Measurement & Analysis in the Real World

Tools for Cleaning Messy Data Will Hayes – SEI Robert Stoddard – SEI Rhonda Brown – SEI **Software Solutions Conference 2015** November 16–18, 2015

Carnegie Mellon University

© 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release; Distribution is Unlimited Copyright 2015 Carnegie Mellon University

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[Distribution Statement A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

DM-0003055



Agenda



Introduction Matching Information Needs Getting to the Data Cumulative Flow Diagrams Tool Demonstration Predictive Modeling

Complementary but Different Focus

Government Program Office

- Assess forecasted risk
- Manage to outcomes
- Responsible for total cost of ownership (and current cost)
- Obliged to seek out and communicate user needs

Strive to avoid directing the contractor on HOW to work

Development Contractor

- Predict performance
- Control performance drivers
- Responsible for meeting current commitments
- Subject to re-direction based on user needs

Influence on WHAT to build may be constrained by contract



Different Audiences for Metrics and Status

Program office personnel who interact directly with contractors

- Generally need insight at a finer level of detail
- Must maintain visibility/continuity over time

Stakeholders in the program, beyond Program Management

- May focus on specific topics to the exclusion of all else
- Participate in less frequent status discussions, perhaps

Senior leadership who oversee the program office

- Focus on performance of the program, not just this contract
- Frame of reference may be broader and more long term



Matching Information with Needs **Re-Casting Metrics for** the Target Audience



© 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release Distribution is Unlimited



Time-Horizon and Specificity

The chart below shows the trend in estimated size, with thresholds for potential corrective action.

SIZE ESTIMATE TRENDLINE



VARIANCE FROM FSTIMATE Estimate 10% Threshold 20% Threshold 350 300 250 KSLOC 200 150 100 50 0 2 3 5 6 7 8 9

The simplified version above shows only 9 weeks, focusing only on variance from original estimate.

Last 9 Weeks

Choose time-horizon and specificity to meet audience needs



Software Engineering Institute | Carnegie Mellon University

Measurement and Analysis in the Real World November 18, 2015 © 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Converging Indicators



4000 3500 3000 2500 2000 1500 1000 500 0 Initial Size Estimate Current Size Estimate

Baseline Added

Estimated Size (KSLOC)

Some information is visible only when you combine data



Software Engineering Institute **Carnegie Mellon University** Measurement and Analysis in the Real World November 18, 2015 © 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Releases Distribution is Unlimited

Useful Graphical Tool Cumulative Flow Diagram





Software Engineering Institute Carnegie Mellon University © 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release Distribution is Unlimited

Constructing a Cumulative Flow Diagram₁





 Measurement and Analysis in the Real World November 18, 2015
 10

 © 2015 Carnegie Mellon University
 10

 Distribution Statement A: Approved for Public Release;
 10

Constructing a Cumulative Flow Diagram₂



 Measurement and Analysis in the Real World

 November 18, 2015

 © 2015 Carnegie Mellon University

 Distribution Statement A: Approved for Public Release;

 Distribution Statement H: Approved for Public Release;

Constructing a Cumulative Flow Diagram₃

... adding the next 7 times





Software Engineering Institute **Carnegie Mellon University** Measurement and Analysis in the Real World November 18, 2015 12 © 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Releases Distribution is Unlimited

Constructing a Cumulative Flow Diagram₄

... now we are looking at the flow from "identified"... to "Closed"... This view starts to show patterns a little easier...



Tell-Tale Signals



Measurement and Analysis in the Real World November 18, 2015 © 2015 Carnegie Mellon University 14 Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Exercise: What is Going on Here?



 Measurement and Analysis in the Real World

 November 18, 2015

 © 2015 Carnegie Mellon University

 Distribution Statement A: Approved for Public Release;

 Distribution Unlimited

Exercise: What *MIGHT BE* **Happening₁**



At time 2, and then again at time 4, the number of items "In Process" goes to zero.

- Have we lost the resource(s) that were preparing the items in the "Waiting" state?
- Is this intentional, due to limited resource(s) who can work on items in the "In Process" state?

Exercise: What *MIGHT BE* Happening₂

The number of items that are "In Process" is growing over time.

- The rate at which things enter "In Process" is greater than the rate at which things leave "In Process."
- Are people moving onto new items without completing their work?
- Are new resources being added, who start new work at each time period?
- Are things moving into the "Done" state quickly enough?





Getting to the Data Mining a Configuration Management Database

or Application Lifecycle Management Tool





Software Engineering Institute | Carnegie Mellon University

© 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Activity Flow: Mining the Database



Weekly analysis activities comprised of these steps:

- Data pulled directly from Configuration Management system
- Inventory 'change records' to verify completeness and accuracy
- Tabulate by pre-defined time intervals and validate totals
- Chart data using Cumulative Flow Diagrams to analyze progress



 Measurement and Analysis in the Real World

 November 18, 2015

 © 2015 Carnegie Mellon University

 Distribution Statement A: Approved for Public Release;

 Distribution Statement H: Approved for Public Release;

Details: Process View



Process flow for a defect being worked

- Entry/exit criteria for each step
- Individual assigned to work each one
- Progress through the process tracked
- Database fields used to record
 - Current state in the process
 - History of progression through the states
 - Date/time stamp for each state change
 - ... and lots of other information

Details: Raw Data

Main Data Table

Defect ID	Title	Description	Severity	
1000001	Dropped data	Message traffic is overwritten when buffer size not specified in	1	
1000002	Missing header	File never read at initialization due to missing pointer in	2	
1000003	Unpredictable close	Process XYZ terminates while opening file	1	

Change Auditing Table

ID	Old State	New State	TimeStamp	LOTS of other data
1000001	New	Open	mm/dd/yy hh:mm:ss	
1000001	Open	Assign	mm/dd/yy hh:mm:ss	
1000001	Assign	Test	mm/dd/yy hh:mm:ss	
1000002	New	Open	mm/dd/yy hh:mm:ss	



Details: Mining the Change Auditing Table

This database table provides:

- Date and time when each item entered a given state
- History of all such transitions since the record was created

Using that information, we can derive:

- How many records are in each state at a given time
- How long each item stayed in any particular state

This allows us to:

- Draw Cumulative Flow Diagrams to show flow
- Model the state-transition activity with a predictive model



Leveraging Excel and Access with VB **Tool Demonstration**



© 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release; Distribution is Unlimited





Measurement and Analysis in the Real World November 18, 2015 © 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Useful Statistical Tool Predictive Modeling





Software Engineering Institute Carnegie Mellon University © 2015 Carnegie Mellon University Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Building Models

Data derived from time stamps

- Duration associated with each state in the sequence
- Information about range of time seen in the past
- Benchmarks for durations can aid in planning

A variety of modeling techniques can be applied



Predicting Change Request Closure



- Software Engineering Institute | Carnegie Mellon University

 Measurement and Analysis in the Real World

 November 18, 2015
 2015

 © 2015 Carnegie Mellon University
 27

 Distribution Statement A: Approved for Public Release;
 Distribution Statement A: Approved for Public Release;

Predicting Remaining Changes to Close



Software Engineering Institute Carnegie Mellon University

Measurement and Analysis in the Real World November 18, 2015 © 2015 Carnegie Mellon University 28 Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Time in State Compared to Past Release



Software Engineering Institute Carnegie Mellon University

 Measurement and Analysis in the Real World

 November 18, 2015
 29

 © 2015 Carnegie Mellon University
 29

 Distribution Statement A: Approved for Public Release;
 Distribution Statement A: Approved for Public Release;

Tracking Software Quality Trends



Software Engineering Institute Carnegie Mellon University

Measurement and Analysis in the Real World November 18, 2015 © 2015 Carnegie Mellon University 30 Distribution Statement A: Approved for Public Release; Distribution is Unlimited

Modeling Flow of Software Change Requests



 Measurement and Analysis in the Real World November 18, 2015
 Main Statement A: 2015
 Main Statement A: Approved for Public Release; Distribution is Unlimited
 Main Statement A: Approved for Public Release;
 Main Statement A: Appro

Contact Information

Presenters

Will Hayes Client Technical Solutions Division Telephone: +1 412.268.6398 Email: wh@sei.cmu.edu

Rhonda Brown Software Engineering Measurement and Analysis Telephone: + 1 412.268.3963 Email: <u>rbrown@sei.cmu.edu</u> Robert Stoddard Software Engineering Measurement and Analysis Telephone: + 1 412.268.1121 Email: <u>rws@sei.cmu.edu</u>

