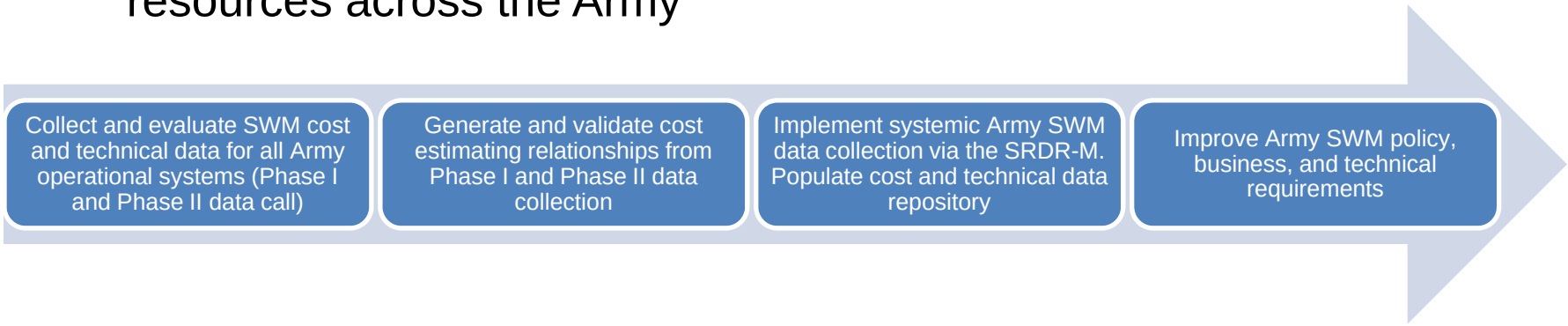
Deriving Software Sustainment Cost Estimating Relationships in a Diverse Army Execution Environment

PSM Users' Group Workshop
14 June 2017

SWM Initiative Objective and Strategy

Accurately estimate Army system software maintenance costs to:

- Effectively project and justify software and system life cycle costs
- Objectively evaluate Army system software maintenance execution costs
- Inform and optimize the allocation of available maintenance resources across the Army



Effective software maintenance cost estimation is the basis for Army system software life cycle cost management

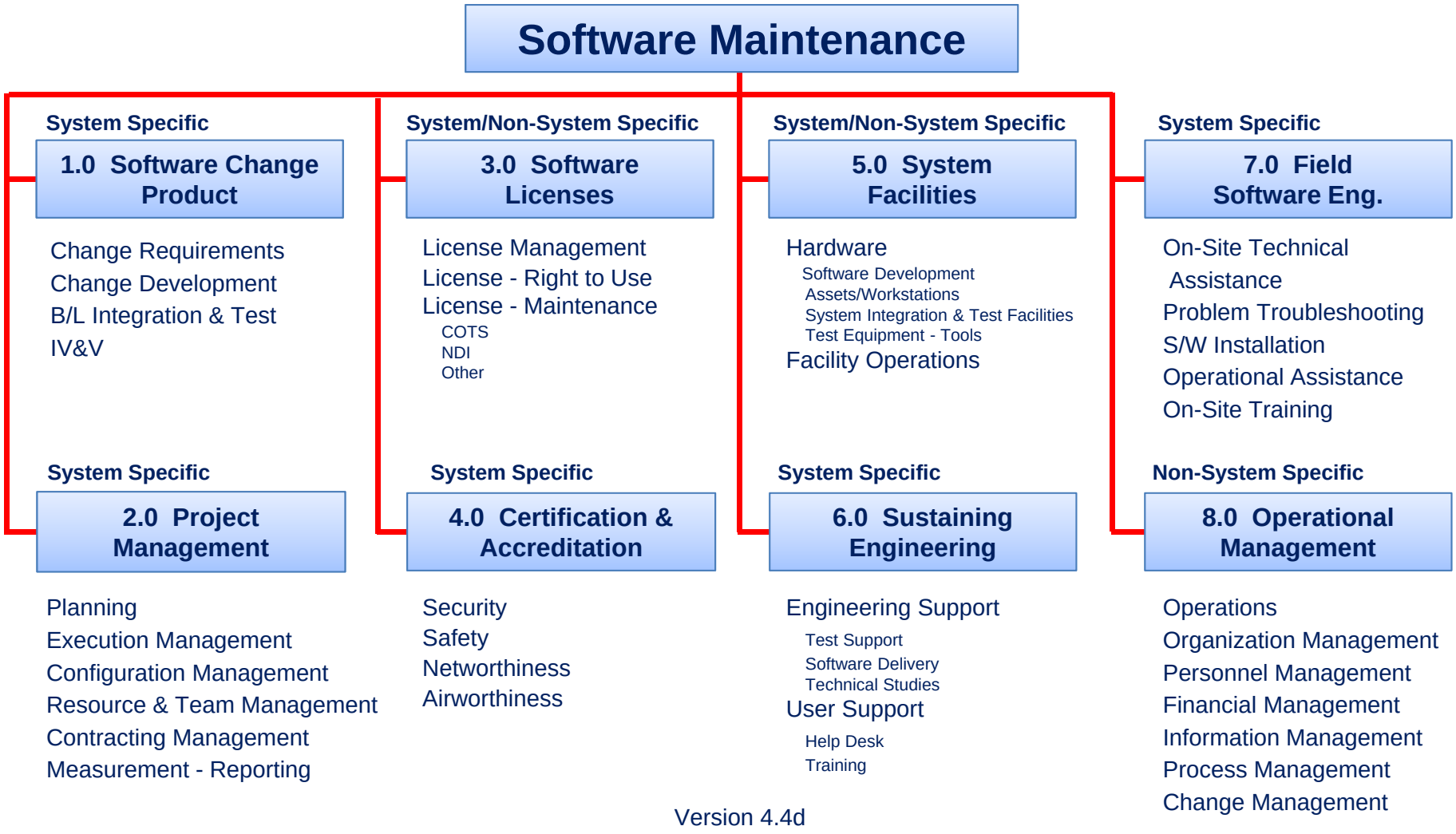
Phase I
Data Collection and Evaluation

Army Software Maintenance Definition

For this effort, software maintenance is defined as:

- Software maintenance includes all software change activities and products associated with modifying a software system after EMD has completed and a software release has been provided to an external party
- The release is the primary SWM change product - a composite of one or more changes - it can be either a formal release or an engineering release
- SWM includes software enhancements and software corrections/adaptations
- SWM includes activities and change products funded by multiple funding sources
- Fixed and Variable costs accrued at both the system and organizational levels by both organic and contractor resources
- Software maintenance and software sustainment are considered to be synonymous

Army Software Maintenance WBS



Version 4.4d

Phase I Data Requirements

System Context

- System description
- Organizations involved
- Maintenance process
- CMMI rating
- Number of software baselines
- Number of hardware platforms/number of users
- Analogous systems

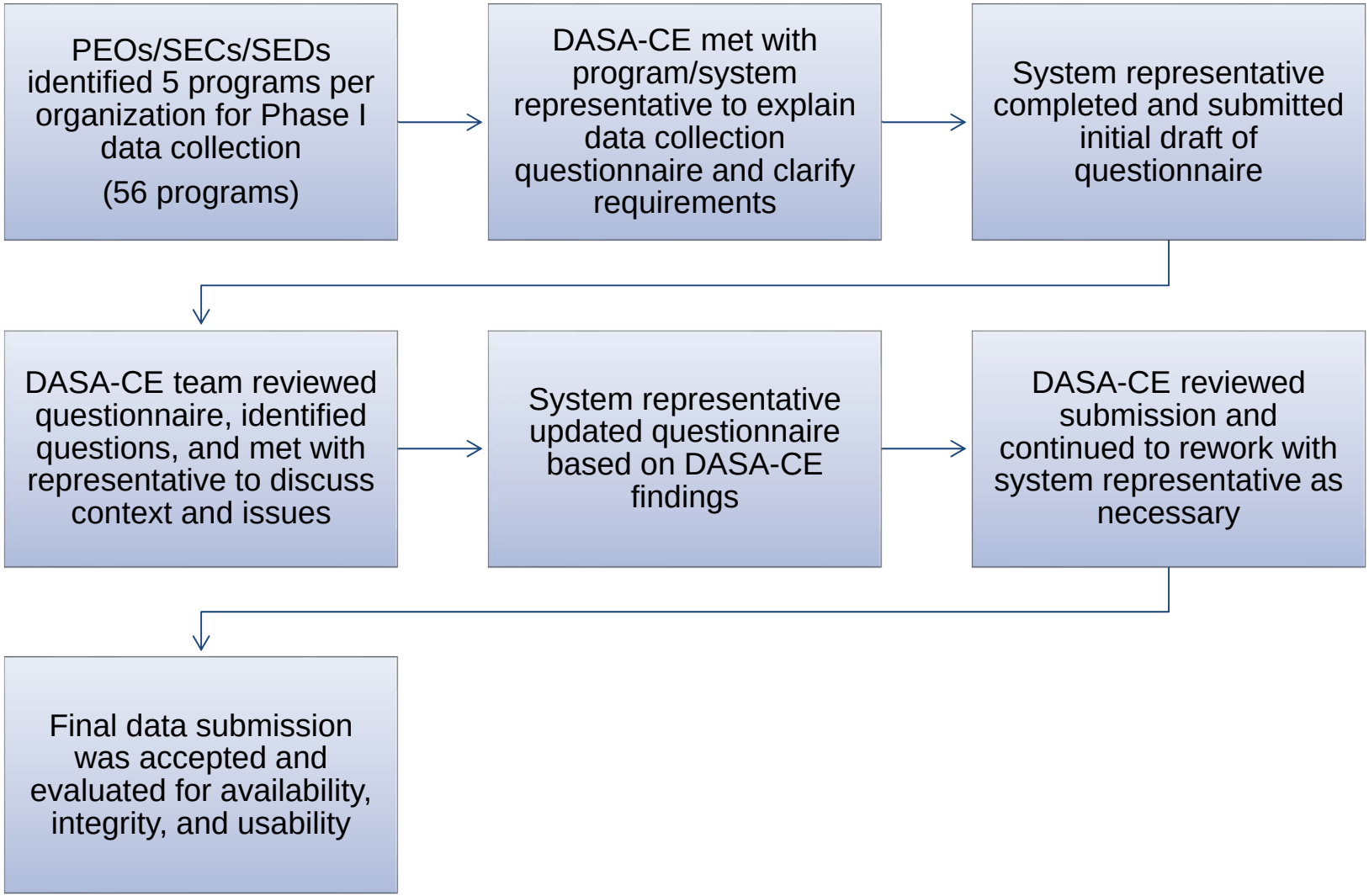
Program Level

- Annual effort/cost data (total annual plus WBS elements #2 through #8) broken out by government and contractor (3 years of data)
- Labor rates
- Hourly basis for FTEs
- Software licenses

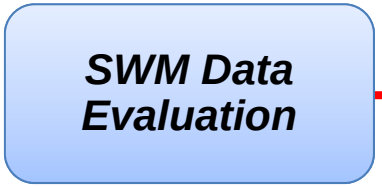
Release Level

- Release context information
- Application domain
- Operating environment
- Release effort / cost
- Schedule - start and end dates
- Size data (those that apply)
  - Software requirements
  - External interfaces
  - Source Lines of Code (SLOC)
  - Non-SLOC based size (e.g. RICE-FW, use cases, story points)
  - Software changes counts by priority (e.g. change requests, problem reports, defects)
  - IAVAs

Phase I Data Collection Process



Software Maintenance Data Evaluation



Availability

- Completeness of required data set
- Underlying SWM business and technical processes are well enough defined to produce objective data on a periodic and/or event driven basis
- IT systems and tools exist to enable systematic and timely data collection

Integrity

- Data are derivatives of actual SWM technical and management processes
- All data (measures) are explicitly defined - measurement contexts are known
- Cost data is directly correlated with the WBS defined output products and activities
- Data is consistent - methods exist to address system conflicts (normalization)

Usability

- Data is aligned with stakeholder decision information needs
- Data can be objectively characterized and interpreted
- Mapping and aggregation structures and methods exist to combine data
- Potential emerging information requirements have been considered

Detailed Data Evaluation

				Initial System Overall				Detailed System Assessment							
Collect Phase	PEO	SEC	System	Definable Maint. Process	Total Program Effort/Cost	WBS 2-8	Project Mgmt (WBS-2)	License Management (WBS-3)	C&A Support (WBS-4)	System Facilities Management (WBS-5)	Sustaining Engineering (WBS-6)	Field S/W Engineering (WBS-7)	Operational Management (WBS-8)	License Costs
1	(PEO 4)	SEC 3	System 1	G	G	G	G	N/A	G	G	G	N/A	R	N/A
1	(PEO 4)	SEC 3	System 2	G	G	Y	Y	R	Y	Y	Y	N/A	R	G
1	(PEO 1)	SEC 2	System 3	R	G	Y	R	G	G	Y	Y	N/A	Y	R
1	PEO 1	(SEC 2)	System 4	G	R	O	R	R	R	R	R	N/A	R	R
1	PEO 1	(SEC 2)	System 5	G	G	Y	R	R	G	Y	G	G	G	R
1	PEO 1	(SEC 2)	System 6	G	G	O	R	R	Y	R	R	Y	R	G
1	PEO 1	(SEC 2)	System 7	G	G	G	G	G	G	G	G	G	G	G
1	PEO 1	(SEC 2)	System 8	G	R	R	R	R	R	R	R	R	R	R

				Initial Release Overall		Detailed Release Assessment							
PEO	SEC	System	Release	CER Usability	SER Usability	Size: Requirements	Size: External Interfaces	Size: SLOC	Size: Non-SLOC	Size: SW Changes	IAVAs	Effort (WBS-1)	Schedule (WBS-1&2)
(PEO 4)	SEC 3	System 1	Release 1	Y	Y	G	G	G	N/A	G	G	Y	G
(PEO 4)	SEC 3	System 1	Release 2	Y	Y	G	G	G	N/A	G	G	Y	G
(PEO 4)	SEC 3	System 2	Release 1	G	G	R	R	Y	N/A	G	G	G	G
(PEO 4)	SEC 3	System 3	Release 1	G	G	G	N/A	G	N/A	G	N/A	G	R
(PEO 4)	SEC 3	System 4	Release 1	G	R	R	N/A	G	N/A	G	N/A	G	R
(PEO 4)	SEC 3	System 4	Release 2	G	R	R	N/A	G	N/A	G	N/A	G	R
(PEO 1)	SEC 2	System 5	Release 1	Y	R	R	R	G	R	G	N/A	Y	R
PEO 1	(SEC 2)	System 5	Release 2	G	G	R	R	G	N/A	R	R	G	G
PEO 1	(SEC 2)	System 5	Release 3	G	G	R	R	G	N/A	R	R	G	G
PEO 1	(SEC 2)	System 5	Release 4	G	G	R	R	G	N/A	R	R	G	G
PEO 1	(SEC 2)	System 5	Release 5	G	G	R	R	G	N/A	R	R	G	G
PEO 1	(SEC 2)	System 5	Release 6	G	G	R	R	G	N/A	R	R	G	G
PEO 1	(SEC 2)	System 6	Release 1	R	R	G	G	Y	N/A	G	R	R	G
PEO 1	(SEC 2)	System 6	Release 2	R	R	G	G	Y	N/A	G	R	R	G
PEO 1	(SEC 2)	System 6	Release 3	R	R	G	G	Y	N/A	Y	R	R	G
PEO 1	(SEC 2)	System 7	Release 1	R	R	G	Y	G	N/A	G	R	O	R
PEO 1	(SEC 2)	System 7	Release 2	R	R	G	Y	G	N/A	G	R	O	R
PEO 1	(SEC 2)	System 8	Release 1	G	G	G	G	G	N/A	G	G	G	G
PEO 1	(SEC 2)	System 9	Release 1	R	R	Y	G	G	N/A	Y	N/A	R	R
PEO 1	(SEC 2)	System 9	Release 2	R	R	Y	G	R	N/A	N/A	N/A	R	G
PEO 1	(SEC 2)	System 9	Release 3	R	R	Y	G	G	N/A	R	N/A	R	G

Data Evaluation Phase I - Summary

	Initial System Overall				Detailed System Assessment							
	Definable Maint. Process	Total Program Effort/Cost	WBS 2-8	Project Mgmt (WBS-2)	License Mgmt (WBS-3)	C&A Support (WBS-4)	System Facilities Mgmt (WBS-5)	Sustaining Engineering (WBS-6)	Field S/W Engineering (WBS-7)	Operational Mgmt (WBS-8)	License Costs
R	14	11	11	28	30	11	21	27	17	32	12
O	0	2	15	2	2	8	6	3	2	1	3
Y	1	12	16	8	4	14	11	9	8	10	1
G	40	31	13	18	12	20	16	16	6	11	35
N/A	1	0	1	0	8	3	2	1	23	2	5
Total	56	56	56	56	56	56	56	56	56	56	56

- Formal data evaluation process was used to rate the data
  - Data was collected from 56 programs*
    - 43 programs provided total system SWM costs (G, Y)
  - Rating criteria is shown below:

Table 1. Data Quality Levels

Color	Definition	Value
R	Red indicates there is no planning or actual data reported.	0
O	Orange indicates only planning data was reported.	1
Y	Yellow indicates FTE or partial, actual data was reported	2
G	Green indicates that actual data was reported.	3

*Detailed breakout of data evaluation by data point provided in backup

Data Evaluation - WBS 1.0 Release Data

Initial Release Overall			Detailed Release Assessment							
	CER Usability	SER Usability	Size: Requirements	Size: External Interfaces	Size: SLOC	Size: Non-SLOC	Size: SW Changes	IAVAs	Effort (WBS-1)	Schedule (WBS-1)
R	71	77	101	79	46	28	39	67	60	40
O	44	43	5	2	5	0	3	3	47	12
Y	23	22	3	2	4	2	5	0	23	27
G	79	75	70	55	69	12	116	106	87	138
N/A	1	1	39	80	94	176	55	42	1	1
Total	218	218	218	218	218	218	218	218	218	218

- Data was collected from 218 releases
  - 146 releases had sufficient data to use in CER cost calculations (G,Y,O)
  - Size data was not always consistently tracked and generally was not mapped to resource (effort/cost/schedule) information
    - 124 releases tracked some sort of software change counts (defects, PTRs)
    - 109 releases tracked IAVAs
    - Systems in different super-domains used different size measures
      - Many weapon systems tracked SLOC data

*Detailed breakout of data evaluation by data point provided in backup

Phase I SWM Data Analysis

Analysis Background

- The analysis covers Phase I data only
  - Phase II data will result in updated CERs and data demographics
- Estimating approach is specific to the SWM WBS. For any relationship identified, the WBS coverage should be noted
- Given the data sample size, the super domain classification is used to group similar data points
- All data points and associated classification are listed in the backup
- Utilized data represents both post deployment software support (PDSS) as well as post production software support (PPSS)
- Utilized data was from a variety of appropriations (see normalization for how this was handled)
- All costs shown are in BY 2016 $
- For regression analysis, the following fit statistics were utilized:
  - R^2
  - P-value/T-stat/F-stat
  - Standard Error of the Estimate
  - Pred (30)

Limitations

- Data is not from a formal deliverable from a performing organization or vendor. It was provided by programs via the DASA-CE SWM questionnaire.
- Programs have not historically tracked SWM execution costs according to the DASA-CE SWM WBS. Data was often provided at an aggregate level or broken out using SME judgement.
- Due to the nature of the data collection, it is assumed that reported costs are more accurate than reported effort (hours). Future analysis will also utilize effort data.
- It is assumed the super domain is a meaningful way to aggregate data points.
- Given the data sample size, all data points were used for analysis*

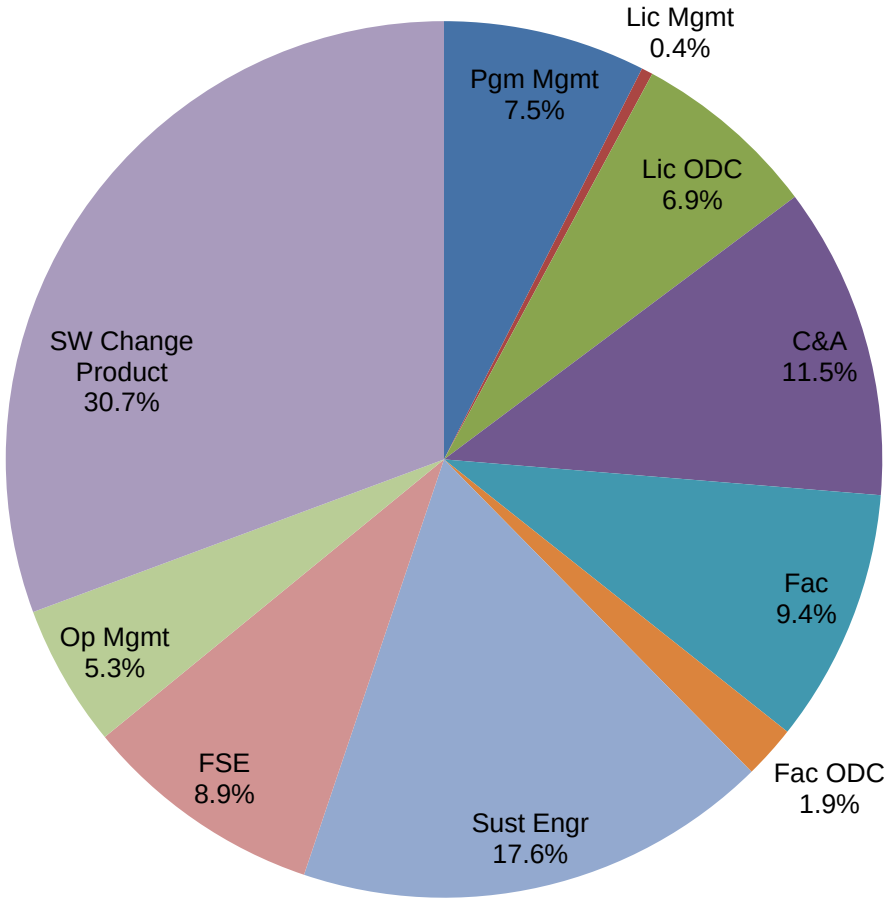
*In a few cases outliers were removed, these instances are noted within the analysis

Super Domain Definitions

Real-Time	Engineering	Support	AIS
Real-Time is the most constrained type of software. These are specific solutions limited by system characteristics such as memory size, performance, or battery life. These projects take the most time and effort due to constraints.	Engineering software operates under less severe constraints than real-time software. This software may take real-time software outputs and further process them to provide human consumable information or automated control of devices. Or the software may perform transformation and aggregation / distribution of data.	Support software assists with operator training and software testing. This software has few constraints.	Automated information system software provides information processing services to humans or software applications. These applications allow the designated authority to exercise control and have access to typical business / intelligence processes and other types of information access. These systems also includes software that facilitates the interface and control among multiple COTS / GOTS software applications.
Application Domains	**Application Domains**	**Application Domains**	**Application Domains**
Microcode & Firmware Signal Processing Vehicle Control/Vehicle Payload Other Real-Time Embedded Command & Control Communications	System Process Control Scientific and Simulation Test, Measurement, Diagnostic and Evaluation	Training Software Tools	Mission Planning Custom AIS Software Enterprise Service Systems Enterprise Information Systems
Examples	**Examples**	**Examples**	**Examples**
Field Programmable Gate Arrays, Flight Control, Missile Control, Radar Altimeter, Network Operations, Signal Electronics, Tracking Sensors, Encryption, Radio Networks, Propulsion	Operating Systems, Image processing, Simulation & Modeling, Test Equipment, File Management, Artificial Intelligence, Manufacturing Process Control	Computer Based Training, Compilers, Programming Aids, Code Generators, Assemblers, Courseware, Test case generation, Linker/loaders, Code Auditors	Scenario Generators, Target Planning, Enterprise Service Management, Enterprise Resource Planning, Transaction Processing, Data Warehousing, Financial Transactions

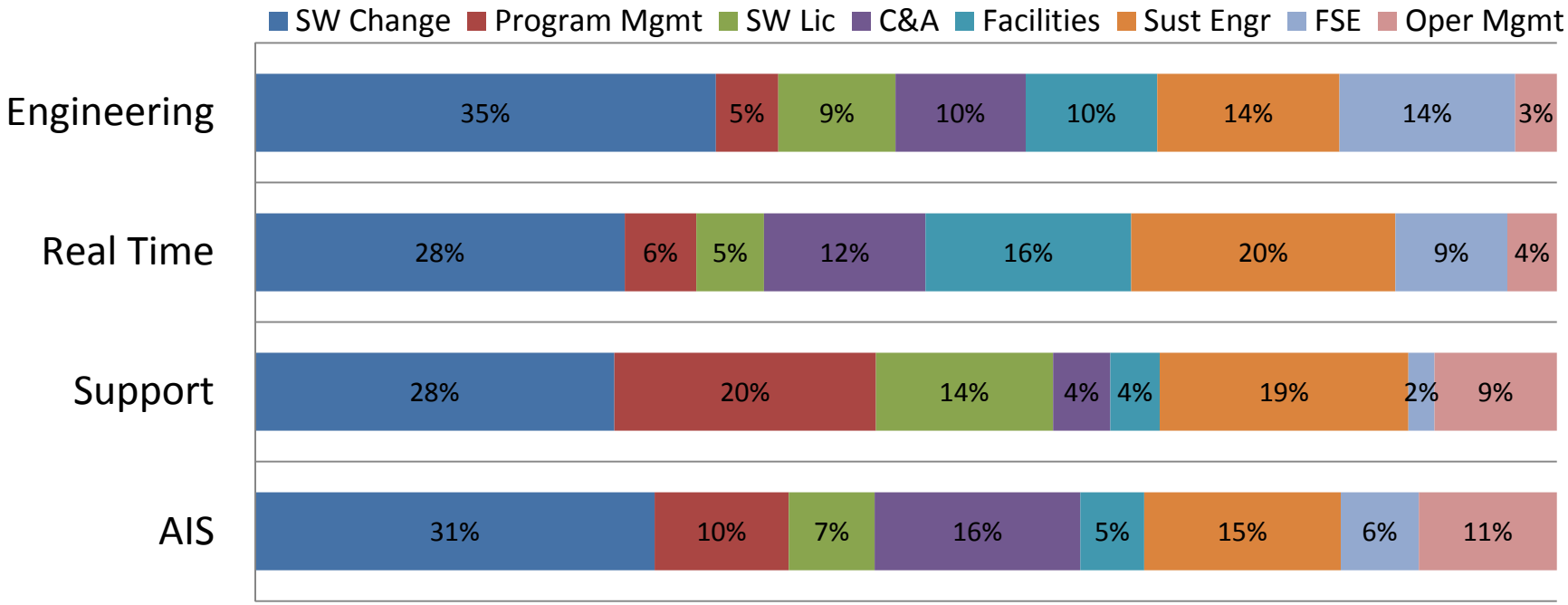
Cost Allocation Across the SWM WBS

All Phase I Systems



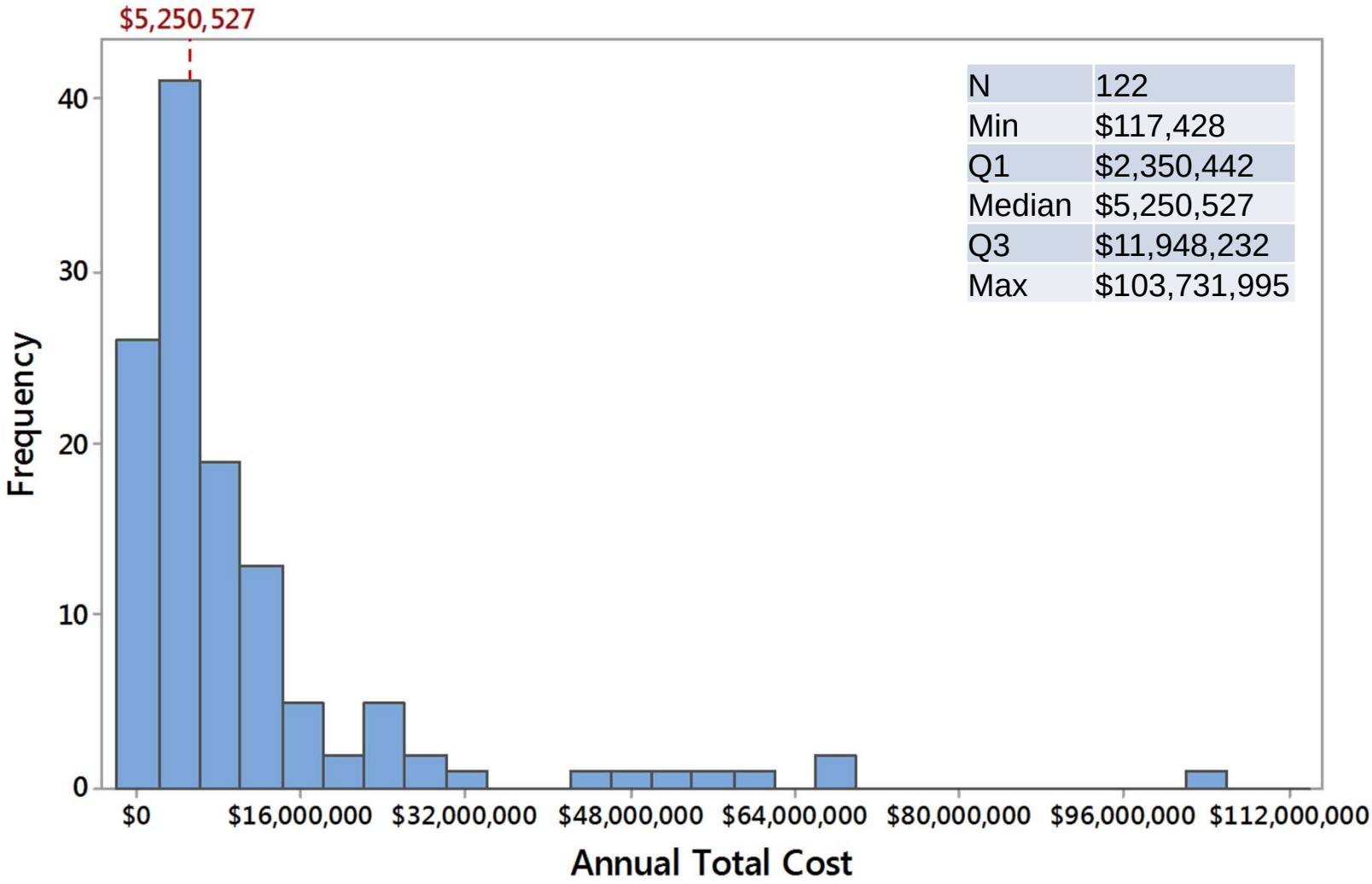
Sample Size: 43 Systems
113 Data Points

Cost Allocation by Super Domain



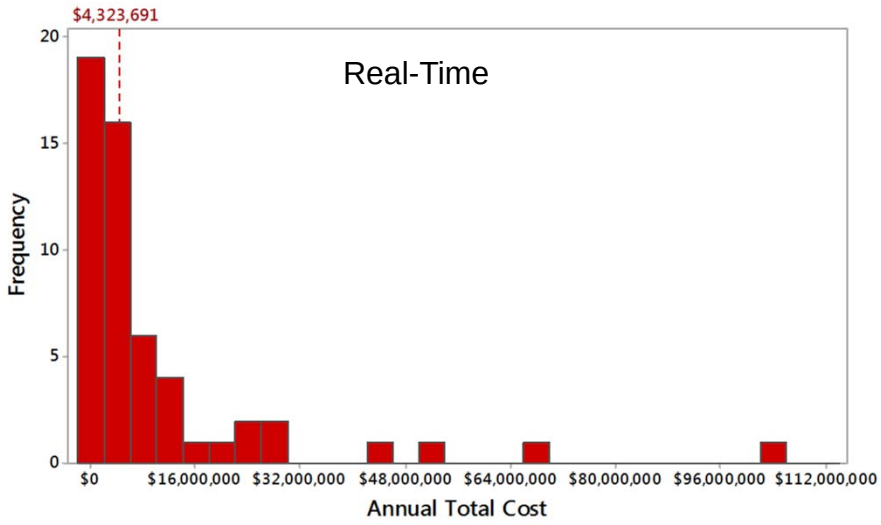
Distribution of Annual Cost

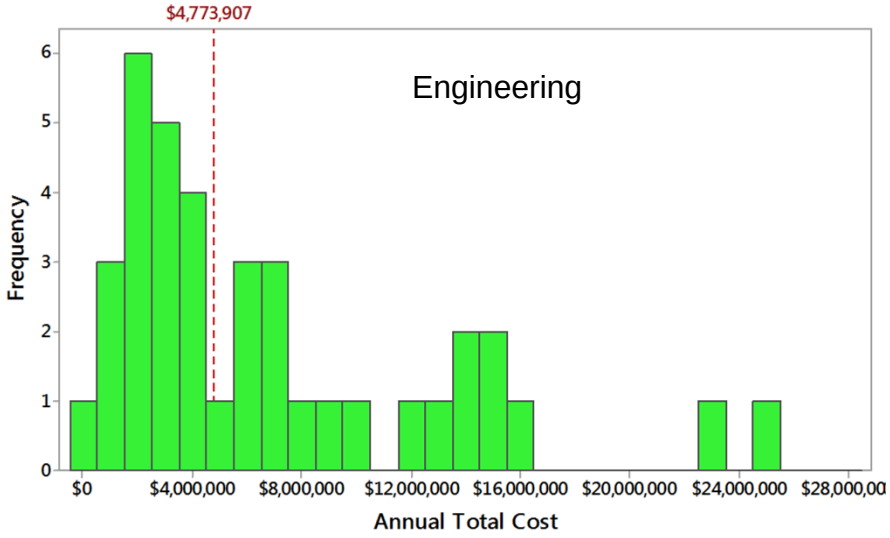
System Annual Cost (WBS 1.0 - 8.0)

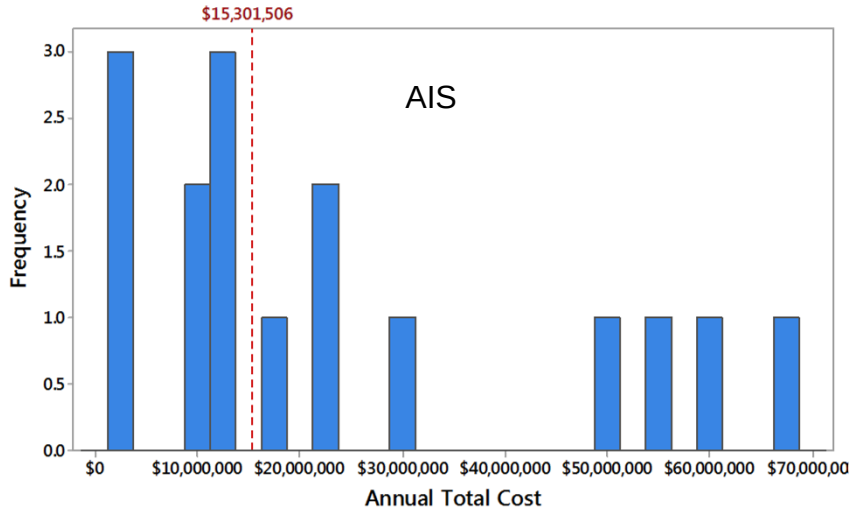


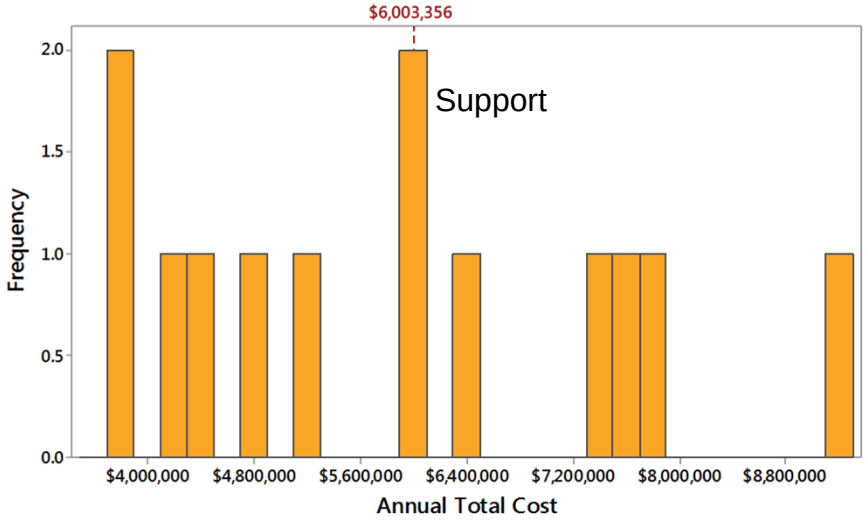
N	122
Min	$117,428
Q1	$2,350,442
Median	$5,250,527
Q3	$11,948,232
Max	$103,731,995

System Annual Cost by Super Domain







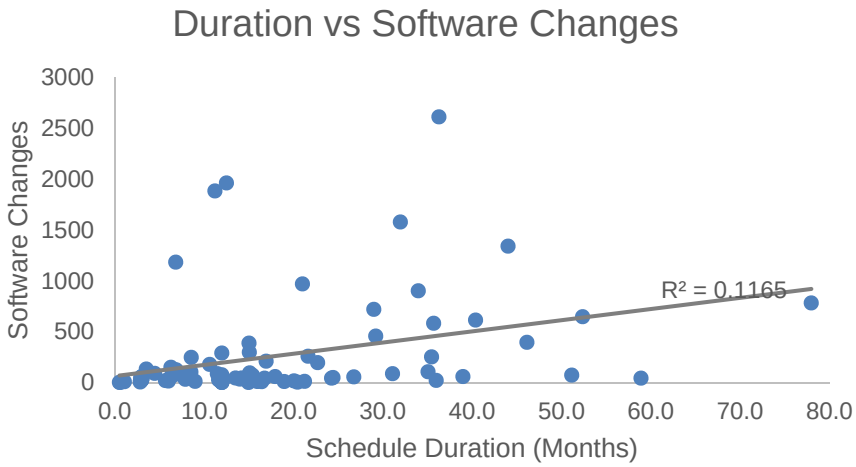


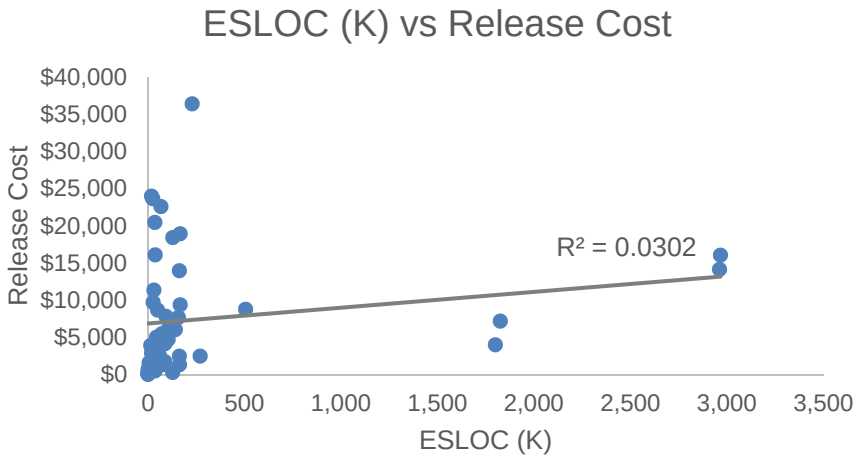
System Annual Cost Summary

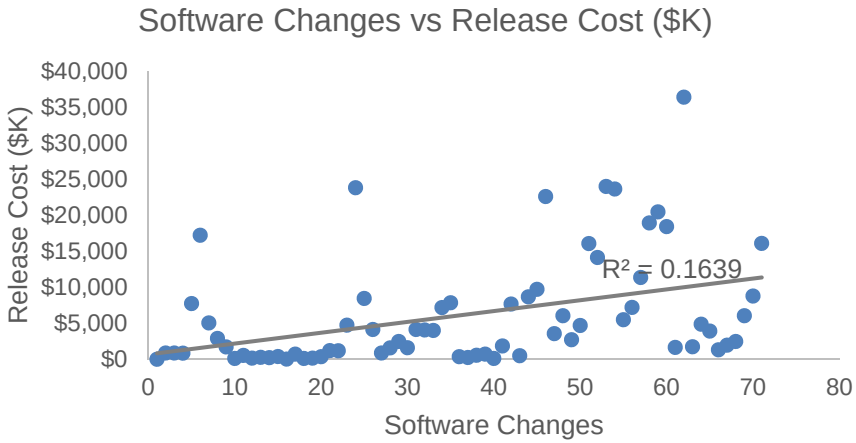
SD	N	Minimum	Q1	Median	Q3	Maximum
AIS	16	$2,783,335	$9,727,349	$15,301,506	$45,125,746	$66,448,489
ENG	38	$250,732	$2,320,639	$4,773,907	$10,734,471	$24,870,059
RT	55	$117,428	$1,363,244	$4,323,691	$10,355,772	$103,731,995
SUP	13	$3,729,674	$4,363,762	$6,003,356	$7,467,262	$9,120,451

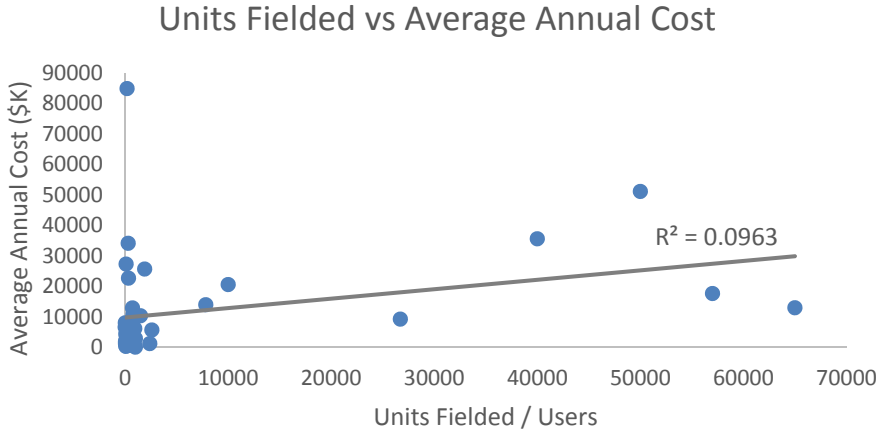
Cost Estimating Relationships

Exploratory Data Analysis









Scatter plots at the top level show significant variance. Phase II should reduce variance and allow analysis on meaningful data subsets.

Initial Phase I CERs

Dependent Variable	Equation	Super Domain	R^2	Sample Size	PRED(30)	SEE
Duration	1.355 * Req^0.3323	AIS	60.04%	16	12.50%	4.7
Total Rel Cost	6,981 * New_Mod^0.4004 * Dur^0.755	All	65.10%	43	27.91%	7,260,456
Total Rel Cost	1,955 * Dur^.6423 * New_Mod ^.5382 * 1.796^RTDummy	All - Outliers Removed*	81.10%	39	25.64%	5,941,321
Total Rel Cost	2,878 * Dur^.8052 * New_Mod ^.4938	All - Outliers Removed*	79.70%	39	35.90%	6,032,352
Ctr Hours	24.49 * New_Mod ^.624	All Non-IAVA	75.15%	38	26.53%	51,539
Total Hours	43.35 * New_Mod ^.5932	All Non-IAVA	71.75%	47	19.12%	180,076
Total Hours	34.67 * New_Mod^0.5911	ENG	76.47%	23	21.74%	44,340
Total Rel Cost	22,159 * New_Mod^0.4362	ENG	73.00%	14	21.43%	3,506,848
Total Rel Cost	28,941 * ESLOC^0.413	ENG	72.80%	14	21.43%	3,093,766
Ctr Hours	29.58 * New_Mod^0.5851	ENG	72.34%	20	15.00%	37,164
Cost per Month	65,626 + 10.82*New_Mod	RT	79.63%	23	34.78%	174,130
Total Rel Cost	4,775 * New_Mod^0.4554 * Dur^0.764	RT	72.00%	27	22.22%	7,332,110
Total Rel Cost	2,697 * ESLOC^0.3728 * Dur^1.058	RT	68.10%	28	28.57%	7,495,672
Total Hours	939.51 * SC^0.5177	SUP	89.91%	13	61.54%	5,309
Ctr Hours	794.69 * SC^0.516	SUP	88.59%	13	69.23%	5,126
Total Rel Cost	47,858 * SC^0.3267 * Dur^0.516	SUP	75.50%	13	46.15%	242,287
Total Rel Cost	123,588 * SC^0.3847	SUP	64.90%	14	28.57%	393,099

Explanation of Variables:

ESLOC = Equivalent Source Lines of Code

New_Mod = Sum of New and Modified Lines of Code

SC = Software Change Count (Problem Reports, Defects, Issues, Change Requests, etc.)

Dur = Release Duration in months

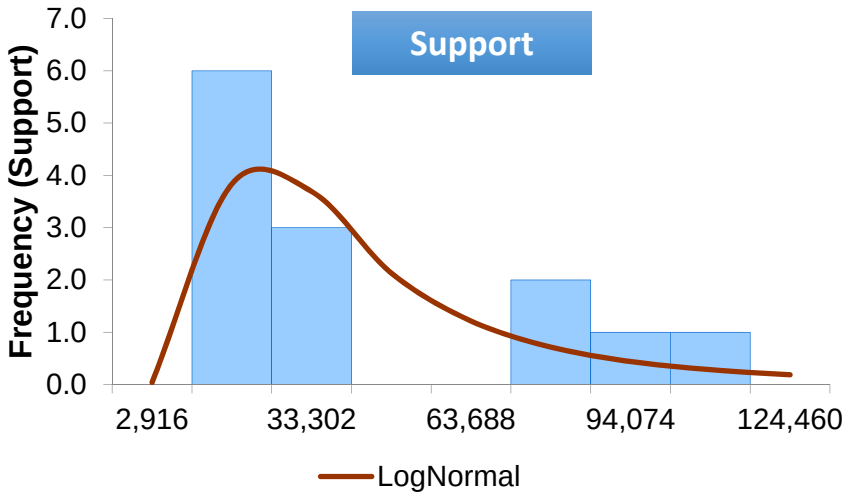
Req = Software Requirements (SRS equivalent requirements)

*All CERs shown have a p-value < .005

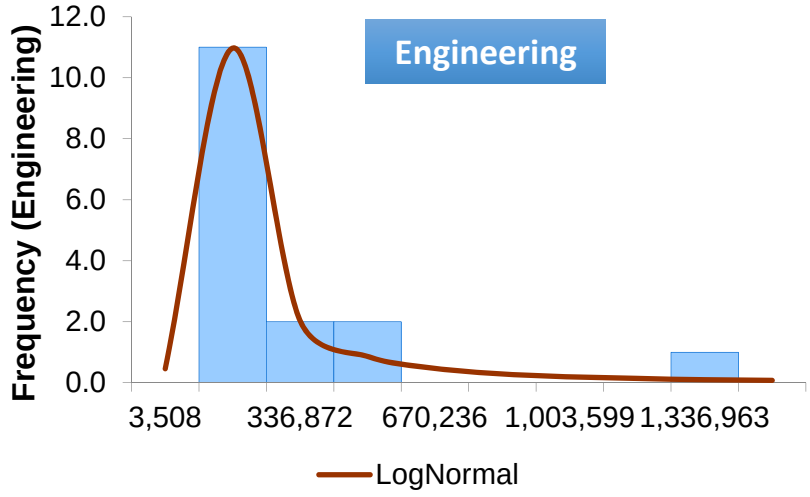
*Min/Max values for each coefficient are shown in backup

Cost Benchmarks

Cost Per Software Change
Support and Engineering Super Domain



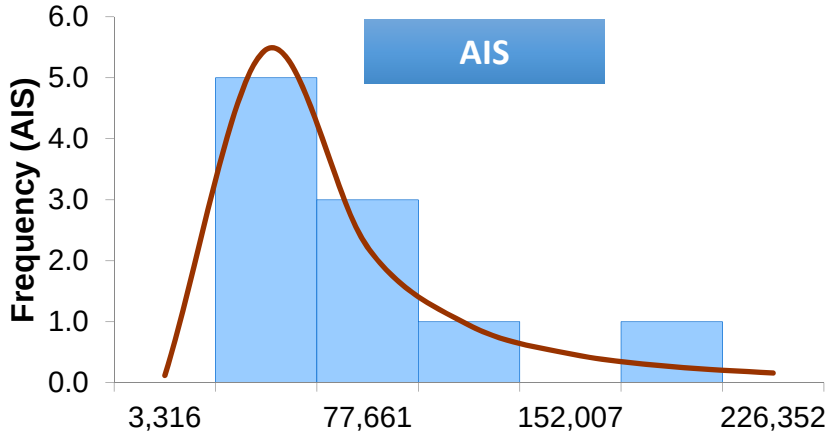
Support (Cost in BY 2016 $)	Sample	Log Normal
Mean	36,095.37	39,483.68
Std Dev	35,829.35	39,076.58
CV	0.99	0.99
Min	2,915.83	
Mode		14,177.15
Max	109,267.40	
Count	13	
Standard Error of Estimate		11,687.15

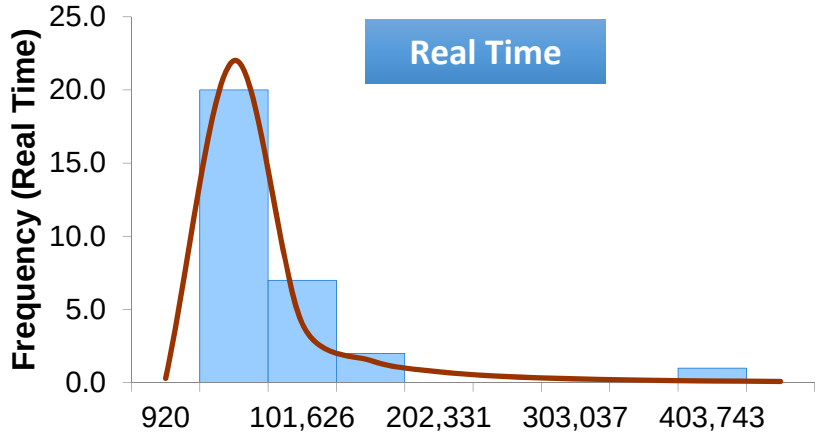


Engineering (Cost in BY 2016 $)	Sample	Log Normal
Mean	203,834.97	240,949.52
Std Dev	329,675.15	795,858.95
CV	1.62	3.30
Min	3,507.83	
Mode		5,862.26
Max	1,336,963.22	
Count	16	
Standard Error of Estimate		41,198.89

- Cost per Software Change is shown by Super Domain
- Software change count only includes program reported software changes. It does not separately include IAVA counts
- Software Changes are also commonly referred to as problem reports, change requests, defects, etc.

Cost Per Software Change
AIS and Real Time Super Domain



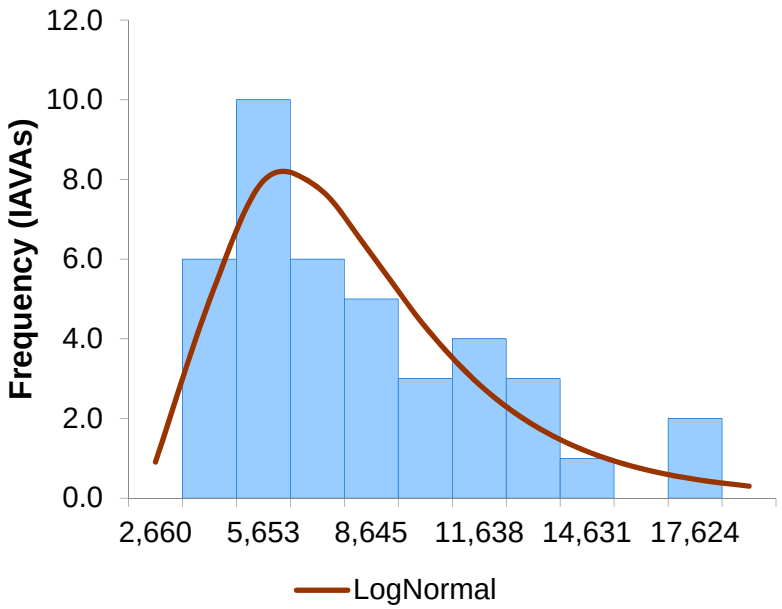


AIS (Cost in BY 2016 $)	Sample	Log Normal
Mean	52,752.21	59,727.81
Std Dev	59,343.51	83,427.32
CV	1.12	1.40
Min	3,315.94	
Mode		11,781.94
Max	189,179.63	
Count	10	
Standard Error of Estimate		10,672.45

Real Time (Cost in BY 2016 $)	Sample	Log Normal
Mean	50,876.79	52,710.70
Std Dev	76,033.47	119,536.17
CV	1.49	2.27
Min	919.76	
Mode		3,462.16
Max	403,742.77	
Count	30	
Standard Error of Estimate		17,432.58

- Cost per Software Change is shown by Super Domain
- One data point was removed from the Real Time dataset for this chart. See backup for distribution with outlier included
- Software change count only includes program reported software changes. It does not separately include IAVA counts
- Software Changes are also commonly referred to as problem reports, change requests, defects etc.

Cost per IAVA

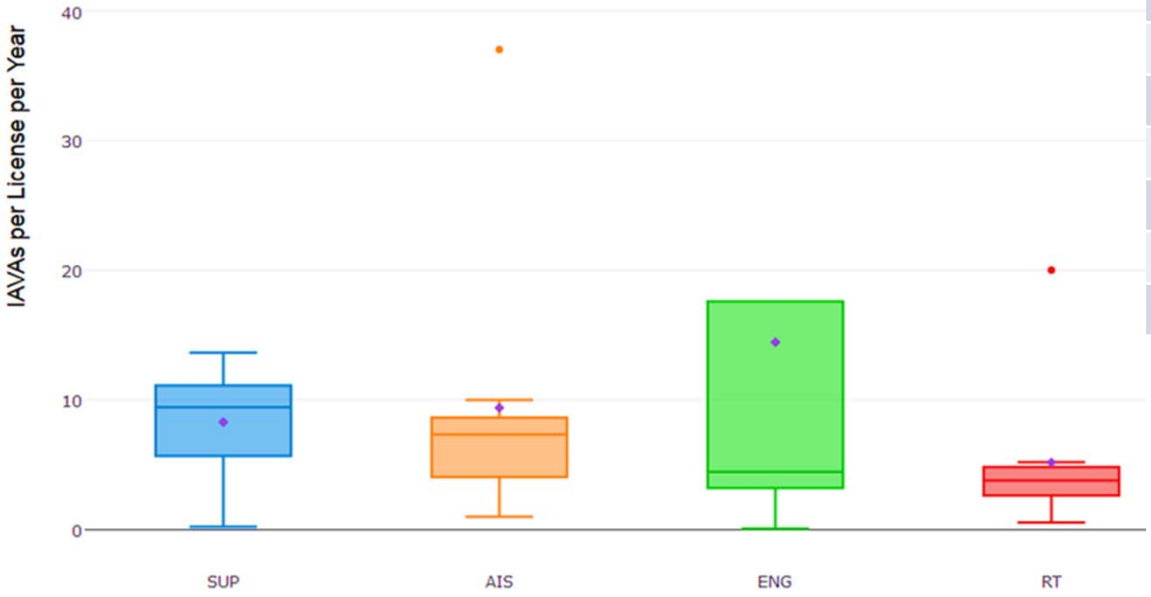


Cost in BY 2016 $	Sample	Log Normal
Mean	7,546.79	7,608.06
Std Dev	3,767.77	3,783.15
CV	0.49	0.49
Min	2,659.91	
Mode		5,461.81
Max	17,623.77	
Count	40	
Standard Error of Estimate		537.34

- Only Information Assurance Vulnerability Alert (IAVA) releases were used, which is a subset of the release data set
- Graph represents (IAVA release cost) / (IAVA count for the release)
  - Includes government and contractor effort

Measurement Benchmarks

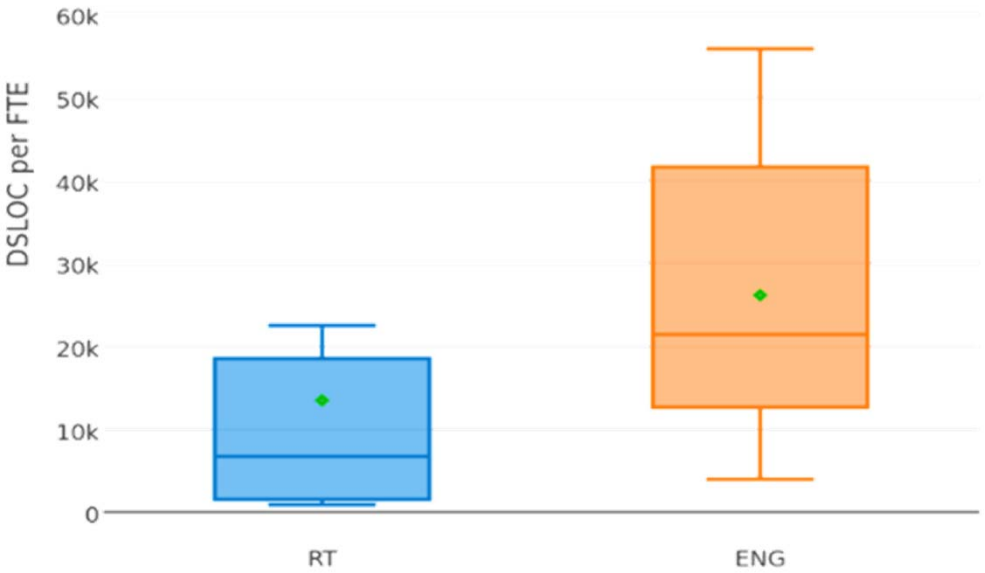
IAVAs Per License Per Year



	SUP	AIS	ENG	RT
Sample Size	8	9	9	9
Min	0.23	1.00	0.08	0.56
Q1	5.68	4.06	3.20	2.64
Median	9.43	7.33	4.46	3.80
Q3	11.11	8.63	17.59	4.82
Max	13.64	37.0	55.89	20.00
Mean	8.29	9.39	14.45	5.18

- Graph illustrates the number of IAVAs per license per year for each Super Domain
- IAVA release rhythm is different for each program. Data is normalized to a yearly amount
- Two Outliers removed (Engineering and Support Domains). See backup for analysis with outliers included

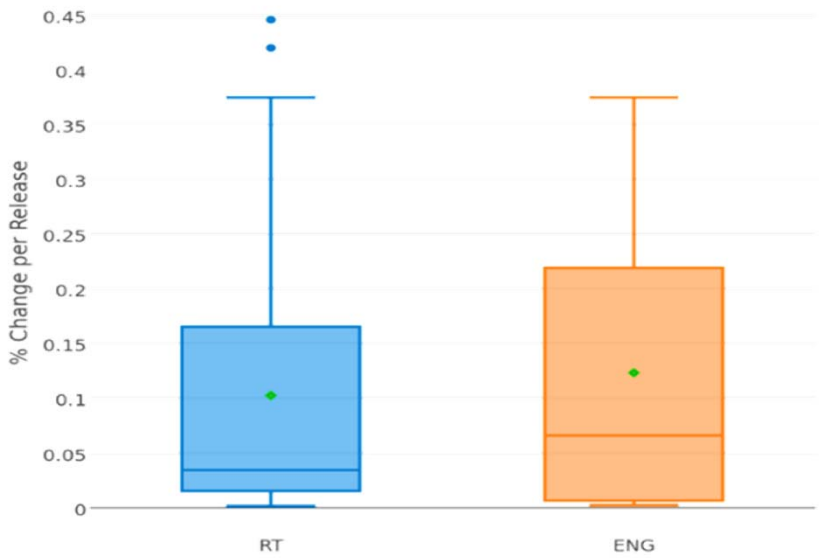
DSLOC per FTE



	RT	ENG
Sample Size	16	6
Min	877.8	3,983.1
Q1	1,558.4	12,687.0
Median	6,736.0	21,436.4
Q3	18,534.1	41,624.9
Max	80,734.1	55,863.18
Mean	13,501.5	26,171.9

- DSLOC represents Delivered Source Lines of Code which counts all code equally
- The earliest baseline size reported was used to represent DSLOC
- Full Time Equivalent (FTE) counts were derived by including the following WBS Elements: SW Change Product (1.0), Program Management (2.0), Sustaining Engineering (5.0), and Certification and Accreditation (4.0)
- FTEs were derived by using labor hours per man-year and labor rate reported for each program
- Only Real Time and Engineering had sufficient data to derive DSLOC/FTE

Baseline Percent Change



	RT	ENG
Sample Size	38	18
Min	0.2%	0.2%
Q1	1.5%	0.7%
Median	3.4%	6.6%
Q3	16.5%	21.9%
Max	44.6%	37.5%
Mean	10.3%	12.3%

- Baseline percent change was calculated for each release as follows:
  - (New Code+ Modified code) / Delivered Code (DSLOC)
  - The earliest baseline size reported was used to represent DSLOC
- Only Real Time and Engineering had sufficient data to derive Baseline Percent Change

Summary of Phase I Data Issues

- Lack of standardized process for data collection for software maintenance
- Inability to map executed cost/effort data to software maintenance output activities and software change products
- Volatile change requirements and execution priorities hinder execution tracking
- Multiple funding streams are often separately managed
- For many systems, the government is heavily leveraged on contractors which limits insight into cost data

Next Steps

- Phase II data collection (in-progress)
  - Phase II includes an additional 208 Army programs
  - Examples of future research using Phase II data:
    - Refined CERs by application domain, organization, operating environment, etc.
    - Schedule Estimating Relationships (SERs)
    - Release rhythm analysis
    - Release characterization (enhancement, defects, cybersecurity) analysis on WBS 1.0 SW Change Product
    - Software Maintenance cost model
  - Phase II data will be used to validate CERs and Phase I analysis
- Systemic data collection
  - The Software Resources Data Reporting for Maintenance (SRDR-M*) closely aligns to the DASA-CE SWM WBS and data requirements
  - Moving forward, the SRDR-M will be utilized to collect SWM data from a large number of programs across the Army
  - Ongoing analysis will be performed as data is made available through the SRDR-M

*See http://cade.osd.mil/policy/dids for more information

Systemic Data Collection Issues

- Implementation of the SRDR-M for most Army programs (>$1M)
  - Requires Army policy for SRDR-M requirement
  - CSDR plan formulation and approval for a large number of programs
    - Relies on use of O&S plan standard for efficient plan generation
    - RFP release notification
- Use of a consistent SWM WBS across programs/services
- Government labor tracking
  - May require forcing function within the Army to ensure government organizations submit SRDR-M
- Cost Model/Training

Contributors

James Judy
NISEC Division Chief
ODASA-CE
703-697-1612

Cheryl Jones
Software Measurement Analyst
US Army ARDEC
973-724-2644

John Staiger
Principal Consultant
Quantitative Software Management, Inc.
(703) 790-0055

James Doswell
Senior Operations Research Analyst
ODASA-CE
703-697-1572

John McGarry
Software Measurement and Analysis Lead
US Army ARDEC
973-724-2644

Robert Charette
President
ITABHI Corporation
(540) 972-8150

Jenna Meyers
Senior Operations Research Analyst
ODASA-CE
703-697-1645

Brad Clark
Vice President
Software Metrics, Inc.
(703) 754-0115

Colin Stratakes
Operations Research Analyst
ODASA-CE
703-697-1626

Doug Putnam
Principal Consultant
Quantitative Software Management, Inc.
(703) 790-0055