In Pursuit of the One True Software Resources Data Reporting (SRDR) Database

ICEAA Conference, IT Track Friday, June 13<sup>th</sup>, 2014, 10:30 a.m. MDT Zach McGregor-Dorsey, Kristen Wingrove, Remmie Arnold, Peter Braxton, Technomics James Doswell, Michael Duarte, ODASA-CE



## Abstract

For many years, Software Resources Data Reports, collected by the Defense Cost and Resource Center (DCARC) on Major Defense Acquisition Programs (MDAPs), have been widely acknowledged as an important source of software sizing, effort, cost, and schedule data to support estimating. However, using SRDRs presents a number of data collection, normalization, and analysis challenges, which would in large part be obviated by a single robust relational database. The authors set out to build just such a database, and this paper describes their journey, pitfalls encountered along the way, and success in bringing to fruition a living artifact that can be of tremendous utility to the defense software estimating community.

SRDRs contain a wealth of data and metadata, and various attempts have been made by such luminaries in the field as Dr. Wilson Rosa and Mr. Mike Popp to excerpt and summarize the "good" data from SRDRs and make them available to the community. Such summaries typically involve subjective interpretations of the raw data, and by their nature are snapshots in time and may not distinguish between final data and those for which updates are expected.

The primary goal of this project was to develop an Access database, which would both store the raw source data in its original form at an atomic level, exactly as submitted by WBS element and reporting event, and allow evaluations, interpretations, and annotations of the data, including appropriate pairing of Initial and Final reports; mapping of SLOC to standard categories for the purposes of determining ESLOC; normalization of software activities to a standard set of activities; and storage of previous assessments, such as those of the aforementioned experts. The database design not only provides flexible queries for quick, reliable access to the desired data to support analysis, it also incorporates the DCARC record of submitted and expected SRDRs in order to track missing past data and anticipate future data.

The database is structured by Service, Program, Contract, Organization, CSDR Plan, and Reporting Event, and is flexible enough to include non-SRDR data. Perhaps its most innovative feature is the implementation of "movable" entities, wherein quantities such as Requirements, Effort, and SLOC, and qualities such as Language, Application Type, and Development Process can be reported at multiple levels and "rolled up" appropriately using a sophisticated set of queries. These movable entities enable the database to easily accommodate future changes made to the suggested format or reporting requirement found in the SRDR Data Item Description (DID).

This work was sponsored by the Office of the Deputy Assistant Secretary of the Army for Cost and Economics, and represents a continuation of the effort that produced the ICEAA 2013 Best Paper in the IT track, "ODASA-CE Software Growth Research." A key motivation of the database is to be able to provide real-time updates to both that Software Growth Model and ODASA-CE's Software Estimating Workbook. We are also collaborating with the SRDR Working Group on continual improvements to the database and how best to make it available to the broader community.



# Outline

- Where we are: Multiple data sources, each with their own limitations
  - Defense Cost and Resource Center (DCARC) SRDRs
  - Popp/Rosa data and evaluations
  - Difficulty in mapping between DCARC data and Popp/Rosa data and evaluations
- Where we are going: Single Relational Database
- How we are getting there:
  - Database overview
  - Challenges

Technomics

- Future goals
- How far we have gotten: Stats on database population

### Where we are...

- DCARC: Defense Automated Cost Information Management System (DACIMS) provides a central repository, but is not a database
  - Authoritative source
  - Non-normalized (not "analysis ready")
  - Inconsistent content and format of reports
    - Abandonment of DD 2630
    - Evolving Data Item Description (DID)
  - Not easily searchable/retrievable
- Popp/Rosa Database:
  - Mike Popp (NAVAIR/Omnitec) has done a yeoman's job of compiling SRDR data as a shareable Flat File (spreadsheet)
  - Further annotated by Dr. Wilson Rosa (then-AFCAA)
  - Non-authoritative source
  - Normalized (analysis ready, maybe?)
- Difficulty in mapping between sources



## **DACIMS** is a Repository

- SRDRs are stored in a file structure tantamount to the one seen on the right
- Manually have to retrieve SRDRs one at a time
- No convenient way to search/filter SRDRs based on data needs





### Popp/Rosa Database

 Popp and Rosa database provides much needed evaluation of SRDRs stored in DACIMS



**Popp Evaluation:** SLOC Represents Build 2 only, but hours are cumulative, 2630-3 for Build 2 adds all previous SLOC into the base



### Popp/Rosa Database

- Mapping Difficulty
  - Popp/Rosa Database does not include CSDR Plan numbers
  - Contractor names often differ between sources
  - Contract names sometimes differ between sources
- Lack of Validation/Verification
  - Simple check to make sure data was correctly transferred from original source to database
  - Are normalization techniques those desired by the end user?



# Where We Are Going...

- Motivation: One Software (SW) Database to support multiple...
  - Models (SW Estimating Workbook, Growth Model, etc.)
  - Analyses (estimates, studies, etc.)

echnomics

- Organizations (ODASA-CE, OSD CAPE, et al.)
- The time is ripe for a more sophisticated tool to support better coordination
  - ODASA-CE actively participating in SRDR Working Group led by Ms. Ranae Woods (AFCAA TD)



- Address both Functionality and Content (and interactions)
- Balance capability and complexity within limited resources

 $E_a (\rightarrow) Y X$ 

 $E_{a}(\rightarrow)XY$ 

# SRDRWG Vision

 "One OSD-hosted, central, user-friendly, authoritative, real-time software cost database and tool"

> - Ms. Ranae Woods AFCAA, Chair Aviation CIPT, May 2014

- OSD-hosted = integrated with CADE
- Central = configuration-controlled, mutually accessible annotations
- User-friendly = queries from relational database, producing "analysis-ready" results
- Authoritative = "community-approved" data traceable back to original submissions
- Real-time = up to date with latest submissions

echnomics

Consistent with OSD CAPE vision for CSDR overhaul

# Having Our Cake...

- Unified Software Database is for:
  - The <u>ODASA-CE Client</u>, built with their data (Army) and models in mind, but the Community\* can leverage both the functionality and content of the database (e.g., OSD CAPE for CADE)
  - The <u>Community</u>, built with a broad (and ever-broadening) perspective, and ODASA-CE can directly benefit from their involvement
- Unified Software Database is:
  - A database proper, to store, relate, and annotate primary source information
  - A <u>data analysis tool</u>, primarily via automated queries to extract and export data in the desired format
- Unified Software Database contains:
  - <u>SRDR data</u>, the official DoD software data source
  - Non-SRDR data, as collected by ODASA-CE/Technomics
- Unified Software Database is:

Technomics

- <u>Backward-looking</u>, capturing legacy data in various formats and annotations thereof
- Forward-looking, enabling improved data collection in the future



# Unified SW Database Vision

#### • A single *relational* Access database that contains:

- Raw source data (fully traceable)
- Data at the level at which it is reported (WBS element, "atomic level")
- Both "initial" and "final" instances of a reporting event
- DCARC CSDR Plan information for reporting events that are still missing or expected in the future
- Assumptions and context about the data that facilitate analysis (e.g., Pairing ID)
- Evaluations of the quality of the data (e.g., knowing that counting rules are not provided in the data dictionary)
- New database provides the ability to:
  - Quickly query data at both the lowest level and summary-levels in order to track progress in obtaining missing data
  - Use the level of data most appropriate for the analysis (e.g., contract vs. plan vs. event)
  - Tag and store "Roll-ups" of data
  - Tag and store Initial/Final pairings of data points
  - Interface with and "feed" multiple workbooks that serve different analytic purposes (without touching or modifying the original data)
  - "Save" queries and dashboards that allow analyst to quickly access often-used sets of data



### Unified SW Database Strengths

- Preserve atomic raw un-normalized SRDR data
- Relational database

echnomics

- Data integrity, flexible queries, etc.
- Enables "crowd-sourcing" community-best version of SRDR database (under aegis of CADE?)
  - Quality assessments, annotations, etc.
- More efficient data ingest
  - $XML \rightarrow DCARC \rightarrow SWDB$
  - Accommodates DID changes, known and unknown
- More rigorous access control and DB exports
  - Full-context versions where NDAs exist
  - Anonymized version (only valuable if you trust the source)

### How We Are Getting There...

- Maintain trace to original data
  - Raw = exactly as submitted (unadulterated)
  - Atomic = at the lowest level submitted
  - Un-normalized = neither mapped, nor rolled up, nor summarized (e.g., ESLOC)
- Provide direct link to source files
- Use "moveable entities" to accommodate reporting at various levels and in non-standard categories



### High-Level Relational Database Structure



### Access Database

Over 50 tables make up the complete relational database in Access. Below is a small sample.



### Database Status

#### Software Report Database

Database Status Report										
	Services to include in status report									
	Army Navy Air Force									
	Department of Defense Missile Defense Agency									
These status reported in relation to DACIMS Summary Report Relevant records in Future reports database for DACIMS expected by Accepted reports DACIMS										
То	otal		All	Some	None					
	<u>69</u>	Programs	2	<u>30</u>	<u>32</u>	44				
1	122	Contracts	<u>12</u>	<u>42</u>	<u>59</u>	<u>53</u>				
2	218	Plans	<u>30</u>	<u>78</u>	<u>100</u>	<u>69</u>				
8	863	Events	<u>306</u>		<u>524</u>	<u>149</u>				

Navigation
Browse/Query Data
Edit/Enter Data
Evaluate Data

#### View Software Estimating Workbook Data



A tool for estimating final effort hours and schedule months for a software development effort based on the contractor's initial estimates. To import data from the database into the workbook, simply open a new instance of the workbook, which is located TBD.



# Searching for SRDRs

#### High-Level Search

Services	Air Force Army Department of Missile Defens Navy	Defense e Agency	Selecting none is the same as selecting all.				
Programs	3DELRR AEHF	Three-Dimensional Expedition Advanced Extremely High Freq	ary Long-Range Radar Juency (AEHF) Satellite Program				
	AMF JTRS	Joint Tactical Radio System Airl	borne & Maritime/Fixed Station				
	AMPV	Armored Multi-Purpose Vehicl	e				
	AOC-WS	Air and Space Operations Cent	er-Weapon System	-			
Contracts	FA8726-08-C-00	008	CSDR Plan Numbers	D-08-A-C1(R1)			
	TBD			D-08-A-C1-S3			
	TBD			D-08-A-C1-S3(R)			
	TBD						

Reporting		\EASS\AMF JTRS\FA8726-08-C-0008\JTRS AMF Sub (Northrop 🔺
Events		\EASS\AMF JTRS\FA8726-08-C-0008\JTRS AMF Sub (Northrop 🗌
	AMF JTRS	\EASS\AMF JTRS\FA8726-08-C-0008\JTRS AMF Sub (Northrop
	AMF JTRS	\EASS\AMF JTRS\FA8726-08-C-0008\JTRS AMF Sub (Northrop
	AMFJTRS	\EASS\AMF JTRS\FA8726-08-C-0008\JTRS AMF Sub (Northrop 👻
	Go to SRDR View	Go to Original



# Viewing\Entering Data

SRDR View	ID			45			
Find Recor	d Previous	Record Next	t Record				
Unclassified  SECURITY CLASSIFICATION		Fields in all c	apital letters	exist on 20	)11 SRDR form.	All others are a	vailable for legacy forms.
SOFTWARE	RESOURCES	S DATA REPOR	RTING: Initiand 60 days afte	I ▼I r start of any	DEVELOPER RE	PORT	
Section 3.1 Section 3.2 Section 3.3 Section 3.4 Section 3.5 Other Reporting Event Contract Initial Estimate							
Secti	on 3.1 REPO	RT CONTEXT	AND DEVEL	OPMENT O	RGANIZATIO	N	
MAJOR PROGRAM a. NAME: U.S. Army's Br	igade Combat	t Team Modern	ization (BCTN	/l) Grour	b. PHASE/M	LESTONE A (T	echnology Development I
REPORTING ORGANIZATION TYPE	NAME/ADDRESS a. REPORTING ORGA			ANIZATION: Systems Division 201 Electronics Blvd., Huntsville, AL 35824			on Systems, Defense le, AL 35824
GOVERNMENT	b. DIVISION:			Same as Reporting Organization			
APPROVED PLAN NUMBER A-10-A-C1-S1	CUSTOMER BAE Systems				СО	NTRACT TYPE	FFP
WBS ELEMENT CODE 1.0	WBS REF	PORTING ELEMENT Ground Combat Vehicle Program					
TYPE ACTION a. CONTRACT NO.: W56HZV-	L1-C-C001	c. SOLICITA	TION NO.	W56HZV	V-11-R-0001	e. TASK ORDE	R / DELIVERY ORDER NO.
b LATEST MODIFICATION: 3		d. NAME:	U.S. Army's	Brigade Cor	nbat Team Mod	N/A	
PERIOD OF PERFORMANCE	PERIOD OF PERFORMANCE APP				SUBMISSION NUMBER		
a. START DATE: 20110528		RDT&E		RESUBMI	SSION NUMBER	0	
b END DATE: 20121205		Procurement		Supersedes Number			
20131205		0&M		REPORT #	S OF	20120127	
	Revisio	on Date		DATE PRE	PARED	20121031	
NAME (Last, First, Middle Initial) DEPART	MENT		1	ELEPHONE	Fax	EMAIL AD	DRESS



## Queries

Choose the records you would like to see:

	🗷 Army	🗷 Navy	🖉 Air Force	
	🗹 Departm	ent of Defense  🕅 M	issile Defense Agency	
Orga to ir Sele Des	nizations nclude ect All elect All	Total Effort Hours Between: Total Size (Total SLG Between:	And:	
Lock Agu BAE Ball Bell	staWest System: Aerospa Helicop	Schedule Months Between:	And:	
Bell Helicop Boeing CAE Inc General Dyn General Elec		MUST include com	uplete data for:	
Hill Hon IBM	AFB eywell	Roll-up Level	Reporting Event	
L-3	-	Platform Type		Run Query
		Commodity Type		

Blank indicates no restricition



### SW Data – Accommodating Different Structures

Language is usually a child of the WBS element and code count is reported separately

2. Product and Development Description	Pero Pro	cent of oduct
1. Primary Application Type: Guidance & Control	2.	100 %
5. Secondary Application Type:	6.	%
9. Third Application Type:	10.	%
13. Fourth Application Type:	14.	%
17-Primary Language Used: C	18.	108%
1 <del>9. Seco</del> ndary Language Used:	20.	%



4. Amount of New Auto-Generated Code developed and delivered (Size in <u>Snc</u> )	205,306
5. Amount of New Hand Written Code developed and delivered (Size in <u>Snc</u> )	6,036

### Here code counts is a sub-element (child) of language

Longange	fine	Generosoft (Total)	Deteriod	Madified	See	Foried	Resort	Des Tatal SLOC	Product Tetal SELOC
Jovial	100,000		2,500	3,000	8,000		6,000		117,999
Assembler	20,000				100	299			20,300

#### Effort is usually reported by Activity

Resource	and Schedule Reporting	Provide Actuals at Final Delivery				
Counting for for each a shown.	rom month 1 at contract award, provide Actual Start and End Month citivity shown. Provide the Actual Total Labor Hours for each activity	Start Month	End Month	Total Hours		
The follo	where covers items should account for all direct hours charged to the set	Twace developer	uent.			
project ( contribu	surce and Schedule Reporting ting from month 1 at contract award, provide Actual Start and End Month ch activity shown. Provide the Actual Total Labor Hours for each activity m. a following seven items should account for all direct hours charged to the s sect use item 7 for any direct hours not accounted for in items 1 through o tribution of indirect hours in the associated Data Dictionary. ware Requirements Analysis d Code and Database Code Design	Explain any				
project ( contribu 1. Software P	use item 7 for any direct hours not accounted for in items 1 through 6), tion of indirect hours in the associated Data Dictionary. Tequirements Analysis	Explain any 3/22/2006	8/27/2008	5,304		

#### Here effort is reported by language

Svs Eng			e		
Phases	Terlahel?	The of Load, Diff Decomposite Street	Textulativ	To of Total Art Development Pilet	
Astall3 (Statl-Months)	31.60				

## Flexible Data Structure

Database captures initial/final "pairings" for analyses that require both. Also provides flexibility to tag and store "roll-ups" using different sets of business rules





21

## Flexible Data Structure

New structure allows us to store "all" the data (multiple sources, multiple levels); provides for total flexibility to compare or merge data from different sources and retrieve the level of data most appropriate for the analysis



22

### Tracking Missing Data – What's the "Universe"?

Database incorporates DCARC-provided tracking sheet that contains all delivered and expected SRDRs for programs still active after 2009

Program	Contract #	CSDR Plan #	Reporting Event	As-Of Date	Due Date	Received Date	ODASA-CE DB?	
Program I	Contract I	Plan I	Event I	12/2007	1/2008	1/2008	YES	Allows us to track
			Event 2	3/2009	4/2009	4/2009	NO	our SRDR data against all
	Contract 2	Plan 2	Event 3	5/2009	6/2009	7/2009	YES	"possible" data
			Event 4	8/2010	9/2010	9/2010	NO	
Program 2	Contract 3	Plan 3	Event 5	7/2011	8/2011	8/2011	NO	
			Event 6	8/2012	9/2012	10/2012	NO	
		Plan 4	Event 7	6/2011	7/2011	8/2011	YES	
			Event 8	9/2014	10/2014	N/A	NO	
			Event 9	5/2012	6/2012	7/2012	NO	
			Event 10	10/2015	11/2015	N/A	NO	



### How it All Fits Together



### Data Normalization Approach

- Dr. Wilson Rosa and Dr. Brad Clark
  - Inspect Data
    - Context Information
    - Effort Data
    - Schedule Data
    - Project Identifiers
  - Correct Data, Evaluate Quality
  - Normalize Data
    - Adjust SLOC data (physical to Logical, ESLOC)
    - Adjust for Missing Effort Data



## Level I Evaluation

- Purpose: Initial evaluation and "organization" of the data needed to get the data in a more usable form
  - With the tags and user-provided data in the Level I evaluation, the database user can develop initial queries of data that can be used to support estimates and other analyses
- Sample items in Level I:
  - Initial/Final pairing tags
  - Identification (and potential addition) of contract-level and build level roll-ups
  - Data dictionary availability
  - Evaluation of the scope of effort represented in the event



# Level I Evaluation: Roll-Up Types

- It is important to clearly define and implement different types of "Roll-Ups" based on data field
  - May require subtle adjustment of database queries
- Summation (distinct)
  - Total SLOC, effort hours, e.g.
- Max/Most Recent (monotonically increasing)
  - Total SLOC
- Max of Max
  - Peak staff, e.g.
- Extremes (Min/Max)
  - Schedule start and end months, e.g.
- Plurality
  - Programming Language, Application Type, e.g.

## Data Quality Analysis

- Leverage to maximum extent previous work of Popp, Rosa, et al.
  - Import where possible, manual review and (re)entry where necessary
- Annotations vs. additional instances of data points (revised/ corrected)





## Level II Evaluation – Vision

- Purpose: To get the data "analysis-ready"\*
- Sample items in Level II:

Technomics

- Mapping of SLOC to our ESLOC categories so that ESLOC can be quickly calculated for each data point
- Mapping of activities to a "standard set of activities" that can be used for effort normalization and cross-data comparisons
- Evaluation of Wilson Rosa/Mike Popp comments and storage of these assessments in a standard fashion (so they can be quickly used to exclude/include certain data points)
- Review of Data Dictionary and entry of standard information from the dictionary in our database (examples: code counting logic, definition of each activity)
- Evaluation and entry of additional "contextual" information that can help with analysis such as Operating Environment and Productivity Type

<sup>\*&</sup>lt;u>Note</u>: Before Level II Evaluations are completed, database can be used to quickly query for a set of data points that meet initial criteria but some of the activities listed below would still need to be conducted manually before the data could be used to support an estimate or as part of a study like the Growth Study. The Level II Evaluation simply completes these steps beforehand.

### How far we have gotten...

Multiple iterations with ODASA-CE client

- Demonstration of incremental capability

- Parallel data entry for Army SRDRs
  - Import of legacy non-SRDR data, All SRDRs metadata
- Version 1.0 incorporates all essential functionality
  - Drill-down
  - Data entry / SRDR view
  - Evaluations (Level I and Level 2)
  - Query
  - Go to Original
- Accompanying User Guide
- Prioritize future enhancements and content updates

## Data Inventory

- 1007 Total Reporting Events (according to DCARC reports)
  - 863 Accepted Events
  - 144 Due in the Future

echnomics

- We have all 863 accepted events obtained in a bulk download request from DCARC
- Approximately 306 of these 863 have been entered into the database
- Dashboard and Drill-Down functionality in current database support further exploration

### Contents of Database



### **Database** Population

- Based on client (ODASA-CE) and community (OSD CAPE, SRDRWG) priorities
- Leverage existing resources to maximum extent possible
  - Import Mike Popp spreadsheet, e.g.
- Analyst involvement still crucial
  - At a minimum, validate against original submissions



### Comparison: SEI SCAR

	SCAR	Unified SWDB
Sponsor	USD(AT&L)?	ODASA-CE
Developer	Software Engineering Institute (SEI)	Technomics
Data	5 programs (pilot)	18 programs (Army), 58 programs (Total)
Metadata	??	All SRDRs (DCARC import), including Future
Data Entry	Scraper (DD 2630 only)	Import/manual
Platform	Web-based	Microsoft Access
Popp/Rosa	Separate repository?	Direct incorporation/annotation
Database Components	4 Databases,	2 Databases,

"Software Cost Analysis Repository" webinar, Brad Clark, Jim McCurley, Software Engineering Institute (SEI), July 2, 2013

Disclaimer: Direct insight into SCAR is limited at this time.

## The Bigger Picture

- Improve Accessibility and Quality of existing data (Past)
- Improve guidelines for ongoing data collection,
   i.e., SRDR DID (Present, Pull)
- Improve capture for incoming SRDRs (Present, Push)
- Improve mechanism for data collection on new programs, i.e., XML (Future)



# Bibliography

- Cost and Software Data Reporting (CSDR) Manual, 4000.04-M-1, CAPE, November 2011
- Initial Software Developer Report, DI–MGMT-81739B (PDF), 20110525, DI–MGMT-81739B
- Final Software Developer Report, DI-MGMT-81740A (PDF), 20110518, DI-MGMT-81740A
- "How I learned to stop worrying and love the Software Resource Data Report", Michael Popp, DoDCAS 2012
- "Data Inspection and Normalization Guide: Software Resource Data Reports (SRDR)," Wilson Rosa (DHS), Joseph Dean (AFCAA), and Brad Clark (AFCAA)
- "ODASA-CE Software Growth Research", Lauren Nolte, Kevin Cincotta, Eric Lofgren, Remmie Arnold, ICEAA 2013 (Best Paper, IT Track)
- "Software Cost Analysis Repository" webinar, Brad Clark, Jim McCurley, Software Engineering Institute (SEI), July 2, 2013
- "Software Estimating Handbook Version 1.0," Office of the Deputy Assistant Secretary of the Army, Cost and Economics (ODASA-CE), September 2012
- "Software Database User Guide," Office of the Deputy Assistant Secretary of the Army, Cost and Economics (ODASA-CE), March 2014

# In Pursuit of the One True Software Resource Data Reporting (SRDR) Database

Backup



# Army SRDR Programs Summary

- Ground Vehicles
  - GCV
  - JLTV
  - PIM
- Missiles and Munitions
  - Excalibur
  - JAGM
  - GMLRS
- Aircraft
  - Apache
  - UH-60M
  - ARH

- Electronics
  - JTRS-GMR
  - WIN-T Increments 2 and 3
  - DCGS-A
  - FBCB2
- System of Systems
  - JLENS
  - IAMD
  - FCS
  - GCSS
  - GFEBS



# Army Non-SRDR Programs Summary

#### Ground Vehicles

- EFV
- Missiles and Munitions
  - AIM-9X Block II
  - AARGM
  - SM-6
  - SDB II
- Aircraft
  - B-2 EHF
  - VH-71
  - Super Hornet
  - C-130 AMP
  - Hercules
  - H-I Upgrades
  - B-2 RMP
  - E-2D AHE
  - F-22
  - KC-46A
  - B-2 DMS
  - CH-53K
  - MH-60R
  - EA-18G

Technomics

- Electronics
  - NMT
  - JATAS
  - CAC2S
  - G/ATOR
  - MPS
  - NAVY ERP
  - MP RTIP
  - IDECM
  - FAB-T
  - ADS
  - CEC
- UAV
  - VTUAV
  - MQ-4C
- Ships
  - LCS
  - Cobra Judy Replacement
- Space
  - SBIRS HIGH
  - GPS OCX
  - NAVSTAR GPS
  - EPS
  - MUOS