evolution will be presented first to explain the importance of timely drawings.

Before the shipyard period starts, the plan for a scheduled availability must be completed. To complete the plan for an availability, NAVSEA must approve the contractor-provided estimate (DoN, 2012b). To generate the estimate, however, the contractor must review all the drawings (first-tier and second-tier) associated with the work to be performed (D. Furey, personal communication, September 9 and 11, 2013). First-tier drawings are the main focus of the modernization, whereas second-tier drawings involve infrastructure and subsystems related to the work. For a particular modernization, if all the drawings are not completed, then the contractor cannot create the estimate and an approved plan will not exist. In addition, availabilities must sometimes commence on a partial solution; otherwise, all work would be completed late. In the situation without an approved plan, the project completion date (PCD) has a larger margin of error, and schedule changes are more likely to occur.

This was the case with the CANES installation in the USS *Iwo Jima* availability. CANES, or Consolidated Afloat Networks and Enterprise Services, as its name implies, is a program created to consolidate many networks and services aboard ships into a single information technology system. Although not all of the drawings were received, the maintenance period started anyway. There was other work to perform; CANES was not the only reason for the USS *Iwo Jima* to visit the shipyard. As drawings for CANES were completed, they were then provided to the contractor. However, the plan for CANES could not be approved until all the drawings were received, the contractor generated the estimate, and NAVSEA accepted the plan.

In the case of the USS *Wasp*, the estimated modernization cost was extremely high at $250 million to $300 million. The high cost was partially due to modernizations needed to accommodate the F-35 Joint Strike Fighter (JSF) since the USS *Wasp* was to be the first ship to test the JSF and part of the flight deck had to be strengthened. Not only was the structural reinforcement of the after flight deck a large package, but the ship was also undergoing many other modernizations. Unfortunately, the USS *Wasp* also started its availability without a complete plan. Twenty modernization packages were not included in the plan, including the structural reinforcement of the after flight deck, because the drawings had not yet been delivered.

In addition, the NSSA erroneously included one large work item in the plan for which second-tier drawings had not yet been received. The contractor brought the discrepancy to the NSSA's attention explaining that they, the contractor, would not be able to complete an estimate before the plan was completed (also known as *100% lock*). The NSSA had two options, either extend the lock or pull the work item out and add it back in later as NW. They chose the latter.

In both these vignettes, modernizations had significant potential to severely affect the scheduled PCD because the drawings were not completed. Two questions arise associated with the implications of missing PCD on ship maintenance costs:

- Is there a cost premium to new work? In other words, do costs increase because a modernization was added after 100% lock?
- Are LODs caused by planning or scope? In other words, is it the planning process or the scope of work which is to blame for missing PCD?

Summary

The U.S. Navy ship maintenance process is already an enormously expensive endeavor. Situations which result in NW or DM only add cost to the process in the form of budget and schedule overruns. The information regarding those overruns is available to decision-makers, but only in cumbersome, static spreadsheets and in very large quantities. Executive-level ship maintenance decision-makers need a way to easily and intuitively understand the information available to them so that decisions can be made which would reduce the occurrence of NW and DM. Ship maintenance executives require a big data technology that would provide a clear understanding of the relationships among all the variables, specifically those which cause increased costs and schedule overruns. In the next section of this report, software is used to analyze the historical maintenance information of a selected group of U.S. Navy ships. It will show how big data technology could be used to provide decision-makers with a clear, intuitive visualization of ship maintenance costs.

Ship Maintenance Simulations

Overview

A team from the Naval Postgraduate School (NPS) was tasked by PEO Ships to work with naval ship maintenance metrics groups to provide additional options regarding how large datasets could be optimized. In particular, presentation methods were requested succinctly showing a ship's maintenance status, including all operational costs and schedule deviations from planned maintenance. Project sponsors also sought suggestions for improving how key information could be summarized and ultimately used in making critical maintenance allocation decisions. The current process for presenting data on the more than 150 parameters measuring ship performance maintenance costs and processes, containing billions of data points, is still done with static, cumbersome spreadsheets.

The project was conducted in three distinct phases as seen in Figure 25. First, data were collected on 19 U.S. Navy guided missile destroyers (DDG) with maintenance periods spanning a few years, 2010 to mid-2013. Data were collected on 21 maintenance availabilities for those DDGs and included definitized estimates prepared by SMEs in the planning process, along with the actual cost and availability data on three maintenance categories. In Phase 2, a hypothesis was tested and two simulations were run using the Knowledge Value Added (KVA) methodology. In Simulation 1, we tested the potential impact of incorporating three-dimensional printing (3DP) on ship maintenance programs while in Simulation 2 we evaluated the combination of 3DP plus two more technologies (3D laser scanning technology [3D LST] and Collaborative Product Lifecycle Management [CPLM]). In Phase 3, a visualization tool offered by an independent software vendor was selected to show how large volumes of data could be shown in a succinct manner.

Phase 1 : Data Collection

- Definitized estimates for 19 guided missile destroyers (DDG)
- Twenty-one maintenance availabilities from 2010 to mid-2013
- Actual costs from Surface Team One Metrics System (ST1MS)
- Cost categories of Growth, New Growth, New Work, Original

Phase 2: Simulations

- **Simulation 1** Three-dimensional printing technology (3DP)
- **Simulation 2** Three-dimensional printing technology (3DP)
  Three-dimensional laser scanning technology (3D LST)
  Collaborative Product Lifecycle Management (CPLM)

Phase 3: Analysis & Results

- Definitized cost estimates for maintenance work ($313.7 million)
- Actual costs for maintenance work ($435.5 million)
- Cost estimates after simulations incorporating technologies ($271.1 million)
- Potential cost savings of 37.7% ($164.4 million)

Figure 25. Project Phases

The visualization software provides a higher level of visual clarity, enabling faster and more intuitive interpretation of ship maintenance data by presenting the data relationships in diagrams, graphs, and charts. Relationships among variables are more readily discoverable and, more importantly, those relationships can be used in forecasting to develop more accurate maintenance data, estimates that are based on historical data. Decision-makers are able to see analytical results quickly with visualization software, which allows them to find relevance among millions of variables, communicate concepts and hypotheses to others, and even forecast possible scenarios.

This section of the report is divided into several topics. First, maintenance categories and the data collection process are reviewed. Final simulation results are highlighted to provide a framework for understanding the power of visualization software, followed by a general discussion of the original definitized cost estimate (Figures 27–29). Actual costs are then compared with the definitized cost estimates and discrepancies between the two are discussed

(Figures 30–34). An analysis of the potential effect on ship maintenance costs by incorporating specific technologies in Simulation 1 and Simulation 2 is discussed in greater detail (Figures 35–38). Alternative presentation methods, which drill down into specific detail, are then explored (Figures 39–41). Next, a description and analysis of a common ship maintenance metric, lost operating days (LODs), is given along with a recommendation of a more useful metric, availability density (Figures 42–44). This section concludes with further examples of visualization tools' abilities to drill down into specific details. (Figures 45 and 46).

Maintenance Cost Categories

There are several cost categories for ship maintenance: original work (OW), growth (G), new work (NW), and new growth (NG). OW is the estimated ship maintenance cost (shipyard or contractor, labor, and material costs) at the completion of planning and is also known as the definitized cost estimate. The definitized cost estimate is a figure provided by a SME in the planning process.

G is an expansion of OW and can result from many factors, including undiscovered discrepancies or an increase in scope. For example, the OW plan for a hypothetical ship called for preservation work on the ship's hull. While conducting the preservation work, the maintenance technician discovered hull damage that required minor repair. The minor repair work would be classified as G.

NW is maintenance that is added to a ship's availability after planning has been completed (i.e., not part of the original work maintenance package). NW can result from discrepancies that have not yet been discovered and are unrelated to previously planned maintenance or from work that was not added to the availability work package until after planning was complete. For example, while conducting preservation work on the hypothetical ship, the maintenance technician discovered damage to a communication antenna. The resulting repair work would be classified as NW.

NG is the growth resulting from an expansion in NW, similar to the relationship between G and OW. For example, the antenna maintenance technician conducting antenna repair work discovered that the antenna was beyond repair and needed to be replaced. Replacement of the antenna would be considered NG.

Data Collection

Data for this analysis were derived from the ST1MS)website (https://mfom-shipmain.nmci.navy.mil). In particular, ship availabilities were selected for examination based on several factors designed to establish a proof of concept for the use of big data to shape executive-level decisions. The availabilities were restricted to only U.S. Navy DDGs whose maintenance period started by 2010 and whose final reports were closed and completed by the time this study began in 2013. Ships whose close-out reports were incomplete or missing data were not included in the analysis.

The figures in this section are screenshots of solar graph results that were captured while using the visualization software program to process the ship maintenance data obtained from the ST1MS website. The data consist of 21 maintenance availabilities for the DDGs.

Final Simulation Results Incorporating Different Combinations of Technologies into U.S. Navy Ship Maintenance Programs

Two simulations were run to show the potential cost savings of incorporating specific technologies. In Simulation 1, only 3DP technology was evaluated while in Simulation 2, three combined technologies were evaluated. Tables 5 and 6 reflect the differences between definitized costs, actual costs, and projected costs for Simulations 1 and 2. The definitized cost estimate was $313.7 million, compared to the actual cost of $435.5 million. If 3DP, 3D LST, and CPLM technologies combined were incorporated into the ship maintenance processes, the costs would have been reduced to an estimated $271.1 million.

Table 5. Cost Comparison by Ship

(Based on J. Kornitsky, personal communication, November, 2013)

Cost Comparison by Ship (dollars represented in millions)										
Ship	Definitized Cost Estimates	Actual Costs	% vs. Definitized	Simulation 1 (3DP)	% vs. Definitized	% vs. Actual	Simulation 2 Radical (3DP, 3D LST, CPLM)	% vs. Definitized	% vs. Actual	% vs. 3DP
Barry	$48.0	$70.1	46.0%	$65.8	37.1%	-6.1%	$43.9	-8.5%	-37.4%	-33.3%
Arleigh Burke	$46.9	$58.0	23.7%	$56.4	20.3%	-2.8%	$35.7	-23.9%	-38.4%	-36.7%
Ramage	$46.3	$57.2	23.5%	$55.9	20.7%	-2.3%	$35.7	-22.9%	-37.6%	-36.1%
Donald Cook	$21.4	$36.3	69.6%	$36.2	69.2%	-0.3%	$22.9	7.0%	-36.9%	-36.7%
Stout	$45.4	$63.2	39.2%	$64.1	41.2%	1.4%	$38.6	-15.0%	-38.9%	-39.8%
All Other	$105.6	$150.5	42.5%	$147.5	39.7%	-2.0%	$94.1	-10.9%	-37.5%	-36.2%
TOTAL	$313.6	$435.3	38.8%	$425.9	35.8%	-2.2%	$270.9	-13.6%	-37.8%	-36.4%

Table 6. Cost Comparison by Work
(Based on J. Kornitsky, personal communication, November, 2013)

Cost Comparison by Work (dollars represented in millions)							
Work	Definitized Cost Estimates	Actual Costs	% vs. Definitized	Simulation 1 (3DP)	% vs. Actual	Simulation 2 Radical (3DP, 3D LST, CPLM)	% vs. Actual
Original	$313.7	$313.7	0.0%	$307.3	-2.0%	$195.4	-37.7%
Growth	$0.0	$47.1	100.0%	$45.7	-3.0%	$28.1	-40.2%
New Work	$0.0	$66.8	100.0%	$65.5	-1.9%	$43.0	-35.6%
New Growth	$0.0	$7.7	100.0%	$7.4	-3.9%	$4.5	-41.6%
TOTAL	$313.7	$435.5	38.7%	$426.2	-2.1%	$271.1	-37.7%

Visualization Software Analysis of U.S. Navy Ship Maintenance

Visualization Model

The visualization model (Figure 26) is an overview of how the DDG spreadsheet data was mapped into the software. It shows four cost categories on top, all 19 ships by name in the middle, and their combined availabilities at the bottom. The lines between the boxes depict connection relationships.

The 24 boxes referred to in the model have a number above each that represents the aggregate cost. For example, the box on the middle left side of Figure 26, labeled *Stout*, indicated $28.1 million of aggregate cost attributed to the availability. In addition, the horizontal bar between the cost number and the box represented the relative portion of cost attributed to that availability when compared with all availabilities. The box in the top left corner, labeled Growth, indicated a relative cost which resulted in the length of the bar shown.

At the bottom of each box, the number of connections to all other variables was depicted in two ways. The number displayed in the bottom right of each box and the number of ovals displayed in the bottom left of each box. At the box at the bottom of Figure 26, labeled Avail, indicated 21 connections to all other variables with both the numeral, “21,” and the number of ovals displayed, 21.

At the top of Figure 26, four boxes are depicted and represent one category of cost, type of work. The labels on each box indicate a particular type of work, G, NG, NW, and OW. Each particular type of work accounted for the amount of cost indicated.

In the middle of Figure 26, the 19 boxes labeled with ship names indicate the maintenance cost each ship incurred. For 17 of the ships, the ship maintenance cost was attributed to a single availability. For the *Arleigh Burke* and the *Donald Cook*, the ship maintenance cost was attributed to two availabilities. For example, the box labeled *Arleigh Burke* in Figure 26 indicated $35.7 million in ship maintenance cost, but for two unique availability periods. This can be verified by referencing the number in the lower-right portion of the ship name boxes. For most of the ships, this number was 4 and the number of ovals was four. This represented the number of connections to the kinds of cost. In any single availability, there were four types of work (cost) identified (OW, G, NW, and NG). In the cases of the *Arleigh Burke* and *Donald Cook* ships, there were two availabilities recorded, and, therefore, eight connections to the four types of work (cost) as was indicated by the number, 8, and the eight ovals indicated in either box in Figure 26.

The single box depicted at the bottom of Figure 26 represented the aggregate forecasted cost of all availabilities, $271.1 million.

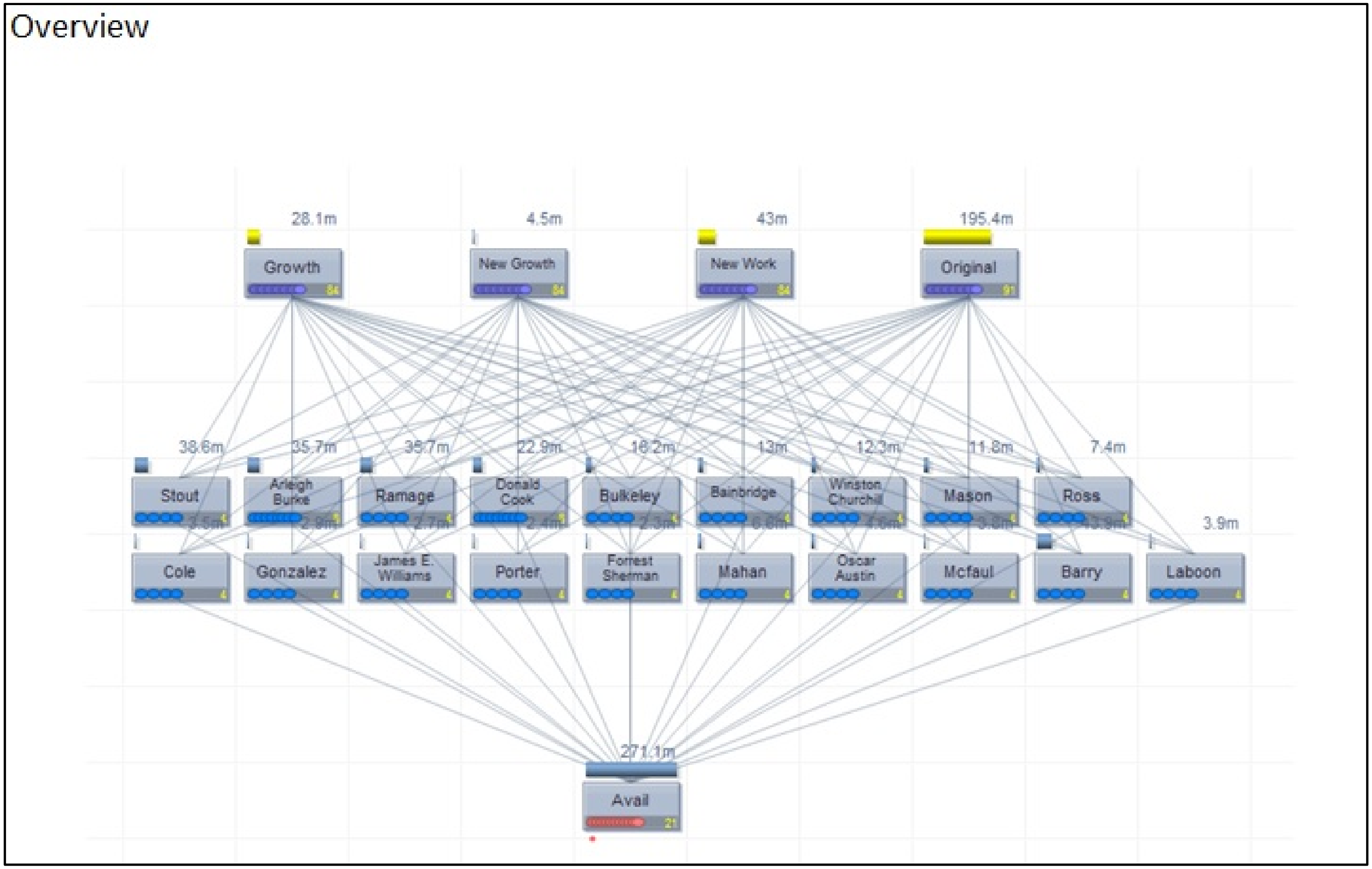


Figure 26. Visualization Model
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimate, All Ships

Definitized estimates are the total projected costs of an availability upon completion of the planning phase of ship maintenance, provided by SMEs in the planning process. According to the *Joint Fleet Maintenance Manual* (DoN, 2012b), the planning phase for an availability for a DDG begins 720 days before the first day of maintenance (A-720). By this day, A-720, an availability must be added to the U.S. Navy surface ship availability schedule. The next milestone, a letter of authorization, occurs on or before A-360 and obligates the stakeholders to specific cost of prorate schedules. Through the next three milestones, A-240 (50%), A-120 (80%), and A-75 (100%), progressively more of the budgeted funds must be allocated, or locked, to specific work items. By A-60 the overall plan for maintenance must be finalized to allow the detailed work schedule to be formulated and cost estimates completed. The final cost estimate, or definitized work package, must be finished by A-35 and represents all costs attributed to OW. After definitization, all additional work items are considered to be G, NW, or NG (DoN, 2012b).

Figure 27, the Definitized Estimate, All Ships, shows how each ship contributed to the total expected cost of all the availabilities analyzed. The total of $313.7 million is greater than the total presented in the previous image, $271.1 million. As explained earlier, this is because the first screenshot shows the total costs after the combined incorporation of three different technologies into the ship maintenance process.

All the figures shown in this section present a parent-child type of relationship hierarchy, similar to object-oriented programming. In Figure 27, there exists only a simple relationship with each instance having assumed a single role. The total definitized estimate of $313.7 million in the center is the parent, while all the ships, and their total maintenance costs, are the children. Multi-role instances, where the single solar graph screenshots can be both parent and child, will be presented in later figures, beginning with Figure 29.

Each ship contributed to the total definitized estimate of $313.7 million. The amount each contributed is presented in three different ways. First, the size of each bubble signifies its cost relative to the total cost bubble in the middle of the screenshot. The larger the relative cost of the ship identified, the larger the bubble. Second, the relative impact of each ship on cost is also identified by a percentage written on the line connecting each ship with the total. Finally, the actual dollar amount of each ship's impact upon the definitized estimate is shown either inside the instance for larger contributors or near the instance for smaller ones.

The *Winston Churchill*, for example, which is located at the eight-o'clock position on Figure 27, was not the largest contributor to the total definitized estimate. However, a brief visual analysis of the entire figure shows it was not the least significant either because many of the ship solar graph screenshots are smaller. The relative sizes and organization of all the instances enable an intuitive understanding to be quickly developed. The *Winston Churchill* screenshot is larger than the four instances directly below it, but it is also smaller than the four instances directly above it. The relative location of the *Winston Churchill* instance enables a decision-maker to quickly identify that the ship's relative contribution to the overall definitized estimate lies somewhere in the middle of the pack.

If further understanding of the relative contribution is needed, the decision-maker would then refer to the percentage indicated along the line connecting the *Winston Churchill* to the total estimate. The *Winston Churchill* accounted for 3.7% of the total definitized estimate. However, if the actual dollar amount contributed to the total is desired, then the decision-maker could refer to the number located within the instance. In the case considered, the *Winston Churchill* accounted for $11.7 million in absolute terms.

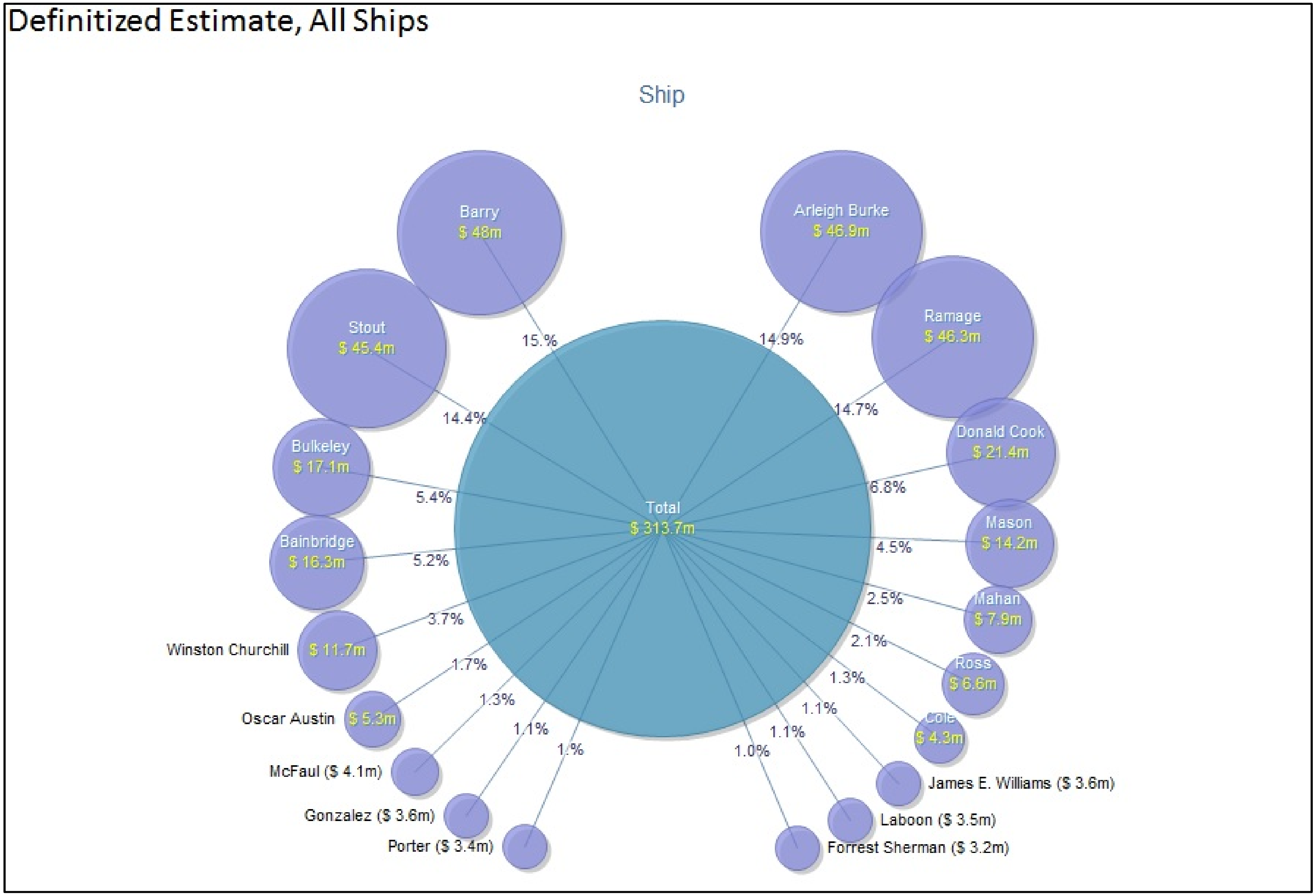


Figure 27. Definitized Estimate, All Ships Screenshot
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimates of the Top Five Ships

Figure 28, Definitized Estimate, Top Five Ships, is nearly identical to Figure 27 except that it has been modified to identify the largest cost contributors. The five largest contributors are shown, and the remaining 14 ships are aggregated into “all other.”

Consider the decision-maker analyzing the presentation. If the executive is only interested in the largest cost contributors, then the addition of the other 14 ships only makes interpretation of the information more difficult. However, the aggregation of the remaining ships into a single instance also provides another view of the data. In this example, the total definitized estimate of the other ships is $105.6 million and represents 33.6% of the entire sum. This view may be significant to a decision-maker who originally thought that the largest cost contributors represented a much larger portion of the total. In this figure, a decision-maker would easily be able to determine that the impact of the remaining 14 ships is much greater than the impact of any single large cost contributor.

Alternatively, if the decision-maker was more interested in determining the sources of the expenses, then an additional level of detail would be necessary. While Figure 28 provided cost information, the costs were aggregated at the ship level. An executive interested in determining the primary drivers of cost would need more detailed information that can be found in Figure 29.

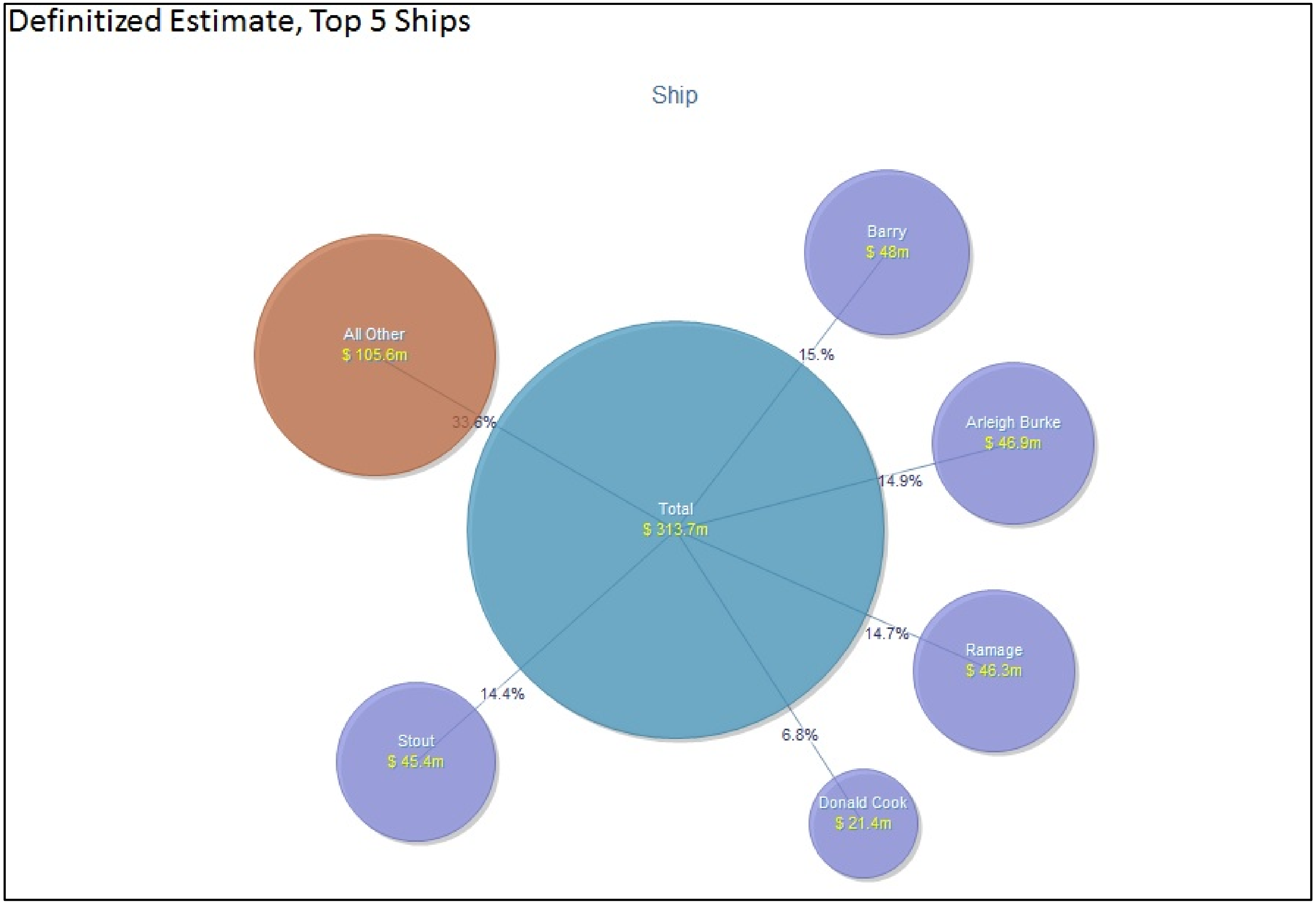


Figure 28. Definitized Estimate, Top 5 Ships Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimates of Top Five Ships by Expense Details

Figure 29, Definitized Estimate, Top 5 Ships, Expense Detail, adds one level of detail to the figure previously discussed. These added dimensions are two cost categories of labor and material, which can be seen radiating further from the graph's center and labeled with the availability's identification number from which it originated. These additional details to the definitized estimate of the top five ships increased the complexity of the parent-child hierarchy and produced different numbers of children among the ship-level instances. For the executive using this solar graph to make important ship maintenance decisions, it is important to understand the changes.

First, the parent-child relationship hierarchy has increased in complexity. With the addition of another level of detail, or another layer of children, the ship name solar screenshots have become both parent and child. The ship names are still children to the parent, total definitized estimate, but are now also parents to the expense details. For example, located at the one-o'clock position in Figure 29, the *Barry* solar graph instance has spawned two children, Labor and Material. The *Barry*, originally only a child to the total definitized estimate, is now also a parent to its two children. However, this concept has produced ship name parents with varying numbers of children and their causes may not be initially intuitive.

Earlier, both the *Arleigh Burke* and the *Donald Cook* ships were identified as being irregular because they represented multiple availabilities. The addition of expense detail has further demonstrated the presence of two separate maintenance periods within each. Just above the three-o'clock position in the solar graph, the *Arleigh Burke* shows four children. Two are labeled as Labor and two are labeled as Material. However, each one labeled Labor is identified by a unique availability identification number, and each one labeled Material has the same unique numbers. The *Arleigh Burke* and *Donald Cook* multiple availability instances produced four children as opposed to the two children generated by the single availability instances of the *Barry*, *Ramage*, and *Stout* ships.

To an executive, the additional level of detail in the solar graph begins to remove ambiguity and provide clear relationships among the sources of cost. But, if the manner and method in which the detail is presented is confusing, then the additional information will only further confound the decision-maker. Understanding why ships produced varying numbers of Labor and Material children is important for the executive to make appropriate decisions regarding ship maintenance based on the solar graph. However, the six children subordinate to the All Other instance at the ten-o'clock position in Figure 29 also require explanation.

The reason the All Other instance produced six children is two-fold. First, the All Other instance includes 14 ships and, therefore, 14 availabilities (since the *Arleigh*

Burke and *Donald Cook* have already been accounted for). The definitized cost estimate for each availability has been categorized into Labor and Material. So, 14 availabilities should have generated 28 expense detail children. There are more than just two or four children available to display. This leads to the second part of the two-fold explanation. In Figure 29, the number of children to be displayed was arbitrarily chosen. The top five largest contributors retained their individual solar screenshots, and the remaining were aggregated into the All Other instance. The choice to display the top five ships in the screenshot with less detail has also affected this graph. The biggest five individual contributors, all which happen to be Labor instances, are displayed while the remaining are aggregated into the All Other instance. Again, the implication for the executive using this solar graph to form ship maintenance policy decisions is that if the manner and method of solar graph creation aren't known, then the insight derived from the graph will be erroneous. For example, if decision-makers assumed that the All Other category displayed all its children, then they would misunderstand the graph and believe that only labor costs were incurred for those 14 ships.

From Figure 28, previously seen, the decision-maker was interested in finding more about the cost sources. Now in Figure 29, with an added level of detail, the decision-maker could make more observations and gain a deeper understanding of cost drivers. For instance, the top five ships all demonstrated that for a given availability, labor impacted cost more than material. Specifically, consider the *Barry*, *Ramage*, and *Stout*. The labor costs accounted for percentages ranging from 70.8% to 73.4%. In this small sample, the decision-maker could develop cost baselines indicating that for a given availability, labor accounted for about 70% of the cost and material accounted for about 30%. Given that the small sample size is an accurate screenshot of DDG ship maintenance, then the definitized cost estimates of future availabilities could be compared to the baseline and predictions generated about how the cost profile might change before work is completed.

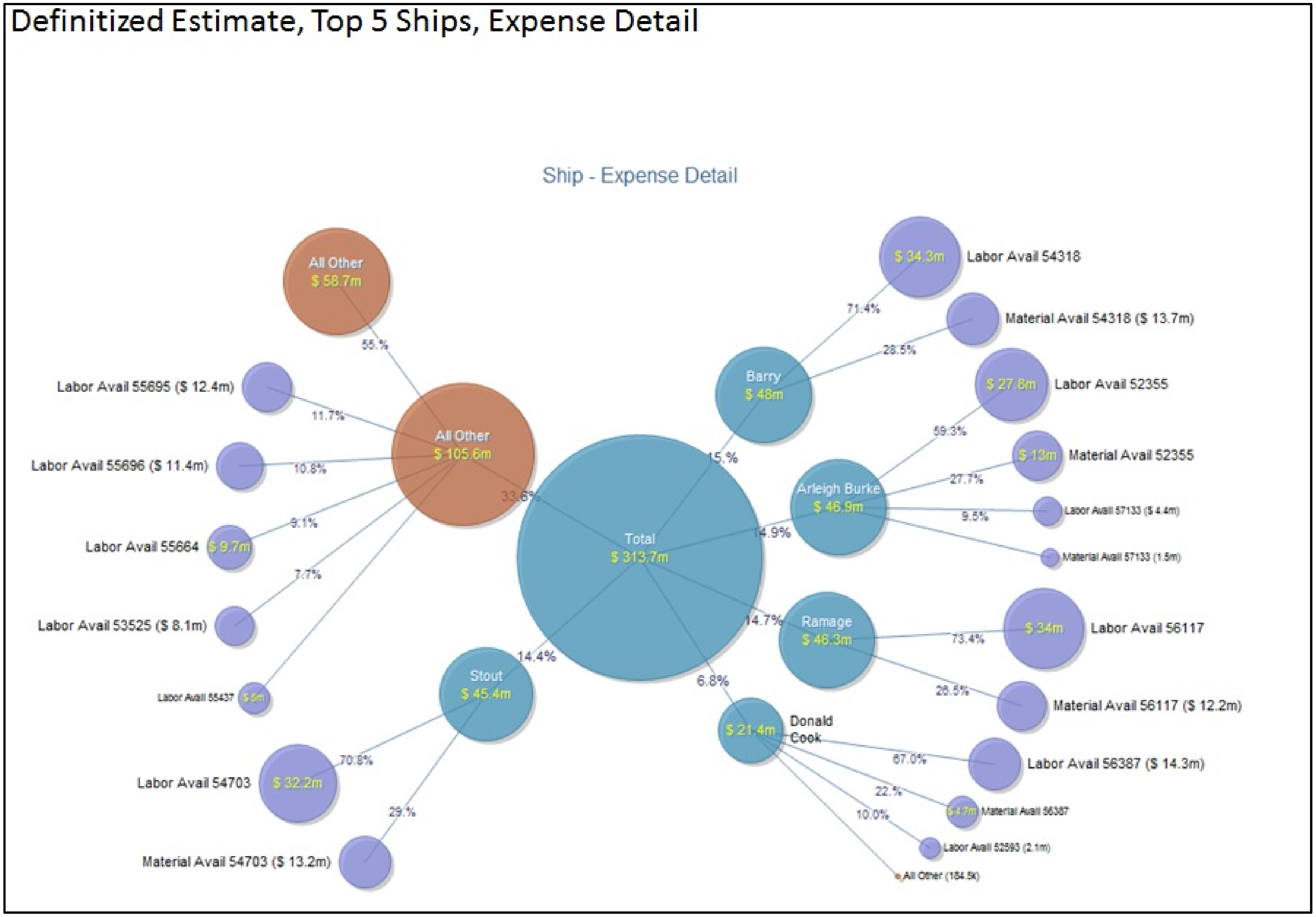


Figure 29. Definitized Estimate, Top 5 Ships, Expense Detail Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Actual Costs of the Top Five Ships by Type Expense

While the definitized cost estimate solar graphs do produce valuable information, they represent only well-educated guesses of the actual cost. The next screenshot (Figure 30) provides actual cost and is organized by the top five ships with an additional level of detail. Figure 30 also provides the additional information of Type of Expense.

Most noticeably, the total actual cost, represented by the largest solar instance in the center of the graph, has increased to $435.5 million. However, referring back to the previous figure (Figure 29), definitized cost was estimated to be $313.7 million so the costs actually increased by 38.8%. A visualization tool enables the decision-maker to drill down further to identify the largest cost drivers.

The types of expenses figure provides the ability to drill down further into the cost sources. Whereas expense detail was broken down into only labor and material categories, type expense splits those into (shipyard) labor, sub (contractor) labor, (shipyard) material, and sub (contractor) material. From here forward, the additional description in parentheses will be excluded, but the terms will retain their definitions. Labor and material, in the context of type expense, refer to the labor and material costs associated with the shipyard hosting the availability. Sub labor and sub material refer to the same costs, but those associated with the expense incurred by subcontractors.

In the *Arleigh Burke*, at the four-o'clock position on the figure, the definitized estimate for this ship was $46.9 million, and the actual cost was $58 million. That represents an increase of 23.7%. However, a decision-maker, knowing that labor is a larger contributor to cost than material, wants to know what type of labor expense is more responsible, the shipyard or the subcontractors. In the case of the *Arleigh Burke*, sub labor accounted for 50.2%, whereas labor represented only 19.2% of total availability cost. Representing a majority of cost for the *Arleigh Burke*, perhaps sub labor should be examined for cost-reduction opportunities.

The bubble charts of either definitized estimates or actual costs provide decision-makers with valuable insight. However, the size difference between estimates and actual costs would provide an understanding of the sources of cost growth. For instance, an executive is interested in determining the primary driver of increased costs. While the previous solar graphs possess the necessary information, further calculations are needed to figure changes in cost. If the relative and actual changes in cost were displayed on the same graph, then the decision-maker would be able to easily identify the primary drivers of cost growth and cost savings. The next four figures (Figures 31–34) demonstrate the concept of representing both the definitized estimates and actual costs simultaneously.

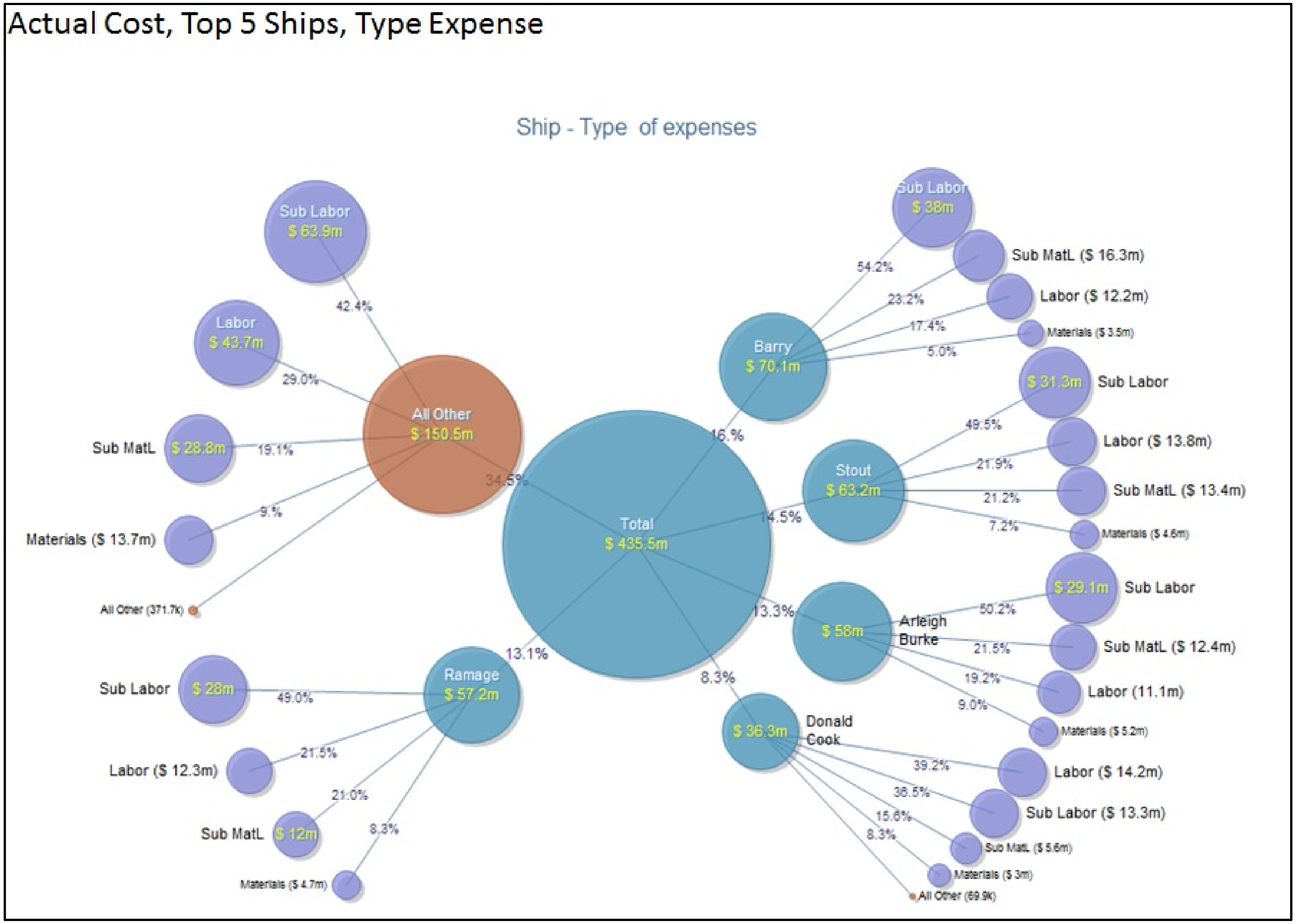


Figure 30. Actual Cost, Top 5 Ships, Type Expense Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimate versus Actual of the Top Five Ships by Type Expense

Figure 31, Definitized Estimate versus Actual, Top 5 Ships, Type Expense, displays a zoomed-in look at the comparison format to provide an introduction to the new characteristics and to review some old ones. Starting at the nine-o'clock position on the bubble chart (Figure 31), the first characteristic examined is the shell. The shell thickness and color represent the difference in amount of change and whether the change was cost growth (red) or cost savings (green).

Proceeding clockwise, the terms are familiar, but their presentation is new. Definitized cost estimate and actual cost refer to the estimated cost at the end of planning and the cost incurred upon completion of the availability, respectively. In this figure, the definitized estimate is represented by the inner layer of the shell and the actual cost by the outer layer. For example, the largest bubble represents total cost. The inner layer shows how large the instance would be if only the total definitized estimate, $313.7 million, was displayed. The outer layer shows how large the instance would be if only the total actual cost, $435.5 million, was displayed. The difference between the layers, or the thickness of the shell, represents the change in cost and is numerically indicated by the percentage shown, 38.7%. The definitized estimate was less than the actual cost, which means that there was cost growth, and is represented by the color red.

Although the next two aspects of the bubble chart are familiar, they require further clarification. First, the number represented in millions of dollars is the final state of the instance. For this comparison between definitized estimate and actual cost of the *Barry*, located at the one-o'clock position in Figure 31, the final state is the actual cost, which was $70.1 million. Second, the percentage immediately below the actual cost value indicates the change from the initial state (definitized estimate) to the final state (actual cost). In the case of the *Barry*, the cost grew by 46% from the definitized estimate to the actual cost.

The final characteristic identified on the close-up is another percentage. Whereas the percentage within the instance represented cost growth, the percentage on the line between parent and child represented the proportion of the parent's cost that the child contributed. In the figure, the dialog box arrow points at the percentage the child instance accounted for with regard to its parent, the *Barry*, or 17.4% of the total actual cost incurred by the *Barry*.

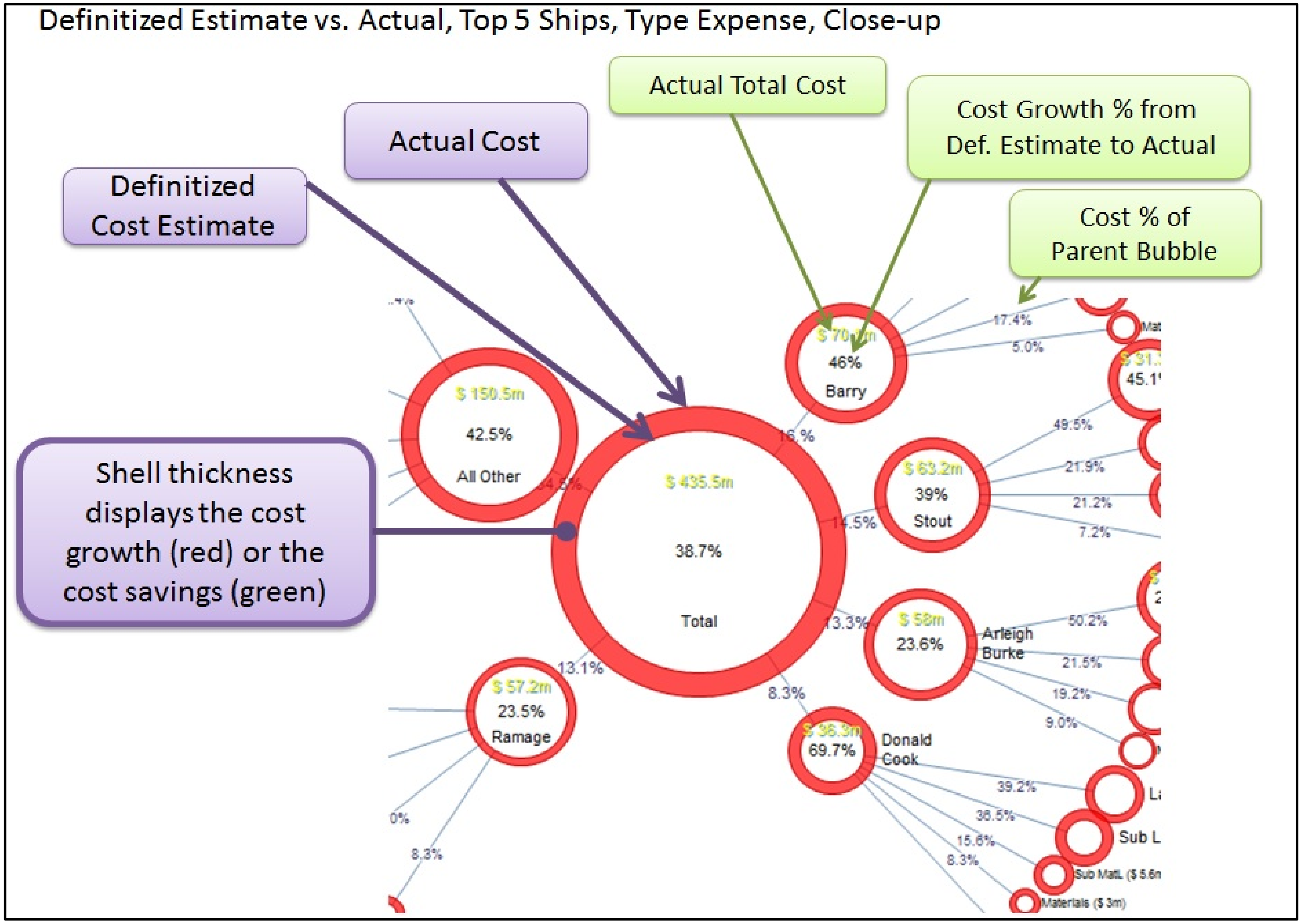


Figure 31. Definitized Estimate Versus Actual, Top 5 Ships, Type Expense, Solar Graph Close-Up
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimate Versus Actual of the Top Five Ships by Type Expense

The most distinguishing feature of Figure 32, Definitized Estimate Versus Actual, Top 5 Ships, Type Expense, which has been organized to display the top five ships to show an additional level of detail according to type expense, is the near absence of green. The two instances of cost savings, both titled All Other and located at about the nine- and five-o'clock positions on the outermost ring, are relatively insignificant, representing only 0.1% of the total cost, $435.5 million. Examples of cost growth are abundant, but an examination of the largest contributor to total cost may produce valuable insight for the executive-level decision-maker.

The All Other instance at the ten-o'clock position represents 14 ships. Those 14 ships accounted for $150.5 million, or 34.5%, of the total actual cost. The red shell and the percentage inside the All Other instance together indicate 42.5% aggregate cost growth for the 14 ships. These numbers reveal that the All Other category would be an area for a decision-maker to examine more closely in an attempt to identify the drivers of cost growth. A cursory glance at the children of the All Other instance shows that subcontractors, both sub labor and sub material, experienced more than 50% cost growth. Therefore, subcontractors are a primary driver of cost growth for at least the 14 ships represented by the All Other instance.

The visualization software provides the ability to delve into the data to discover more detail. For example, if personnel are preparing a presentation based on Definitized Estimate versus Actual, Top 5 Ships, Type Expense data (Figure 32), and the decision-maker asks the question, "What was the definitized estimate for the Barry?," then the answer can be found readily. Rather than regress to previous screenshots, the presenter can simply select the *Barry* instance and pull up a bar chart which, among other information, displays the definitized estimate. Perhaps the decision-maker requests even finer details. The software possesses the ability to drill down five levels of detail and can reproduce the data located on the original spreadsheet. So, more detail is available than just what is displayed on the static screenshots presented here. Refer to the two figures titled Barry Drill Down (Figures 44 and 45) near the end of this section for examples.

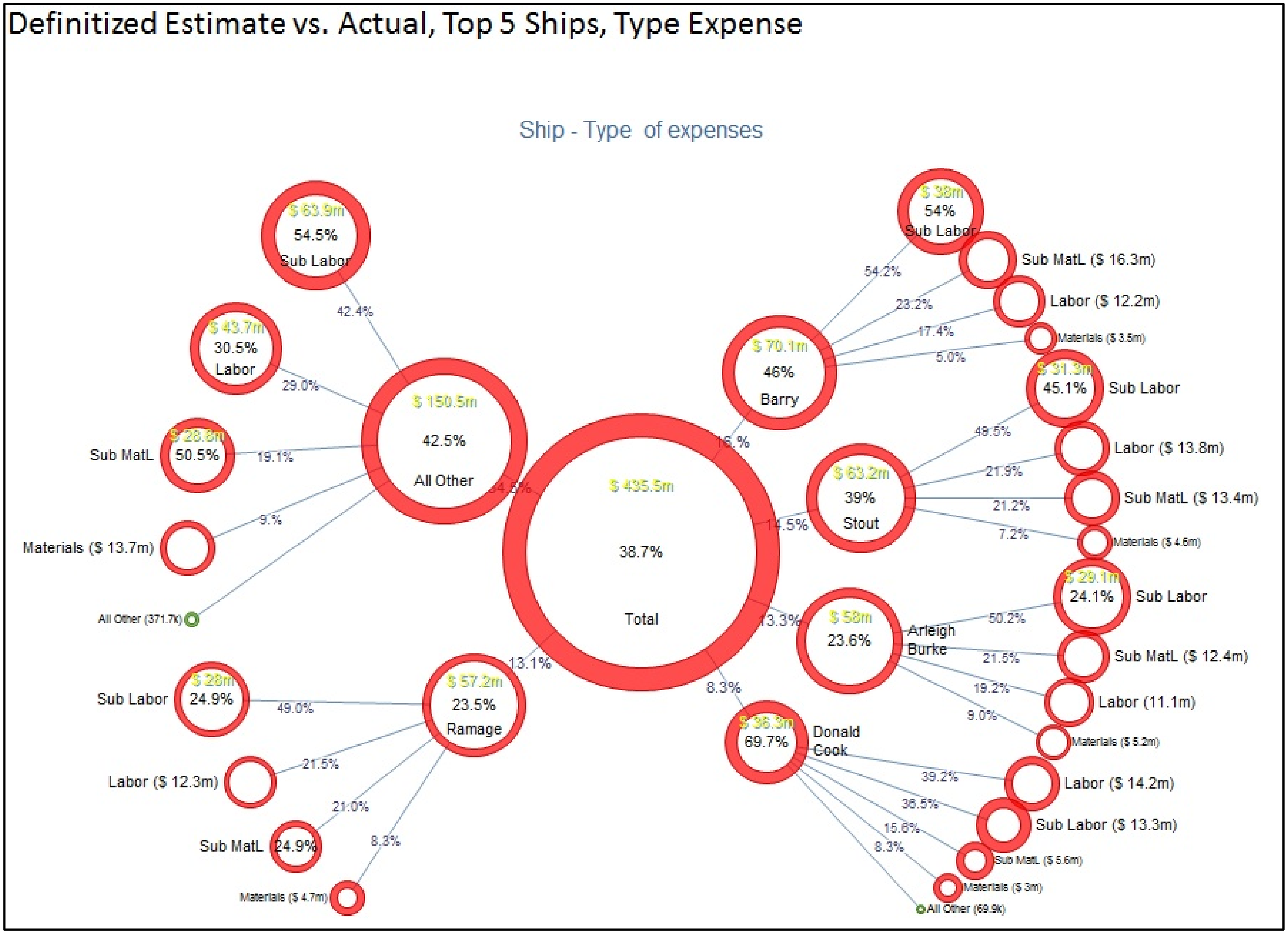


Figure 32. Definitized Estimate Versus Actual, Top 5 Ships, Type Expense Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimate Versus Actual of the Top Five Ships by Work

Figure 33, Definitized Estimate Versus Actual, Top 5 Ships, Work, is the last of three figures showing simultaneous display of both definitized estimates and actual costs, and provides the additional detail of work instead of type of expense. As previously discussed, maintenance work is broken into four types: OW, G, NW, and NG. Changing the detail to allocate cost by work creates a couple of peculiarities, both related to the definitions of the work, and important for the executive-level decision-maker to understand.

There are two anomalies when the data are changed to show work details. The first peculiarity is that there are now a significant number of instances that possibly indicate cost savings. Unfortunately, all the percentages within the green shelled instances are left blank revealing that no change (0%) has taken place. That is because the instances are representing OW, which does not change after the completion of planning, making the percentage within the instance irrelevant. For example, refer to the *Arleigh Burke* solar graph instance at the four-o'clock position in Figure 33. The green shelled child instance attached to the *Arleigh Burke* is labeled Original for OW. The percentage displayed is blank which indicates 0% change in cost has occurred because any change in cost is recorded by the other categories of work. The percentage which is important for the decision-maker to acknowledge, though, is indicated along the line connecting the child to parent. The, 80.8% for OW, indicated what portion of the total actual cost, for the *Arleigh Burke*, that OW accounted for.

The second peculiarity, also a result of definitions, is that the instances for the other three categories of work are all solid red. Solid red indicates that the baseline, or definitized estimate in this case, was $0 and the actual cost is all cost growth. That is because the other three categories of work (G, NW, NG) all result from work needed in addition to the OW and are, therefore, cost growth by definition. Continuing with the examination of the *Arleigh Burke*, its larger solid red child is labeled Growth. The percentage within the instance is blank, but again, it is less important. The significant values important to the decision-maker are the actual cost of G, $8.4 million, and the proportion of the *Arleigh Burke*'s total actual cost that G work accounted for, 14.5%. With the two peculiarities defined and understood, reconsider the previous figure to identify a cost driver.

The decision-maker examined the largest All Other instance more closely and determined that subcontractors were a primary driver of cost growth. The decision-maker might then ask to see the additional detail organized by work to further expand his or her understanding. Again looking at the All Other instance located at the ten-o'clock position in Figure 33, the largest driver of cost growth is NW, which accounted for $34.4 million, or 22%, of the actual costs for All Other 14 ships. Combine the knowledge derived from examining both graphs (Definitized Estimate versus Actual, Top 5 Ships, Type Expense and Work, Figures 32 and 33, respectively) and the keen decision-maker might direct staff personnel to investigate NW performed by subcontractors for cost-savings opportunities.

Figure 33 demonstrates how costs aggregate from the bottom up. Costs are created at the operational level and occur in different forms. Here, the forms are categorized according to the classification of work that created the cost. As the costs move from the outer rings of the solar graph, the costs are aggregated into ship instances that provide less cost detail, but is still useful as another way of looking at cost. Finally, all the ships' actual costs are aggregated into the center solar instance, Total. The visualization software offers the opportunity to view the cost data at many levels of detail, each of which delivers valuable information for decision-makers.

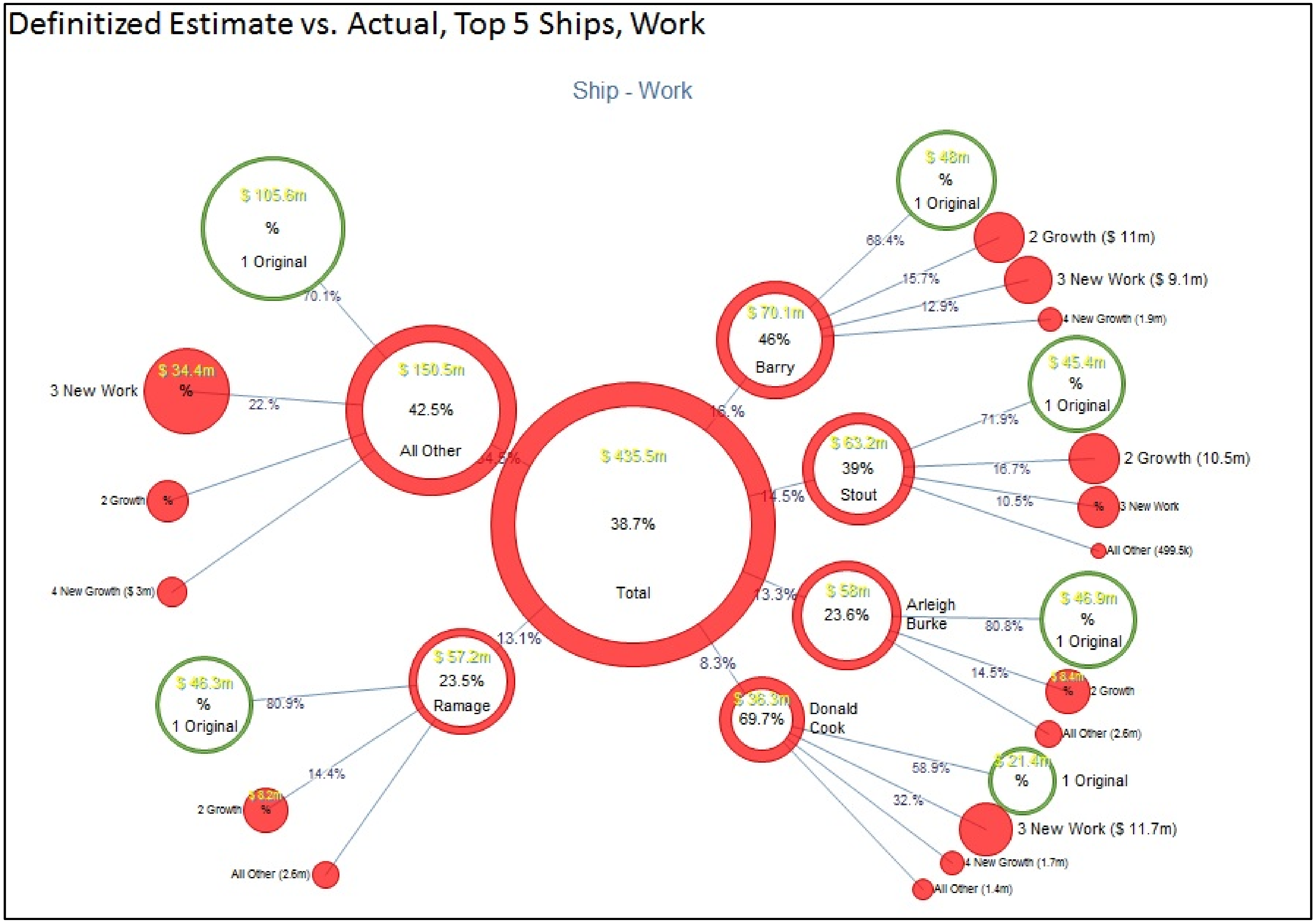


Figure 33. Definitized Estimate Versus Actual, Top Five Ships, Work Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Simulation 1 and 2: Introduction of 3DP and AM Radical

Visualization tools provide decision-makers with insights into historical data, and more importantly, offer forecasting capabilities. Before implementing process changes, which involve risk and uncertainty, an executive could use bubble charts to forecast the effects of such changes. Consider the following example.

The executive-level decision-maker has analyzed the figures previously presented and has concluded that changes to the ship maintenance process are necessary to control cost growth. Three technologies have been identified to reduce costs: 3D Printing (3DP), 3D Laser Scanning Technology (3D LST), and Collaborative Product Lifecycle Management (CPLM). To test this hypothesis, two simulations were conducted with differing implementation strategies. In Simulation 1, 3DP technology only was applied, while in Simulation 2, all three technologies (3DP, 3D LST, and CPLM combined) were applied to the ship maintenance process. Simulation results, which could identify potential cost savings, are discussed further in this section.

To quantify the potential benefits of those technologies, the Knowledge Value Added (KVA) methodology was applied. KVA assigns a value to the knowledge assets of an organization (Housel & Bell, 2001) and was used to forecast the effect that 3DP, 3D LST, and CPLM technologies would have on U.S. Navy ship maintenance programs. In one prior study, the researchers found that 3DP and CPLM could result in cost savings of as much as 81% (Kenney, 2013). Another study determined that cost savings of as much as 84% could result from the use of 3D LST and CPLM in U.S. Navy ship maintenance programs (Komoroski, 2005). The potential impact of these three technologies has been determined to be substantial. Therefore, they were used to demonstrate the ability of the software program to create intuitive screenshots of the cost savings generated by their implementation.

In the previous set of comparison figures, the definitized estimate was the baseline, and actual cost was the value compared. In the next set of four comparison figures (Figures 34–37), the baseline and the value compared are changed to examine the effect of three different technologies on ship maintenance actual cost. In the first two figures to follow (Figure 34 and 35), the actual cost is the baseline and the forecasted effect of 3DP only is the compared value. The next two figures (Figures 36 and 37) visualize the effect that the combination of 3DP, 3D LST, and CPLM , labeled as Additive Manufacturing (AM) Radical, has on ship maintenance costs.

Actual Versus 3DP for the Top Five Ships by Type Expense

Figure 34, Actual Versus 3DP, Top 5 Ships, Type Expense, visualizes the effect on actual cost of implementing 3DP into the ship maintenance process. The familiar top five ship format is maintained, and the additional level of detail is organized by type expense. The baseline is the actual cost incurred and the compared value is the backcasted effect that 3DP would have had on actual cost.

To the executive-level decision-maker analyzing the effect of 3DP on U.S. Navy ship maintenance, this figure provides two important pieces of information. The first is that overall, the actual cost of ship maintenance can be reduced with the implementation of 3DP technology. Figure 34's center instance shows that the effect of 3DP on the ship maintenance process could have reduced the total cost by 2.1% as is indicated by the percentage and the green shell. The cost of ship maintenance with the incorporation of 3DP is now $426.2 million versus the original $435.5 million for a savings of $9.3 million. The *Barry*, again located at the one-o'clock position, is the ship that demonstrates the largest percentage cost savings at 6.1% and reduced costs across all types of expense.

Second, not every ship may benefit from the use of 3DP technology. Just above the three-o'clock position on Figure 34, the *Stout* indicates 1.4% cost growth for a backcasted total cost of $64.1 million, or $0.9 million greater than the original cost. Drilling down one level of detail into expense type, the decision-maker can easily determine that every category of expense contributed to the cost growth for the *Stout*. However, additional levels of detail are available, and the executive may request that more information be displayed to help identify the primary drivers of cost growth for the *Stout* and/or the leading sources of cost savings for the *Barry*. Therefore, Figure 35, organized by work, adds another layer of detail.

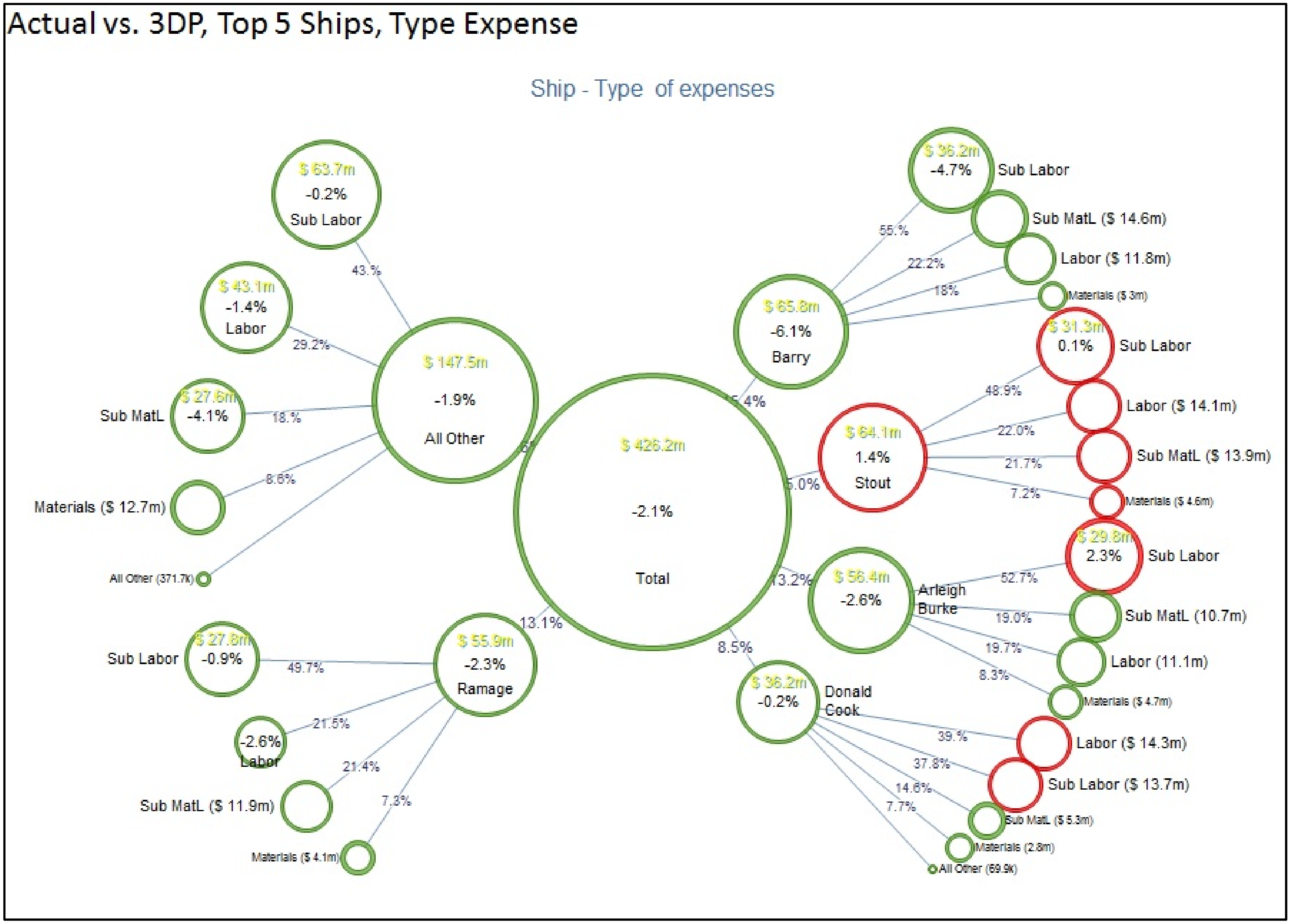


Figure 34. Actual Versus 3DP, Top 5 Ships, Type Expense Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Actual Versus 3DP of the Top 5 Ships by Type Expense, Work

Figure 35, Actual Versus 3DP, Top 5 Ships, Type Expense, Work, is the second in the series of comparison figures. It allows the decision-maker to visually drill down into the data even further. In Figure 34, the *Barry* displayed cost savings across all types of expense, and the *Stout* indicated cost growth with the implementation of 3DP into the ship maintenance process. The addition of the classification of work detail however, indicated where each ship derived its savings or growth with 3DP.

The executive drilling down into the 3DP backcasted cost data for the *Barry* can quickly identify one classification of work, in one type of expense, which produced cost growth. The only red-shelled solar graph instance subordinate to the *Barry* in Figure 35 is the NG instance, subordinate to Sub Labor, which has been backcasted to account for $1.3 million dollars of Sub Labor cost. However, the percentage growth is not displayed because the software limits the presence of information to reduce clutter and increase clarity. But, the executive requiring more information can simply select the red-shelled NG instance, and more information is available immediately, including the percentage of cost growth. If the decision-maker decided to implement the 3DP-only strategy, then the NG work attributed to Sub Labor could be an aspect that should be looked at for improvement.

The executive examining the *Stout*, at the two-o'clock position in Figure 35, more closely can quickly see that even though the aggregate change in cost is cost growth, there are indications of possible cost savings. Immediately subordinate to the Stout, Sub Labor is backcasted to account for $31.3 million. Again, the percentage increase in cost is not displayed, but is available by selecting the solar graph instance. Even though the Sub Labor instance indicates cost growth, there are children subordinate to Sub Labor that signify cost savings. For example, the Growth instance is green shelled and is backcasted to account for $6.4 million. To the decision-maker, this figure is forecasting the possible effect of implementing 3DP into ship maintenance using historical data, and it provides the ability to examine the effect a particular technology might have on cost without the risk and uncertainty involved with actual implementation.

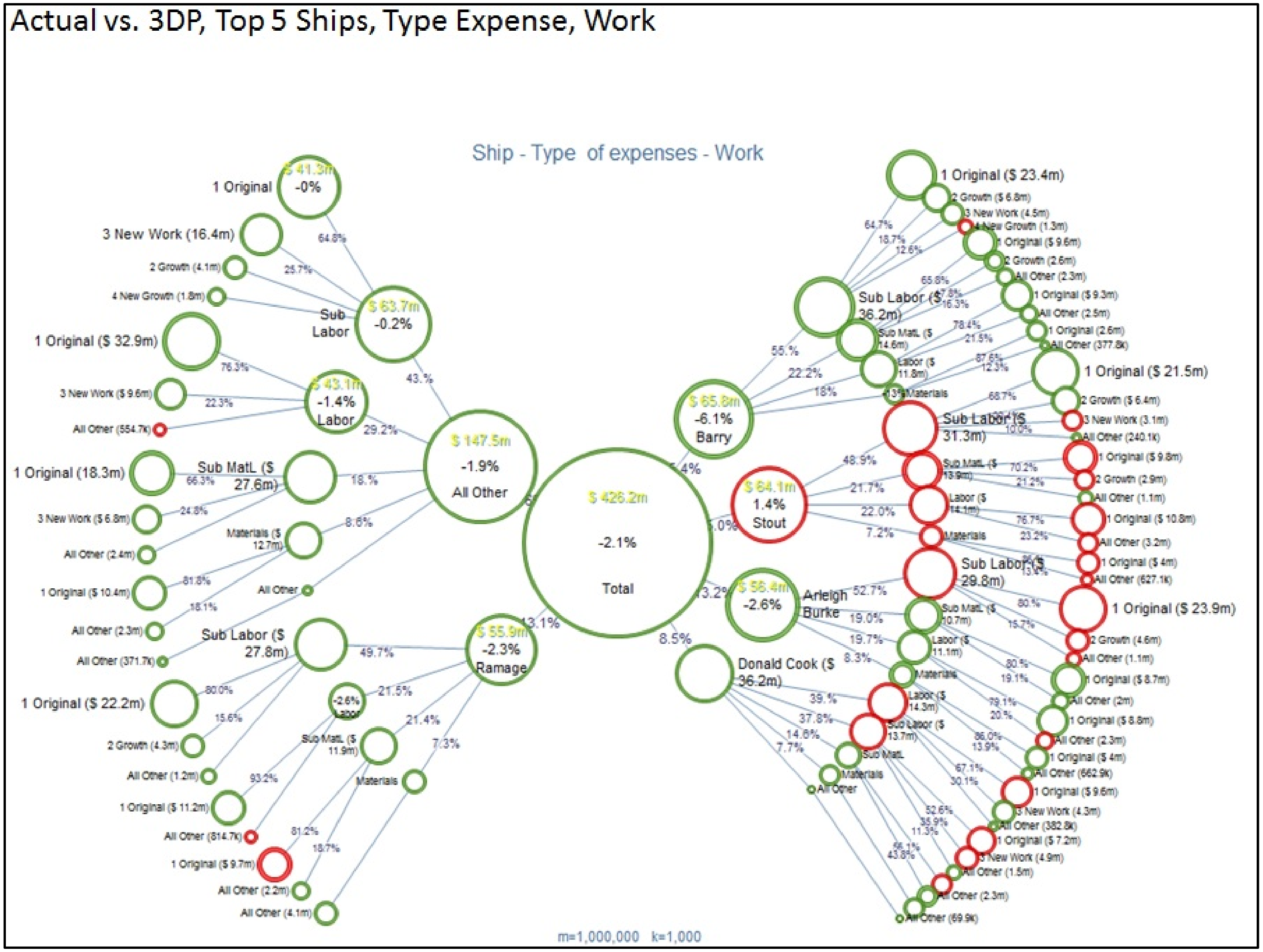


Figure 35. Actual Versus 3DP, Top 5 Ships, Type Expense, Work Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Actual Versus AM Radical of the Top 5 Ships by Type Expense

Figures 36 and 37 compare the baseline, actual cost, to the backcasted effect that the implementation of all three combined technologies might have had on cost. The structure of the graph remains familiar, but the increase in cost savings demonstrates the ability of the software to produce intuitive screenshots that easily communicate the differences in effect on cost. Figure 36, Actual Versus AM Radical, Top 5 Ships, Type Expense reverts back to the format of Actual Versus 3DP, Top 5 Ships, Type Expense (Figure 34) with less detail, but easily demonstrates the difference in cost savings. The substantial increase in cost savings is communicated by, most intuitively, the thickness of the solar graph instance shells, but is also indicated by the absolute and relative values displayed in or near the instance.

To the executive-level decision-maker concerned with cost, the most evident display of cost savings is the center instance. The total backcasted cost of ship maintenance for all 19 ships, had 3DP, 3D LST, and CPLM technologies been implemented, is $271.1 million or 37.7% cost savings under actual cost. The difference, $164.4 million, could have been used to finance other needs such as system upgrades, structural improvements, or reducing the number of maintenance jobs deferred until the next availability due to shrinking fiscal budgets. The decision-maker analyzing the change in cost might also be interested in understanding the difference in cost savings of individual ships.

In contrast to the 3DP-only implementation strategy, which slightly increased cost for one of the top five ships, AM Radical decreased costs for all top five ships. There appear to be substantial cost savings in the All Other solar graph instance located at the ten-o'clock position in Figure 36 as well, but current settings prevent concluding that all 19 ships incurred cost savings. In the case of the *Barry*, cost savings is significantly increased. With implementation of 3DP only, the backcasted cost was $65.8 million, or 6.1% cost savings. With the use of all three technologies, or AM Radical implementation, the backcasted cost for the *Barry* is $43.9 million, a cost savings of 37.3% when compared with actual cost. Drilling down one layer of detail, two of the type expense children subordinate to the *Barry* have thicker green shells than the others, an intuitive indication of substantial cost savings. In fact, Sub Labor and Sub Material account for almost 80% of the increase in the cost savings of AM Radical over the 3DP-only implementation strategy for the *Barry*.

Earlier, in the description of the Definitized Estimate Versus Actual, Top 5 Ships, Type Expense (Figure 32) solar graph, the executive-level decision-maker identified subcontractor labor and material as primary drivers of cost growth. The keen decision-maker might begin to formulate that a possible solution to subcontractor labor and material cost growth is the implementation of all three technologies. However, the addition of another layer of detail is available, and it could provide either supporting or contradictory evidence.

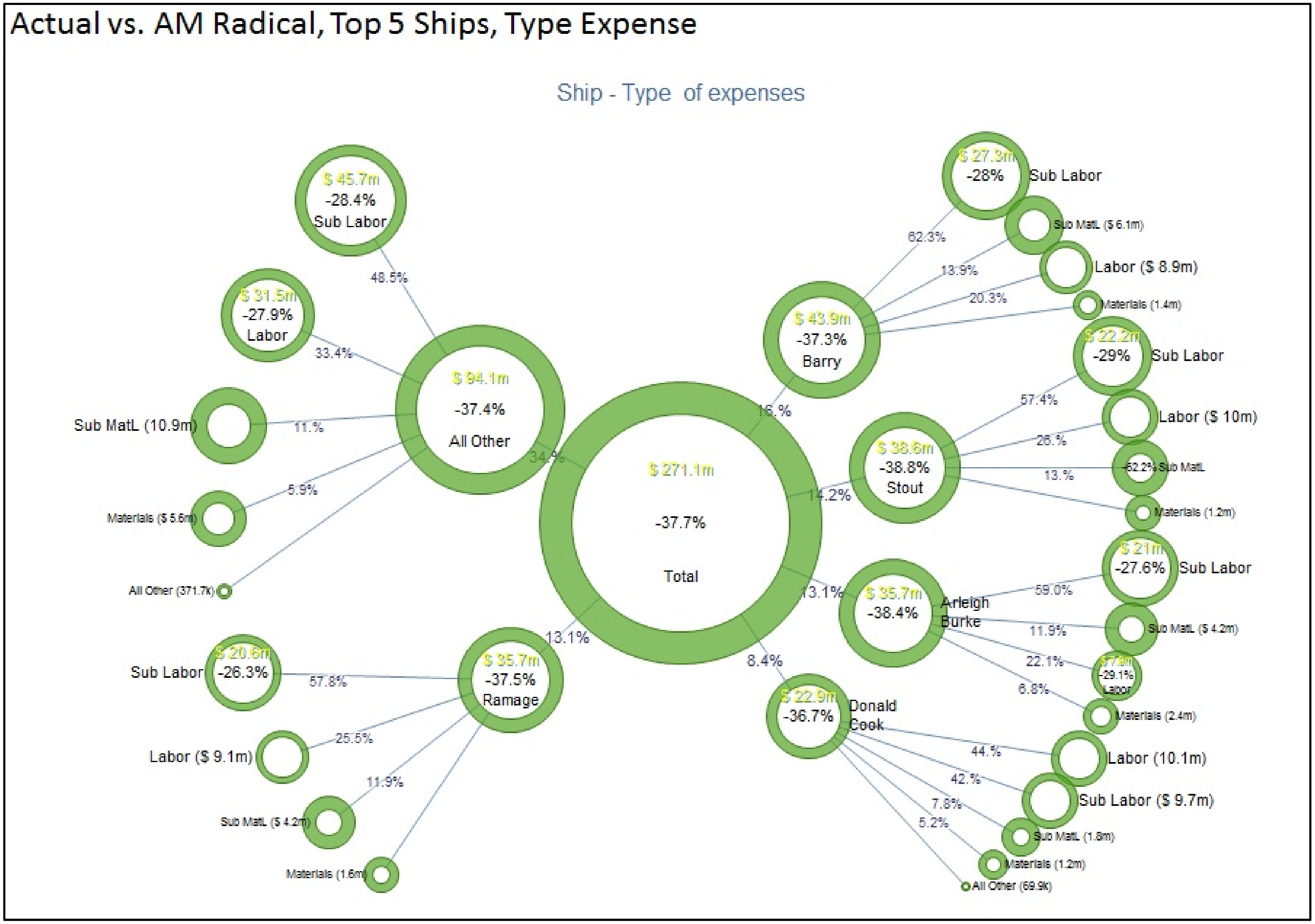


Figure 36. Actual Versus AM Radical, Top 5 Ships, Type Expense Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Actual Versus AM Radical of the Top 5 Ships by Type Expense, Work

Figure 37, Actual Versus AM Radical, Top 5 Ships, Type Expense, Work, the fourth and final screenshot of this comparison series, allows the executive-level decision-maker to drill down visually into the data even further. The additional layer of detail is organized by work and provides more information about the sources of cost savings.

An executive analyzing this figure could notice the most obvious aspect first, the fact that AM Radical implementation creates cost savings throughout the entire dataset. Whereas 3DP-only implementation indicated cost growth in one ship, various type expenses and classifications of work, the backcasted effect AM Radical implementation could have produces cost savings in every instance. For example, with 3DP-only implementation, the Actual Versus 3DP, Top 5 Ships, Type Expense, Work solar graph (Figure 35) identified cost growth in one classification of work, NG, which accounted for $1.3 million of Sub Labor. But with AM Radical implementation, the NG instance subordinate to the *Barry* on this solar graph, Figure 37, indicates cost savings and now accounts for $0.99 million. As stated before, the percentage change is not displayed to reduce clutter; however, it is available by simply selecting the instance. Possibly more interesting to the executive-level decision-maker is the case of the *Stout* which changed from a source of cost growth to a significant driver of cost savings.

In the previous solar graph, Actual Versus 3DP, Top 5 Ships, Type Expense, Work (Figure 35), showing the backcasted effect of 3DP, the *Stout* displayed an absolute cost of $64.1 million and cost growth of 1.4%. The classification of work which contributed most to the cost of the *Stout* was OW, a child of Sub Labor, and indicated an absolute cost of $21.5 million. But with AM Radical implementation, this solar graph (Figure 37) backcasted the cost to $15.3 million for the OW associated with Sub Labor, a cost savings of $6.2 million when compared to 3DP-only implementation.

The *Stout*, as well as the other top five ships, could have produced significant cost savings had the AM Radical approach been implemented. However, the actual costs have already been incurred. The significance of this series of figures is that a decision-maker can visualize the effect the technology implementation strategies might have had on historical data and then make predictions about the effect on future costs. The decision-maker, armed with the predictions derived from the solar graphs, weighs additional executive-level organizational considerations, and then is able to make better cost-control choices for the future of U.S. Navy ship maintenance.

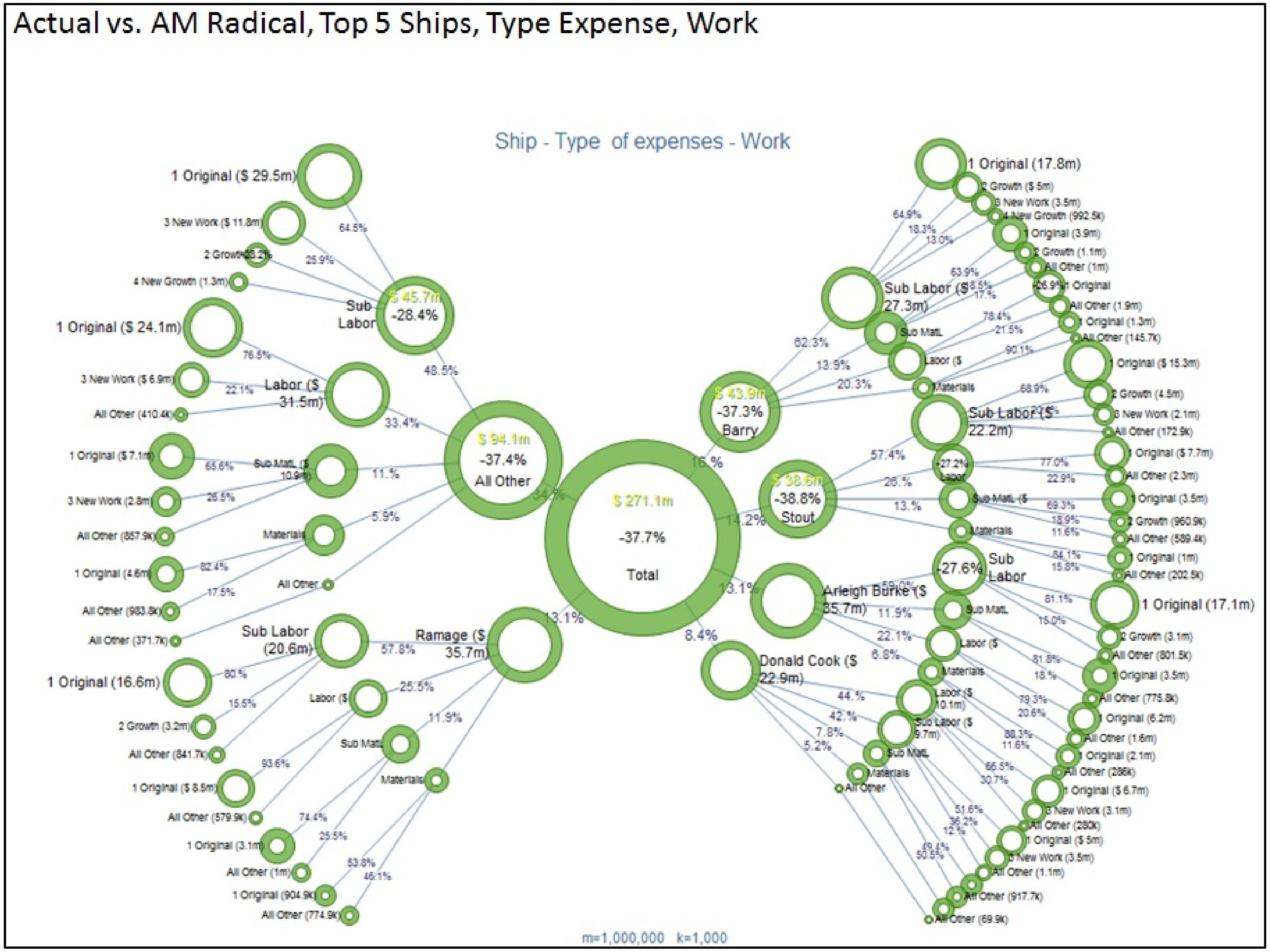


Figure 37. Actual Versus AM Radical, Top 5 Ships, Type Expense, Work Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Alternative Figures

The final series of solar graph figures (Figures 38–40) demonstrates the flexibility of visualization tools, enabling drilling down into specific details. All but the first of the figures described thus far have used the top five ships structuring concept for the first level of detail. While the use of the single method of organizing the first layer of detail made comprehension of the graphs easier, it limited the appearance of the flexibility of the third-party software. Therefore, other methods of organizing and presenting the data are explored in the following three comparison figures.

Definitized Estimate Versus Actual of the Type Expense by Work

Figure 38, Definitized Estimate Versus Actual, Type Expense, Work, is useful for the decision-maker interested in analyzing cost growth without discriminating by ship. This figure reverts back to using the definitized estimate as the baseline and the actual cost as the comparison, as in Definitized Estimate Versus Actual, Top 5 Ships, Type Expense and Definitized Estimate Versus Actual, Top 5 Ships, Work (Figures 32 and 33, respectively). However, the top five ships are not used as an organizing concept. Instead, the first layer of detail is grouped by type expense and the additional layer is organized by work.

Consider the theory arrived at by the executive during analysis of the Definitized Estimate Versus Actual, Top 5 Ships, Type Expense figure (Figure 32). The decision-maker noted that subcontractor labor and material appeared to be primary drivers of cost growth. In this screenshot, Figure 38, the Sub Labor instance appears at ten o'clock and the Sub Material instance at three o'clock. The indicated percentages of cost growth are 45% and 44%, respectively. Compared to the cost growth of Labor and Materials associated with the shipyard, 28.5% and 27.1%, respectively, subcontractors also appear here to be primary drivers of cost growth. The decision-maker is interested in understanding the causes of subcontractor cost growth at a deeper level of detail. Therefore, the executive might analyze the graph further and discover that NW is the largest absolute contributor to both Sub Labor, at $31.5 million, and Sub Material, at $13.5 million.

The same information was derived from the analysis of two sequential solar graphs described earlier. Those were Definitized Estimate Versus Actual, Top 5 Ships, Type Expense and Definitized Estimate Versus Actual, Top 5 Ships, Work (Figures 32 and 33, respectively). The same understanding was derived from two unique presentations, one with two graphs and the other with this one graph. Arriving at the same conclusion from different presentations of the data builds confidence in the decision-maker that the data are accurate and the visualization methods, valid.

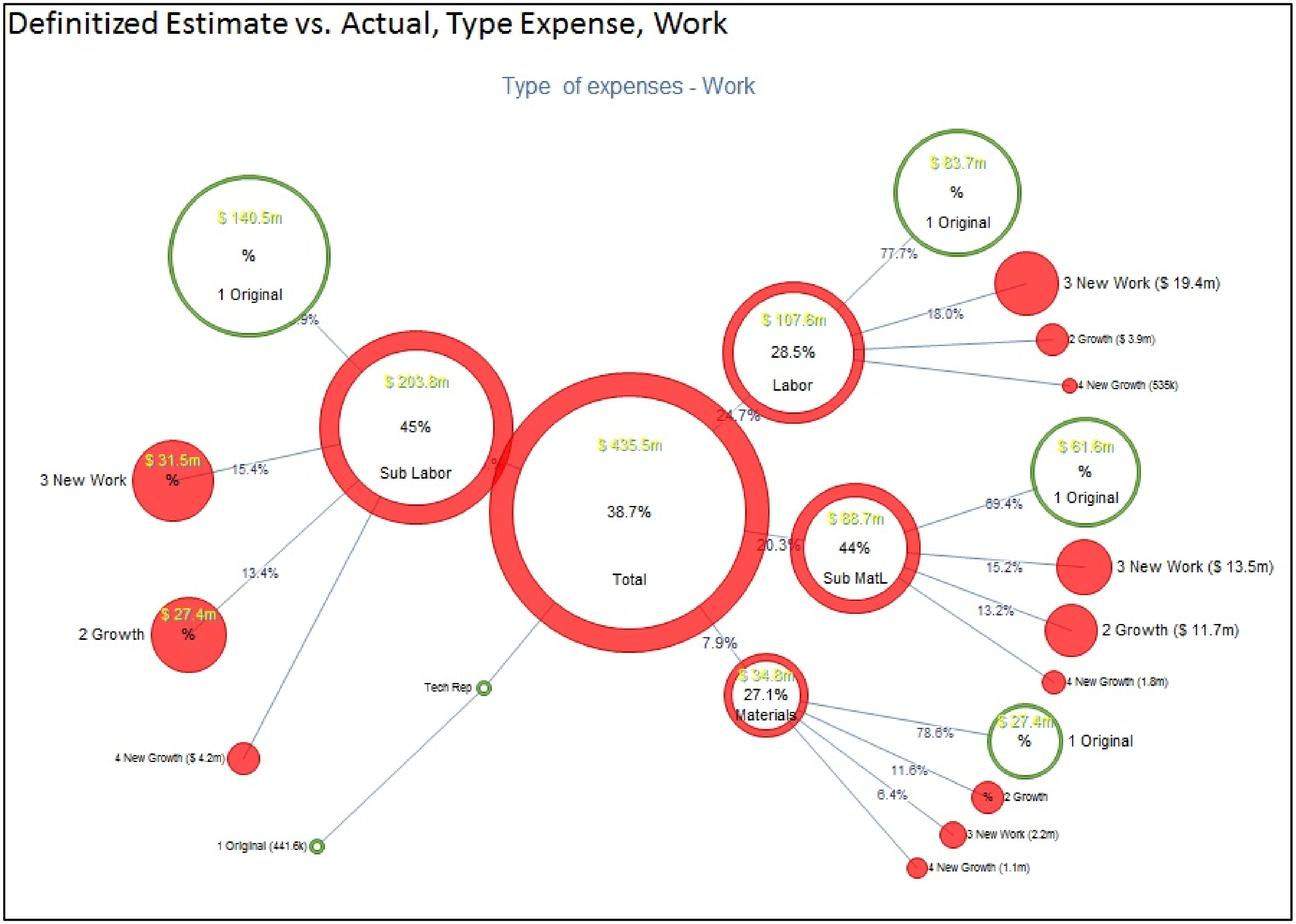


Figure 38. Definitized Estimate Versus Actual, Type Expense, Work Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Definitized Estimate Versus Actual of the Work by Ship

The remaining two alternative figures are complementary. Figure 39, Definitized Estimate Versus Actual, Work, Ship, is a figure that a decision-maker could use to identify problem areas of cost growth based on classification of work. Figure 40, Actual Versus AM Radical, Work, Ship, keeps the same organization format, but enables the decision-maker to analyze how the implementation of the three technologies could have created cost savings.

In Figure 39, which demonstrates cost growth, there is only one peculiarity that has already been explained. All the thin, green-shelled instances on the left side of the graph represent 0% growth because of the definition of OW, which cannot grow in expense. Also, the solid, red-shelled instances on the right side represent only cost growth that occurred and are classified as either NW, G, or NG because of their definitions.

Figure 39 is important to the executive-level decision-maker because it exhibits data already presented in another format. The other format was the Definitized Estimate Versus Actual, Top 5 Ships, Work figure (Figure 33), which organized the first level of detail by ship and the second level by work. In this graph, the organizing concepts have been reversed. If the same deduction can be derived from this screenshot, then the decision-maker's confidence, in their ability to make accurate and valid choices for the future of U.S. Navy ship maintenance processes, increases.

The deduction already made by the decision-maker was that NW, over the other classifications of work, accounted for the largest portion of cost growth. Referring to the Definitized Estimate Versus Actual, Work, Ship solar graph (Figure 39), a quick visual scan over the classification of work instances creates an intuitive understanding. The NW instance is the largest indicator of cost growth. Further examination by the decision-maker provides the dollar values that support the intuitive perception. The NG instance, located at the five-o'clock position, accounted for $7.7 million. The G instance, located just below the three-o'clock position, represented $47.1 million. Finally, the NW instance, located at the two-o'clock position, produced $66.8 million in cost growth. Even though the data were organized and presented differently, the same deduction was reached: NW was the primary driver of cost growth.

If the executive were interested in determining the ships that produced the largest cost growth, then simply referring to the additional level of detail would provide the answer. For example, since NW was the primary driver of cost growth, identification of the largest contributing ship may provide a specific case for further analysis of cost growth. Referring to the NW instance, located at the two-o'clock position in Figure 39, the child ship that represents the largest portion of cost growth is the *Donald Cook*. The decision-maker, remembering that the *Donald Cook* represents two availabilities, would drill down into the next level of detail by selecting the *Donald Cook*. Then, the determination would be made whether either one of the *Donald Cook* availabilities or the next largest individual ship (the *Barry*) was the ship representing the most cost growth for NW. Once the ship was identified, the executive could direct further study into the causes of cost growth.

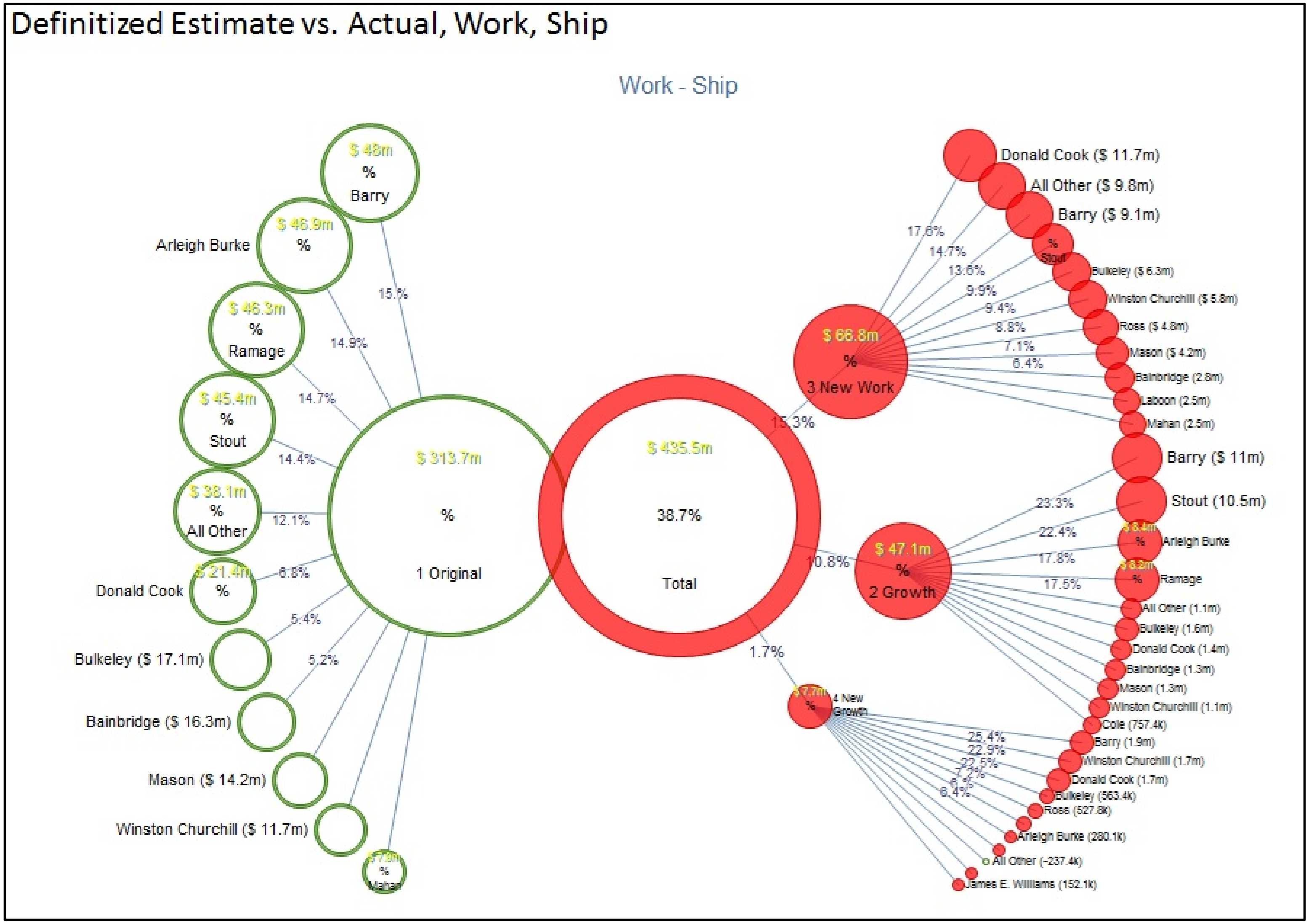


Figure 39. Definitized Estimate Versus Actual, Work, Ship Solar Graph
(J. Kornitsky, personal communication, November, 2013)

Actual Versus AM Radical of the Work by Ship

Actual Versus AM Radical, Work, Ship (Figure 40) shows the backcasted effect that the implementation of all three technologies combined might have had on U.S. Navy ship maintenance costs. Figure 40 maintains the organizing structure of the immediately previous solar graph to provide easy comparison for the executive-level decision-maker.

For example, the decision-maker is interested in figuring out the overall effect that AM Radical implementation has compared to definitized cost. The center instance in Figure 40, Total, indicates the bottom-line cost savings that may have occurred had the AM Radical implementation strategy been employed. At $271.1 million, AM Radical implementation might have resulted in 37.7% cost savings, but that's compared to actual cost. Referring to the OW instance located at the nine-o'clock position on the previous solar graph, Definitized Estimate Versus Actual, Work, Ship (Figure 25) the value is $313.7 million. Because of the definition of OW and the position of the OW instance at the first level of detail, it also represents the total definitized estimate. Simple math shows that AM Radical implementation might have caused the ships analyzed to come in under budget by $42.6 million, or 13.6%. Cost growth could have been turned into cost savings through the backcasted effect that AM Radical implementation might have had on the ships studied. To the executive-level decision-maker, this is important because if the three technologies (3DP, 3D LST, and CPLM combined) were selected for implementation, then future U.S. Navy ship maintenance budgets might be reduced and result in reallocation of funding to higher priority projects.

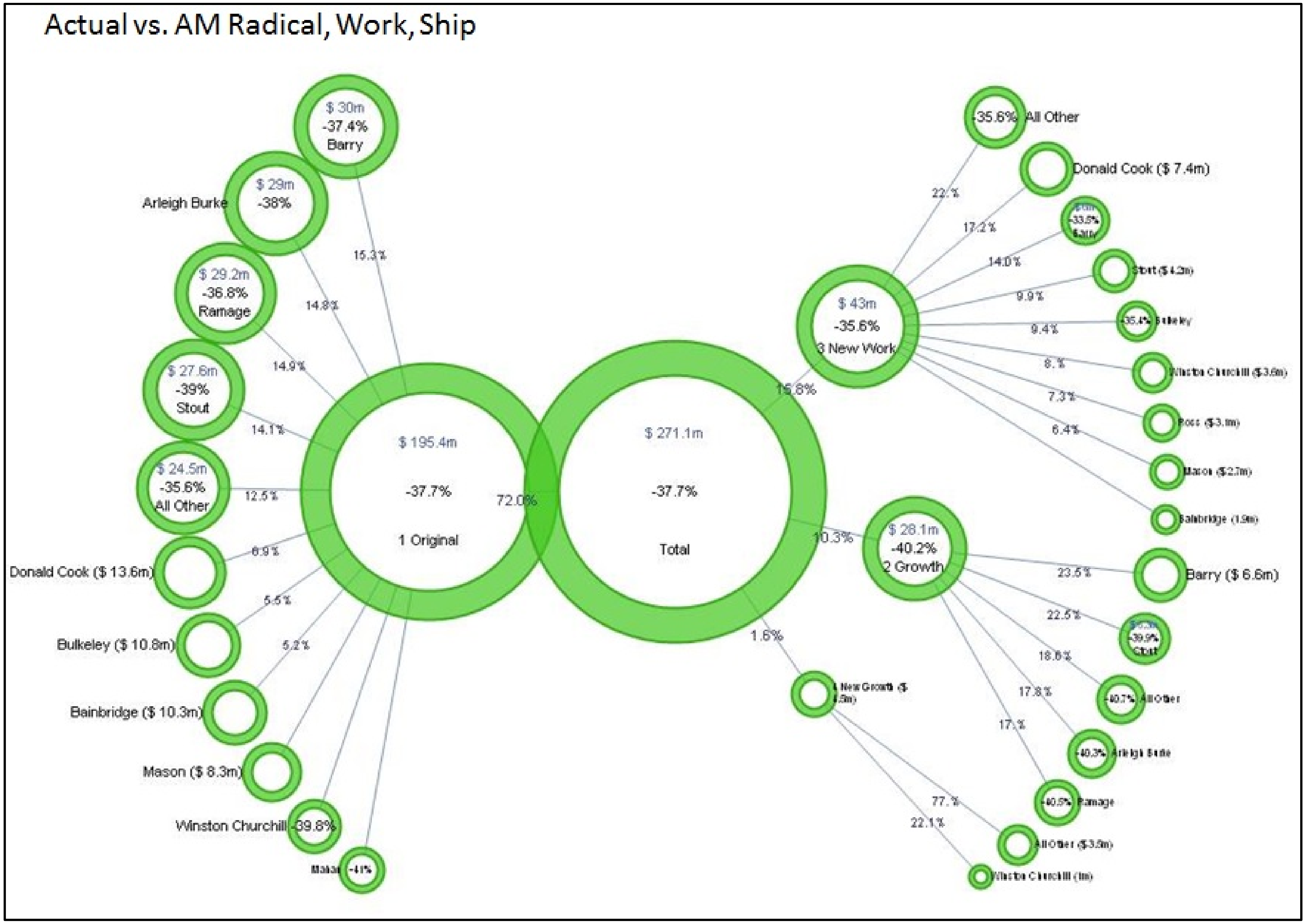


Figure 40. Actual Versus AM Radical, Work, Ship Solar Graph
(J. Kornitsky, personal communication, November, 2013)

LOD and Availability Density Bubble Charts

Lost operating days (LOD) have long been considered by the U.S. Navy ship maintenance metrics groups to be a valuable indication of the performance of the ship maintenance process. The LOD metric is often included in reports made by regional maintenance centers (RMCs) to Naval Sea Systems Command (NAVSEA) as an indication of the effect on ship's schedule caused by delays (M. Leftwich, personal communication, September 4, 2013). However, the LOD metric has been linked to the quality of the definitized estimate and that quality has been determined to be random (T. Laverghetta, personal communication, November 26, 2013). To the executive-level decision-maker, the important aspect of ship maintenance is cost. Availability density is considered a better metric for predicting cost and was provided to this study for further analysis (P. Pascanik, personal communication, November 21, 2013). Of the following three figures (Figures 41–43), the first two highlight the lack of correlation between the LOD metric and actual cost. The third demonstrates the validity of using the availability density metric to indicate actual maintenance cost.

LOD Versus Expense (Actual Cost)

Figure 41, LOD Versus Expense (Actual Cost), is presented first to provide an introduction to the structure of the chart. The next figure, LOD Versus Expense (Actual Cost) – Highlighted (Figure 42), will then be considered as the chart important to the U.S. Navy ship maintenance executive-level decision-maker interested in controlling costs.

In Figure 41, the LOD Versus Expense (Actual Cost) figure is structured as an XY scatter plot. The X-axis represents expense or actual cost of a ship availability and ranges from $0 to $70 million. The Y-axis represents the total LODs incurred during an availability and ranges from 0 to (-107); negative numbers represent operating days *lost*. The data points scattered throughout the chart represent the LOD and expense values for the individual availabilities and are labeled with their unique availability identification numbers. For example, the data point labeled Avail 56387 near the center of the bubble chart represents one of the availabilities for the *Donald Cook*. The LODs incurred during that availability totaled 63, and the total expense was $33.4 million.

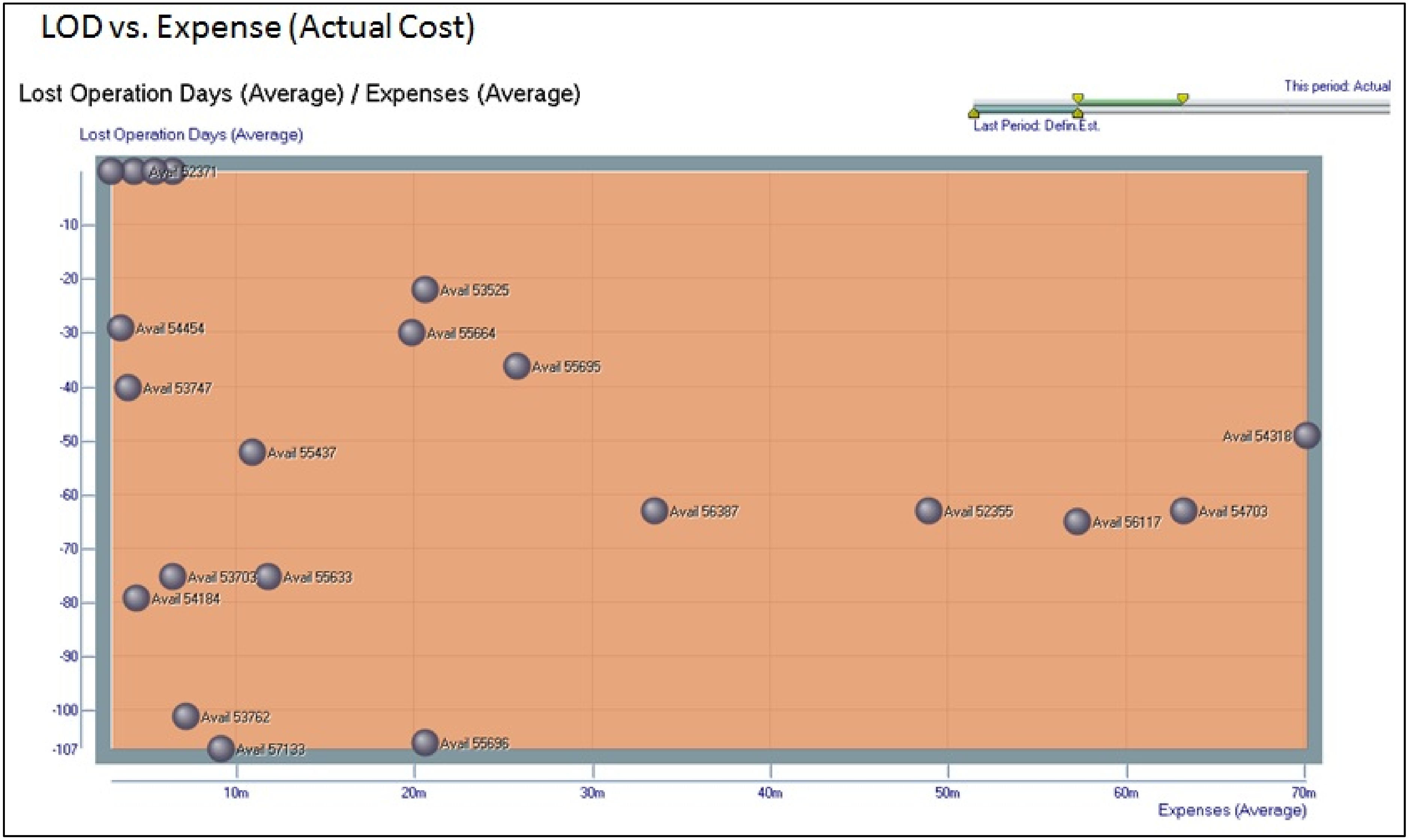


Figure 41. LOD Versus Expense (Actual Cost) Bubble Chart
(J. Kornitsky, personal communication, November, 2013)

LOD Versus Expense (Actual Cost) – Highlighted

Figure 42, LOD Versus Expense (Actual Cost) – Highlighted, is important to the executive-level decision-maker because it demonstrates that the LOD metric is not useful for forecasting the actual cost of an availability. This is shown by both a visual analysis of the chart and by a mathematical calculation of the correlation factor.

Visually, the data points show that smaller availabilities, under $30 million, can result in either the highest number or the lowest number of LODs. For example, the data point labeled Avail 52371 in Figure 42 is for the *James E. Williams* and indicates an actual cost of $4.2 million with a total of zero LODs. Meanwhile, the data point labeled Avail 57133 is for one of the *Arleigh Burke* availabilities and indicates an actual cost of $9.1 million with a total of 107 LODs. In fact, the six data points highlighted in the lower left corner of the bubble chart all represent availabilities of relatively small cost that incurred relatively high numbers of LODs, which prevent the appearance of a linear relationship. Therefore, the LOD metric is not a good indicator of availability cost.

Mathematical calculation also demonstrates the lack of connection between the LOD metric and expense. The expense, or actual cost, of each availability was totaled, to include OW, G, NW, and NG. Then, a correlation factor was calculated between the cost of each availability and the number of LODs incurred during each availability. The correlation factor is (-0.14). This number shows that, mathematically, the LOD metric is not a good indicator of cost. For the executive-level decision-maker, LODs have been visually and mathematically shown not to correlate well with cost. However, the metric that correlates well with cost is the metric provided to this study for further analysis.

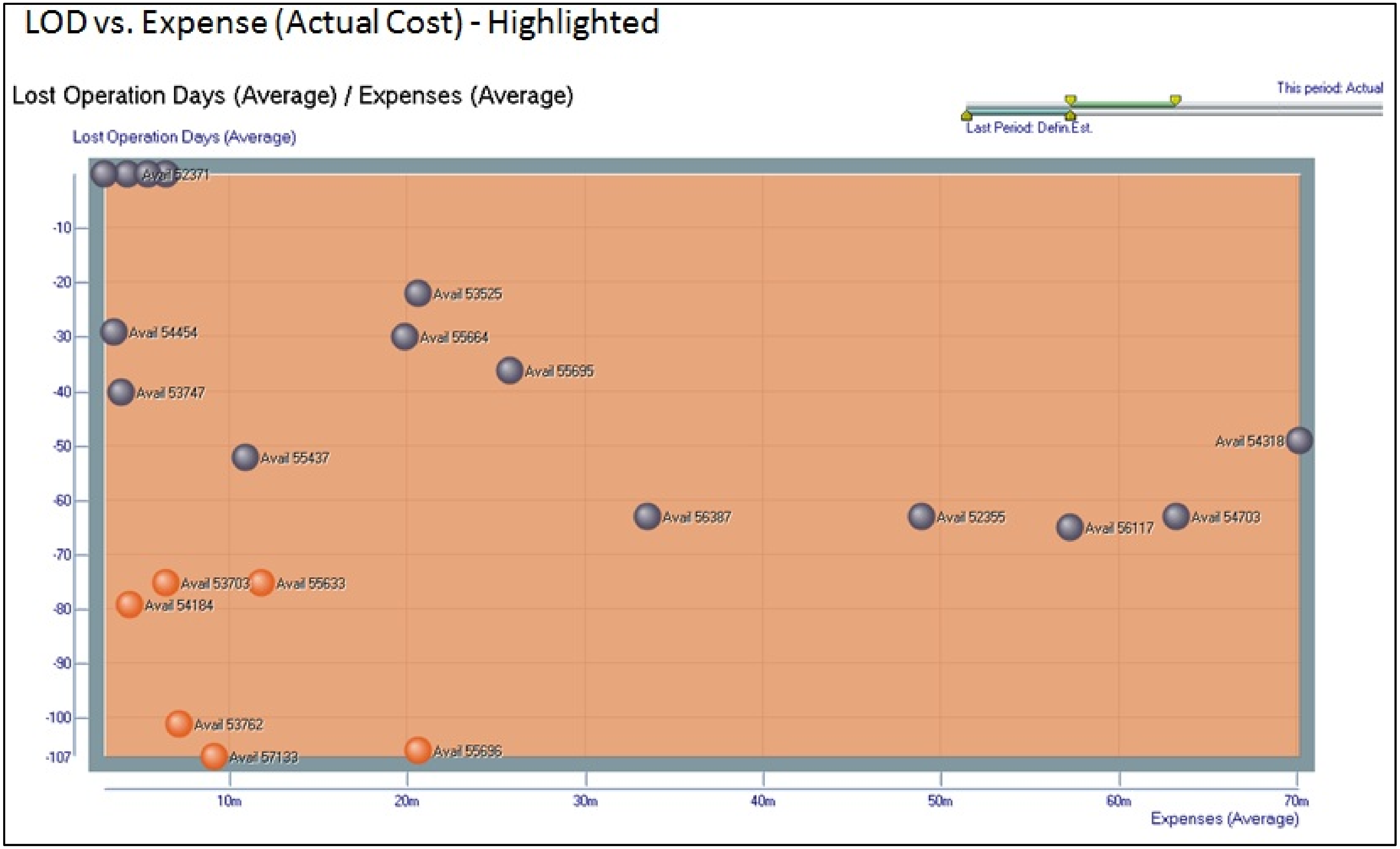


Figure 42. LOD Versus Expense (Actual Cost)—Highlighted Bubble Chart
(J. Kornitsky, personal communication, November, 2013)

Availability Density Versus Expense (Actual Cost)

The metric of availability density was provided to this study and is defined as the Total Actual Man Days divided by the Total Availability Duration Days. In other words, the availability density number represents the average number of man-days performed each calendar day of the availability. For example, in Figure 43, the *Stout* (Avail 54703) used 134,254 man-days to complete its availability, which lasted 275 calendar days. The availability density for the *Stout* is 488. For each calendar day of the *Stout*'s availability, an average of 488 man-days were performed.

Figure 43, Availability Density Versus Expense (Actual Cost), changes one axis to represent the new metric. The X-axis remains as expense, but the Y-axis now represents availability density and ranges from 55 to 611. For example, the *Stout* (Avail 54703) data point near the top-right corner of the chart indicates an average of 488 man-days per availability calendar day and an actual cost of $63.2 million.

Availability density is a better indicator of cost and the Availability Density versus Expense (Actual Cost) bubble chart (Figure 43) proves that visually and mathematically. Visually, availability density correlates with expense. For example, the data point labeled Avail 57133 near the bottom-left portion of the chart represents one of the *Arleigh Burke* availabilities and indicates an availability density of 85 and expense of $9.1 million. In the diagonally opposite corner, *Avail 54318* represents the *Barry* and indicates an availability density of 611 and expense of $70.1 million. Visually, availability density provides a good indication of availability expense through its linear response.

Mathematically, availability density and cost correlate very well. The expense of each availability was again totaled. Then, a correlation factor was calculated between the cost and availability density of each availability. The correlation factor is 0.98. This number shows that, mathematically, the availability density metric is a strong indicator of cost. Availability density is visually and mathematically an accurate indicator of cost.

For the executive-level decision-maker, predicting the actual cost of events in progress is extremely valuable. The metric availability density shows such a strong correlation to cost that it may be able to predict whether a particular current availability is expected to meet or exceed the definitized estimate. The ability to predict the ending actual cost of a ship maintenance evolution in progress would enable decision-makers to avert large cost growth by implementing changes earlier in the U.S. Navy ship maintenance process.

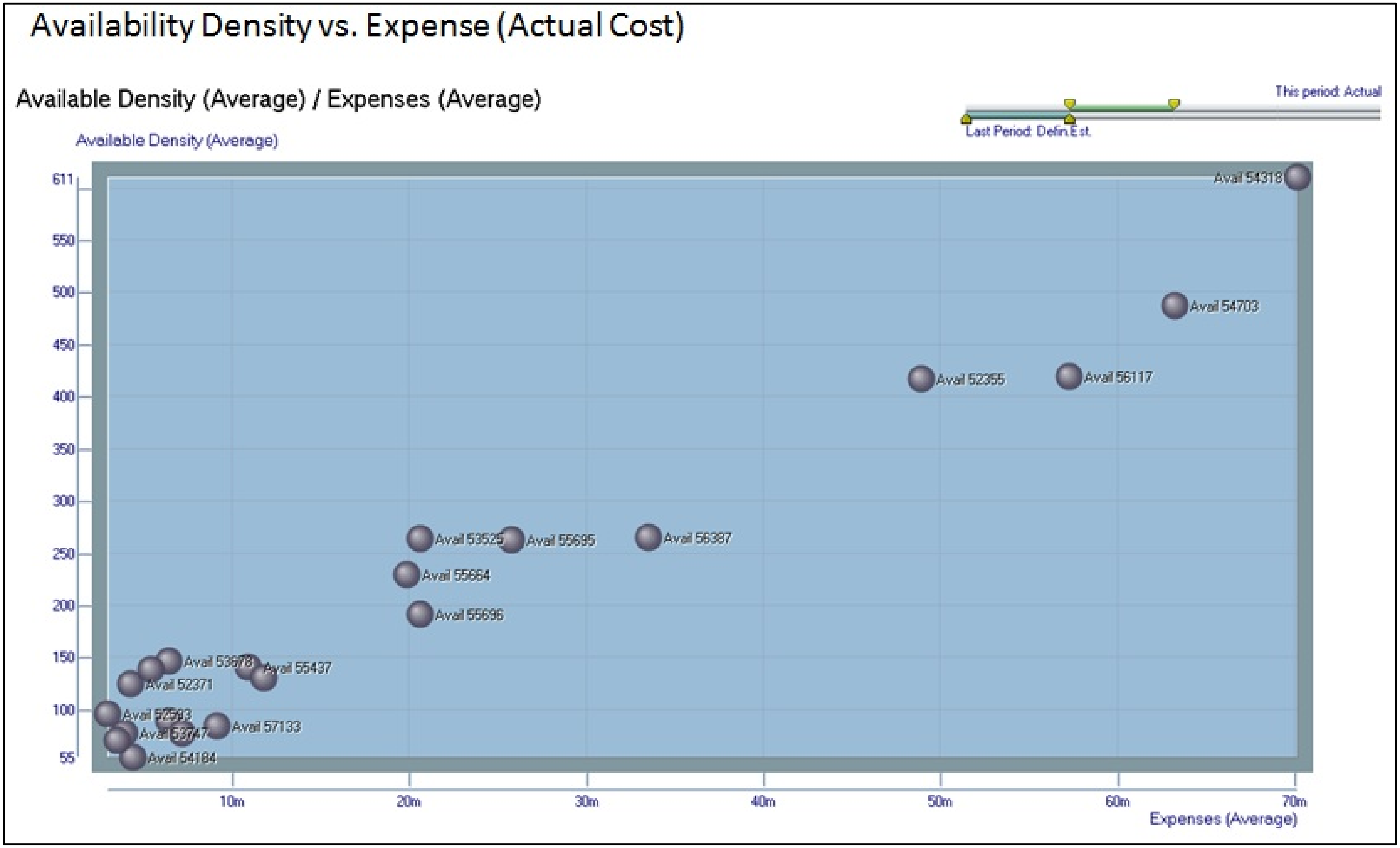


Figure 43. Availability Density Versus Expense (Actual Cost) Bubble Chart
(J. Kornitsky, personal communication, November, 2013)

Drill-Down Spreadsheets

Figures 44 and 45 are examples of drill-down spreadsheets that can be produced from any solar graph instance. The ship selected as the target for drill down was the *Barry* and a time-analysis spreadsheet was produced at two different levels of detail. The time analysis covers the actual cost, 3DP-only implementation backcasted cost, and AM Radical implementation backcasted cost. The titles of each spreadsheet indicate the levels of detail shown; *Barry* Drill Down, 3 Levels of Detail (Figure 44) shows three levels of detail, and *Barry* Drill Down, 4 Levels of Detail (Figure 45) shows four.

These spreadsheets would be valuable to the executive-level decision-maker who wanted to see the numbers that the visualization software translates into intuitive solar graphs. For example, consider the executive analyzing the Actual Versus AM Radical, Top 5 Ships, Type Expense solar graph, Figure 36. The *Barry* instance, located at the one-o'clock position on that graph, indicates a backcasted absolute cost of $43.9 million if all three technologies had been implemented into the ship maintenance process. However, the decision-maker wants to see the absolute values for the actual cost, the 3DP-only backcasted cost, and the AM Radical backcasted cost together for comparison. Then, simple selection of the *Barry* instance would produce the option to generate detailed spreadsheets at varying levels of detail. If the decision-maker only wanted to see a little additional detail, then the *Barry* Drill Down, 3 Levels of Detail (see Figure 44) option might be selected. If the decision-maker really wanted to drill down into the data, then the *Barry* Drill Down, 4 Levels of Detail (see Figure 45) spreadsheet could be generated. Either way, these spreadsheets provide the executive-level decision-maker with drill-down capability sufficient to meet the needs of the most detail-oriented executive.

Barry Drill Down, 3 Levels of Detail

Barry / - Time analysis

Account			Totals				Actual	3DP	AM Radical
Code	Name	Level	Actual Am.	% Revenue	% Cost	% Growth	Amount	Amount	Amount
TEx	Expenses	1	43,941,901		100.00%	-37.35%	70,145,759	65,829,780	43,941,901
LBR	Labor	2	36,325,647		82.66%	-27.79%	50,310,106	48,128,487	36,325,647
LTG	Labor Total Growth (G+NW+NG)	3	11,522,533		26.22%	-27.95%	15,994,526	15,336,894	11,522,533
TOL	Total Original Labor	3	24,803,113		56.44%	-27.72%	34,315,580	32,791,593	24,803,113
MTR	Materials	2	7,616,254		17.33%	-61.60%	19,835,653	17,701,293	7,616,254
MTG	Material Total Growth (G+NW+NG)	3	2,357,236		5.36%	-61.52%	6,126,742	5,381,784	2,357,236
TOM	Total Original Material	3	5,259,017		11.96%	-61.63%	13,708,911	12,319,508	5,259,017
TG	Total Increase (G+NW+NG)	1	13,879,769		31.58%	-37.25%	22,121,268	20,718,678	13,879,769
TO	Total Original	1	30,062,131		68.41%	-37.40%	48,024,491	45,111,101	30,062,131

*** %Growth is AM Radical vs. Actual

Figure 44. *Barry* Drill Down, 3 Levels of Detail
(J. Kornitsky, personal communication, November, 2013)

Barry Drill Down, 4 Levels of Detail

Barry / - Time analysis

Account			Totals				Actual	3DP	AM Radical
Code	Name	Level	Actual Am.	% Revenue	% Cost	% Growth	Amount	Amount	Amount
TEx	Expenses	1	43,941,901		100.00%	-37.35%	70,145,759	65,829,780	43,941,901
LBR	Labor	2	36,325,647		82.66%	-27.79%	50,310,106	48,128,487	36,325,647
LTG	Labor Total Growth (G+NW+NG)	3	11,522,533		26.22%	-27.95%	15,994,526	15,336,894	11,522,533
TLG	Total Growth Labor (Original)	4	5,389,107		12.26%	-29.80%	7,677,244	7,298,394	5,389,107
TLNG	Total Labor New Growth	4	1,026,852		2.33%	-26.03%	1,388,275	1,414,158	1,026,852
TLNW	Total Labor New Work	4	5,106,572		11.62%	-26.30%	6,929,007	6,624,341	5,106,572
TOL	Total Original Labor	3	24,803,113		56.44%	-27.72%	34,315,580	32,791,593	24,803,113
OL	Original Labor	4	7,000,423		15.93%	-27.00%	9,589,621	9,301,932	7,000,423
OSL	Original SubLabor	4	17,802,690		40.51%	-28.00%	24,725,959	23,489,661	17,802,690
MTR	Materials	2	7,616,254		17.33%	-61.60%	19,835,653	17,701,293	7,616,254
MTG	Material Total Growth (G+NW+NG)	3	2,357,236		5.36%	-61.52%	6,126,742	5,381,784	2,357,236
TMG	Total Growth Material (Original)	4	1,240,520		2.82%	-63.08%	3,360,462	2,882,044	1,240,520
TMNG	Total Material New Growth	4	164,808		0.37%	-71.73%	583,025	531,539	164,808
TMNW	Total Material New Work	4	951,907		2.16%	-56.39%	2,183,255	1,968,201	951,907
TOM	Total Original Material	3	5,259,017		11.96%	-61.63%	13,708,911	12,319,508	5,259,017
OM	Original Material	4	1,338,166		3.04%	-57.00%	3,112,015	2,676,332	1,338,166
OSM	Original SubMaterial	4	3,920,851		8.92%	-63.00%	10,596,896	9,643,175	3,920,851
TG	Total Increase (G+NW+NG)	1	13,879,769		31.58%	-37.25%	22,121,268	20,718,678	13,879,769
TO	Total Original	1	30,062,131		68.41%	-37.40%	48,024,491	45,111,101	30,062,131

*** %Growth is AM Radical vs. Actual

Figure 45. *Barry* Drill Down, 4 Levels of Detail
(J. Kornitsky, personal communication, November, 2013)

Summary

Visualization tools make it easier for executive-level decision-makers to determine the status, path, and origin of ship maintenance costs. NPS researchers were asked to identify new ways of summarizing millions of data points that are critical in making maintenance decisions. The visualization software illustrated how additional tools for big data could provide the diagrams, charts, and graphs to facilitate maintenance costs allocation decisions. In addition, we have provided a methodology to help mitigate the risk and uncertainty in decision-making.

The big data collected and stored by ST1MS is a giant trove of information. In this limited study, only 19 DDGs and 21 availabilities were analyzed. Primary drivers of cost growth and possible sources of cost savings were identified. Consider the use of big data visualization methods for not just every ship in the U.S. Navy, but also in the U.S. Coast Guard. These methods could also be applied to aircraft and ground vehicle maintenance. The scope is expandable to any system that collects big data. The ability to intuitively analyze large amounts of information and gain a deeper understanding of the relationships among the components of an entire system makes big data visualization so important for everyone, including U.S. Navy ship maintenance executive-level decision-makers.

Conclusions and Recommendations

Conclusions

PEO Ships asked the team from NPS to work with U.S. Navy ship maintenance metrics groups to provide additional options regarding the optimization of large datasets. Static, cumbersome spreadsheets are no longer suitable for executive-level decision-makers to make strategic choices regarding ship maintenance budgeting and scheduling. The visualization software used to present ship maintenance big data provides a means to aggregate voluminous data in visually intuitive ways to better understand cost drivers and factors that lead to schedule overruns. Big data visualization allows decision-makers to identify trends quickly, develop a better understanding of the problem space, establish defensible baselines for monitoring activities, perform forecasting, and determine the usefulness of metrics.

Visualization software provides decision-makers with a tool that makes quick identification of trends possible. Refer to Figures 32 and 33 (Definitized Estimate vs. Actual, Top 5 Ships, Type Expense and Work, respectively) or Figure 38 (Definitized Estimate vs. Actual, Type Expense, Work). In the example scenarios presented, an executive-level decision-maker was interested in identifying the largest cost contributor. Visual analysis of the figures led the decision-maker to quickly identify that subcontractor labor resulting from NW caused a trend of higher costs.

Better understanding of the problem space is also provided by the visualization of big data. Before decision-makers can make choices about the future of U.S. Navy ship maintenance, they must be able to understand the characteristics of the problem as a whole. Charts, diagrams, and solar graphs enable executives to visualize how all the data points relate to each other, to define which categories of data are of particular interest, and to forecast the impact of policy changes. Through the manipulation of big data visualization tools, decision-makers can develop a better understanding of their specific problem space.

Continued collection of ship maintenance big data would provide for the creation of defensible cost and schedule performance baselines. The sample data analyzed in this project represented a limited number of availabilities and is therefore limited in its ability to represent U.S. Navy ship maintenance as an industry. However, expanded and continued collection of ship maintenance big data would provide data that more accurately reflect the industry. If the collection of data were expanded to include all types of ships and continued to provide for the analysis of many years of data, then the visualization software could be used to create defensible cost and schedule baselines.

Executive-level decision-makers are often concerned with the future impact of their current policy change choices. Historically, executives relied upon the advice of experts and instincts developed over several years of personal experience to select which policy changes would create the effects desired. Through big data visualization software, manipulation of the data is possible to allow for forecasting. In the simulations, which examined the implementation of either 3DP technology only or the combination of multiple technologies (3DP, 3D LST, and CPLM), cost-savings trends, derived from previous research of those technologies, were applied to historical ship maintenance data. The results were then presented, as screenshots from visualization software, in a manner that allowed a decision-maker to intuitively understand the forecasted effect without the need for expensive test cases or extensive research by experts.

Metrics provide an indication of performance as long as they represent a causal relationship. LODs have long been a metric used to indicate ship maintenance performance, but their validity was questioned. Availability density was offered as an alternative metric, but proof of its validity was necessary before being considered as a real substitute for the LOD metric. Through the use of bubble charts, the visualization software created a visually intuitive display that demonstrated the correlation to expense of each of the metrics. The LOD metric was shown to be a poor indicator of cost, and the availability density metric was shown to be a good indicator of cost.

Through the use of big data visualization tools, executive-level decision-makers can identify trends quickly, develop a better understanding of the problem space, establish defensible baselines for monitoring activities, perform forecasting, and determine the usefulness of metrics.

Recommendations

Big data visualization tools are beneficial to executive-level decision-makers responsible for implementing policy throughout their enterprise. For U.S. Navy ship maintenance decision-makers considering the use of visualization software in their industry, the following recommendations are made:

- Continue collection of data. Data that reflect ship maintenance over time will provide greater value and more defensible baselines.
- Expand collection of data. Data that reflect all types of ships in the U.S. Navy would better reflect the industry and better characterize the problem space.
- Identify performance accounting software for tracking. Software packages are available that would provide for a systematic, common, and seamless method for collecting, storing, and analyzing performance data.
- Begin forecasting once performance baselines are established. Forecasting the effects of policy decisions is only as accurate, and therefore valuable, as the baselines used to derive the forecast. Continued and expanded collection of data in a common software package over a period of time must be accomplished before value can be obtained through forecasting.
- Develop a meaningful numerator for evaluating ship maintenance performance. Return on investment (ROI) is calculated by dividing the output by the input. U.S. Navy ship maintenance collects troves of data on the input, the denominator, in the form of dollars of cost. However, there is no output, or benefit, derived from ship maintenance, which is collected as a metric and represented in generic units of output. Without a numerator, the ROI of U.S. Navy ship maintenance cannot be determined.

Through the implementation of these recommendations, U.S. Navy ship maintenance executive-level decision-makers would be well on their way to deriving the benefits of big data visualization. Those benefits include the ability to identify trends quickly, develop a better understanding of the problem space, establish defensible baselines for monitoring activities, perform forecasting, and determine the usefulness of metrics.

References

Avdellas, N., Berry, J., Disano, M., Oaks, D., & Wingrove, W. (2011). *Future capability of DOD maintenance depots* (Report LG901M2). Retrieved from http://armedservices.house.gov/index.cfm/files/serve?File_id=5ea27615-e4fe-41ad-8884-d205988dd772

Choy, J., Chawla, V., & Whitman, L. (2012). *Data visualization techniques: From basics to big data with SAS visual analytics* (White paper). Cary, NC: SAS. Retrieved from http://www.sas.com/offices/NA/canada/downloads/IT-World2013/Data-Visualization-Techniques.pdf

Conway, S. (2012). Big data cloud delivers military intelligence to U.S. Army in Afghanistan [Press release]. Retrieved from http://www.datanami.com/datanami/ 2012-02-06/big_data_cloud_delivers_military_intelligence_to_u.s._army_in_afghanistan.html

Department of Homeland Security (DHS). (2013). Big data storage and analytics technology (Solicitation No. HSHQDC-13-BIG_DATA_RFI). Retrieved from https://www.fbo.gov/index?s=opportunity&mode=form&id=ca5d230b7454615562a026466264acc2&tab=core&_cview=1

Department of the Navy (DoN). (2012a). *Highlights of the DON FY2013 budget*. Retrieved from http://www.finance.hq.navy.mil/fmb/13pres/OMN_Vol1_book.pdf

Department of the Navy (DoN). (2012b). *Joint fleet maintenance manual*. Retrieved from http://www.submepp.navy.mil/

Dijcks, J. P. (2013). *Oracle: Big data for the enterprise* (White paper). Redwood Shores, CA: Oracle. Retrieved from http://www.oracle.com/us/products/database/big-data-for-enterprise-519135.pdf

Donaldson, I. (2013). *Big data implications for enterprise architecture*. Unpublished manuscript, Department of Information Sciences, Naval Postgraduate School, Monterey, California.

Ferguson, M. (2013). *Enterprise information protection: The impact of big data* (White paper). Cheshire, England: Intelligent Business Strategies. Retrieved from http://www-01.ibm.com/software/os/systemz/pdf/Info_Security_and_Big_Data_on_Z_White_Paper_Final.pdf

Gantz, J., & Reinsel, D. (2012). *The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East*. Framingham, MA: IDC Corporation.

Housel, T. J., & Bell, A. (2001). *Measuring and managing knowledge*. Boston, MA: McGraw-Hill.

Kelly, J., Floyer, D., Vellante, D., & Miniman, S. (2013). *Big data vendor market revenue and forecast*. Retrieved from http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2012-2017

Kenney, M. E. (2013). *Cost reduction through the use of additive manufacturing (3D printing) and collaborative product life cycle management technologies to enhance the Navy's maintenance programs* (Master's thesis, Naval Postgraduate School). Retrieved from http://acquisitionresearch.net

King, R. (2013). U.S. government spending on big data to grow exponentially. Retrieved from http://www.biometricupdate.com/201308/u-s-government-spending-on-big-data-to-grow-exponentially

Knowledge Value-Added (KVA) Methodology. (n.d.). International Engineering Consortium. Retrieved from http://cmapspublic3.ihmc.us/rid=1G9L62WTW-NJ5G5D-C84/KVAmethodology.pdf

Komoroski, C. L. (2005). *Reducing cycle time and increasing value through the application of knowledge value added methodology to the U.S. Navy shipyard planning process* (Master's thesis, Naval Postgraduate School). Retrieved from http://acquisitionresearch.net/

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute. Retrieved from http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Office of Science and Technology Policy. (2012). Obama administration unveils big data initiative: Announces $200 million in new R&D investments [Press release]. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf

Office of the Assistant Secretary of Defense for Logistics and Material Readiness (OASD[L&MR]). (2011). *DOD maintenance fact book*. Washington, DC: Author. Retrieved from http://www.acq.osd.mil/log/mpp/factbooks/2011_Fact_Book_ final.pdf

TechAmerica Foundation. (2012). *Demystifying big data: A practical guide to transforming the business of government*. Washington, DC: Author. Retrieved from http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf

Sverdlik, Y. (2012). *Big data gives real time logistics to U.S. Navy: A data analytics platform changes the way Navy thinks about logistics* [Press release]. Retrieved from http://www.datacenterdynamics.com/focus/archive/2012/02/big-data-gives-real-time-logistics-us-navy

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix: Big Data Implications for Enterprise Architecture

(Donaldson, 2013)

Introduction

By April, 2011, the United States (U.S.) Library of Congress had collected 235 terabytes of data. However, of the 17 U.S. business sectors, 15 of them had more data stored than the Library of Congress, per company [McKinsey Global Institute (MGI), 2011]. MGI's report on big data (2011) estimated that a 60% increase in the operating margins of retailers would be possible if big data collection, storage, and analysis techniques were properly utilized. So, how does an enterprise derive value from big data?

Big data spawns from many sources and possesses characteristics which are pertinent to the practitioner of enterprise architecture (EA). The needs of the enterprise and how the data is to be processed determines how an EA should be designed to ensure the enterprise can derive value from big data. The impact of big data results from the volume, variety, velocity, and value traits of the data and influences both the network and capacity considerations of the EA. However, obstacles to implementation exist and are either technical or human, each of which requires a different approach. Should an architect carefully consider, plan, and implement an EA designed to accommodate big data, an enterprise could derive value that affects the bottom line.

Big Data

Big data is generated by a variety of sources. The sources from which big data originate include industry specific transactions, machine/sensor indications, web applications, and text (Ferguson, 2013). Industry specific transactions can include call records and geographic location data. Machines generate extremely large volumes of information everyday and can range in complexity from simple temperature readings to the performance parameters of a gas-turbine engine. Big data on the web also ranges in format from machine language to customer comments on social networks and also is produced in considerably sizeable portions. Text sources can include archived documents, external reports, or customer account information (Ferguson, 2013).

Because big data comes from a variety of sources, it also possesses characteristics which distinguish it from data in the traditional context. Common terms used to define the qualities of big data include volume, variety, velocity, and value (Dijcks, 2013). From the listing of sources above, one can understand that the *volume* of data generated on a daily basis is enormous. For example, Dijcks (2013) stated that just a single jet engine produces 10 terabytes of data in 30 minutes. Extrapolate that example to include all the aircraft currently airborne, then include all the factory infrastructure around the globe collecting data on production, service life, and maintenance requirements, and the enormity of big data volumes begins to emerge. Another characteristic of big data, *variety*, can be directly translated from the various sources into the variety of data formats. In the context of EA, various data formats requires additional consideration to ensure the ability of all systems to share data. *Velocity*, which is related to volume, is the frequency with which big data is created. To illustrate velocity, consider the relative size of a single Twitter feed (140 characters) to the large number of feeds generated in a given time period (Dijcks, 2013). Finally, *value* is the feature of big data which is important to the enterprise.

Big Data is Valuable to the Enterprise

Big data can provide value to an enterprise through various means. Processing and then analyzing big data can help an enterprise better understand its business, the environment in which it operates, and its customers (Dijcks, 2013). Having developed an enhanced perception of itself and the marketplace, an enterprise could stand poised to improve productivity, increase competitive advantage, or develop superior product innovation processes (Dijcks, 2013). All these benefits can translate into significant impacts on the bottom line. The benefits which can be derived from big data are unique to the specific enterprise and, therefore, the manner in which EA design is approached is also unique. However, in general, the proper collection, storage, and analysis of big data is instrumental in the ability of the enterprise to reap value from it.

Impact of Big Data on EA

An EA must be designed properly to provide the capability to an enterprise to derive value from big data. The characteristics of big data - volume, variety, velocity, and value – must all be considered and planned for during the design or update of an EA if the enterprise wishes to use big data to generate value. Bakshi (2012) breaks down the EA considerations into two major groups, network and capacity.

Network considerations include data paths, scalability, buffering, and latency (Bakshi, 2012). Regarding the *data paths*, redundancy provides strength to an EA designed for big data. Data is often located in multiple locations on an enterprise's network. Providing multiple paths among the data locations improves the EA's ability to share data. Should data collection, storage, and processing needs increase, designing an EA to be *scalable* would allow an enterprise the capability to expand or contract as necessary. Considering the volumes with which data will be transmitted, an EA with sufficient *buffers* and *queues* would be beneficial. Without those buffers, a network may become overloaded with data and slow down or even crash. The final point Bakshi (2012) made regarding network considerations was that consistent and predictably low latency must be a trait of an EA designed to handle big data.

Capacity considerations involve dispersed computing and data locations, distributions, and volumes (Bakshi, 2012). The last three aspects are actually all symptoms of the well planned, dispersed computing EA. The main idea of *dispersed computing* is to spread out the data amongst the nodes within the enterprise, possibly within a separate big data warehouse, but more likely throughout the enterprise. With the data *distributed* throughout the EA, the processing power requirements for big data analysis can be shared across the network. Each *location* where data is stored must be able to dependably and reliably collect, store, and analyze *volumes* of information. Therefore, high speed, low latency connections are key throughout the enterprise (Bakshi, 2012). Knowing the implications of big data upon an EA are one thing. Integrating big data into EA is another. Many obstacles exist which ensure the implementation of big data-minded EA is a challenge.

Major Obstacles to Proper Implementation of Big Data into EA

The shift from traditional EA to an EA which is designed for big data is both a technological challenge and a human challenge (M2 Presswire, 2012). The technology aspect involves an architect selecting and introducing IT into an EA which may not have been originally designed to accommodate change or expansion. Obstacles may include incompatible technologies, big data tools which do not address the particular needs of the enterprise, and hidden technology gaps. Careful consideration, planning, and implementation of the data and application architectures into the existing EA is necessary to remedy the existing (and avoid the creation of new) dysfunctionalities and/or technology gaps (M2 Presswire, 2012). Should the technology aspect of EA redesign be executed smoothly and successfully, the human facet must still be addressed.

The stakeholders, both those who finance the EA project and those who are the end-users, represent the human aspect. Among the decision makers, there may exist a lack of awareness regarding the capabilities and risks associated with embarking upon an EA project (M2 Presswire, 2012). There also may exist a resistance by the end-users to change systems already in place. Both of these obstacles can be overcome through gaining stakeholder buy-in. Through education and the inclusion in planning, both decision makers and end-users can be persuaded to support big data changes in EA (M2 Presswire, 2012). One other possible obstacle will be that it might be necessary to change technology interfaces or processes to facilitate the integration of big data into business units. However, skill gap analysis can identify where disconnects between humans and technology exist and training provides the bridge to cross those gaps.

Conclusion

Big data exhibits characteristics which require special consideration when designing an EA. The volume, variety, velocity, and value of big data must be understood by the EA practitioner before embarking upon a project so complex and risky. When compared to traditional methods of designing EA, big data requires networks have redundant data paths, offer scalability, provide sufficient buffering, and exhibit consistent, reliably low latency. As for capacity considerations, successfully implemented dispersed computing environments deliver the necessary data locations, distribution, and volume handling capability required by big data collection, storage, and analysis. When technical or human obstacles arise, an architect which carefully plans, conducts gap analysis, acquires stakeholder buy-in, and provides necessary training will overcome those hurdles. Should an architect follow these guidelines, an EA capable of handling big data could produce improved productivity, increased competitive advantage, and superior product innovation processes for its enterprise.

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net