

Can You Trust Your Data? Establishing the Need for a Measurement and Analysis Infrastructure Diagnostic

Mark Kasunic
James McCurley
David Zubrow

November 2008

TECHNICAL NOTE
CMU/SEI-2008-TN-028

Software Engineering Process Management Program
Unlimited distribution subject to the copyright.

<http://www.sei.cmu.edu>



This report was prepared for the

SEI Administrative Agent
ESC/XPK
5 Eglin Street
Hanscom AFB, MA 01731-2100

The ideas and findings in this report should not be construed as an official DoD position. It is published in the interest of scientific and technical information exchange.

This work is sponsored by the U.S. Department of Defense. The Software Engineering Institute is a federally funded research and development center sponsored by the U.S. Department of Defense.

Copyright 2009 Carnegie Mellon University.

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.

Use of any trademarks in this report is not intended in any way to infringe on the rights of the trademark holder.

Internal use. Permission to reproduce this document and to prepare derivative works from this document for internal use is granted, provided the copyright and "No Warranty" statements are included with all reproductions and derivative works.

External use. This document 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 external and/or commercial use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

This work was created in the performance of Federal Government Contract Number FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. The Government of the United States has a royalty-free government-purpose license to use, duplicate, or disclose the work, in whole or in part and in any manner, and to have or permit others to do so, for government purposes pursuant to the copyright license under the clause at 252.227-7013.

Table of Contents

Acknowledgments	ix
Abstract	1
1 Introduction	3
2 Importance of Measurement and Analysis	4
2.1 Impacts of Poor Data Quality	4
3 Common Errors in Measurement and Analysis	6
3.1 Missing or Unclear Measurement Goals	6
3.2 Lack of Resources and Training	6
3.3 Differing Operational Definitions	6
3.4 The Measurement Method Itself	7
3.5 Lack of Rigor in the Measurement Process	7
3.6 Dysfunctional Incentives	8
3.7 Lack of Priority or Interest in the Measures and Their Analysis	8
3.8 Constrained Precision of Measurement	8
3.9 Failure to Analyze for Variation	8
3.10 Data Entry Errors	8
4 Creating a Measurement and Infrastructure Diagnostic	9
4.1 Defining Information Quality	9
4.2 Measurement Information is the Product of a Process	9
4.3 Improving Information Quality Begins with Understanding Where You Are	10
4.4 A Diagnosis Method for the Measurement Process	11
4.5 Elements of MAID	13
4.5.1 Assessing the Measurement Process	13
4.5.2 Assessing Information Quality	13
4.6 Application Scope – Conducting a MAID Assessment	14
4.7 Looking Ahead	16
References	17

List of Figures

Figure 1:	Simplified Representation of the Measurement Process	10
Figure 2:	SEI IDEAL Model	11
Figure 3:	MAID Focuses on the Diagnosing Phase of the IDEAL Improvement Framework	12
Figure 4:	Inputs to be Considered for Development of MAID	12
Figure 5:	A Hierarchy of Data Quality Dimensions	14
Figure 6:	Representation of Enterprise-Based Entities	15

List of Tables

Table 1:	Resources Used as Guidance to Develop MAID	13
Table 2:	Scope of MAID is Determined by Management Stakeholders	15

Acknowledgments

Thanks to William Nichols for his review and feedback on the draft manuscript. The authors would also like to acknowledge the excellent editing support provided by Erin Harper.

Abstract

An organization's measurement and analysis infrastructure directly impacts the quality of the decisions made by people at all organizational levels. Ensuring information quality is a challenge for most organizations—partly because they might not be fully aware of their own data quality levels. Without this information, they cannot know the full business impact of poor or unknown data quality or determine how to begin improving their data. This report describes common errors in measurement and analysis and the need for a criterion-based assessment method that will allow organizations to evaluate key characteristics of their measurement programs.

1 Introduction

Organizations run on data. They use it to manage projects and the enterprise, select products to develop, make decisions, and guide improvement. Measurement and analysis is a critical enabling activity for superior performance. In the quest to beat the competition by making products and processes better, faster, and cheaper, having and using high quality data and information is an input to every key decision.

But how reliable is the data organizations collect and use? How good is the information generated from the data? Given the importance of measurement and analysis in decision making, organizations need to ensure their data, analyses, and corresponding reports are of high quality and meet the information needs of practitioners and managers. Yet organizations often have no method for evaluating the quality of their measurement and analysis infrastructure and very little trust in their own data.

Attention to data quality has been focused primarily in the areas of customer relationship management (CRM) and business intelligence (BI). Organizations rushed to implement CRM systems in the 1990s, but they soon realized poor data could derail the payoff they expected from their substantial investments in these systems. Industry experts, vendors, and consultants were quick to identify problems and offer solutions [Krill 2001].

While the focus on data quality has not carried over to project measurement data, interest is growing. This report describes the state of the practice in dealing with project- and process-related data and introduces a method for evaluating an organization's measurement and analysis infrastructure. The method draws upon standards, best practices, and techniques from software engineering and process improvement.

2 Importance of Measurement and Analysis

In the context of software and systems development, organizations measure to better understand and control resource use in the production of products and services. Measurement and analysis provides the means for organizations to know, manage, and improve their performance. Essentially, they measure to *characterize, evaluate, predict, and improve* what they are producing and how it is produced.

They need to *characterize* processes, products, and environment to provide baselines for future assessments. Without an understanding of their current state, future changes cannot be evaluated effectively and the use of measurement is ad hoc.

Next, they need to *evaluate* their projects and processes. Once the status is determined, it can be controlled. With suitable control, organizations can *predict* performance more realistically and set achievable goals for quality, schedule, and cost.

Finally, they can *improve* production and products by identifying the root causes of problems and opportunities for improvement. This basic measurement structure has been used for many years, but in software development we find that many of these concepts seem hard to envision, much less implement. Without a good capability for measurement and analysis, an organization is unable to accurately and objectively know, manage, and improve its performance.

2.1 Impacts of Poor Data Quality

The problem of poor-quality data has far-reaching consequences. The Gartner Group predicted that in 2006, at least 25% of the data used by the Fortune 100 companies would be inaccurate [Friedman 2005].

Some of the problems linked to the use of poor-quality data include the following:

- poor project, cost, and schedule estimation
- poor project, cost, and schedule tracking
- inappropriate staffing levels
- flawed product architecture and design decisions
- ineffective and inefficient testing
- fielding of low quality products
- ineffective process change

In short, an organization's measurement and analysis infrastructure directly impacts the quality of the decisions made at all levels, from practitioner and engineer through project and senior managers.

In addition to those directly involved in the development of software intensive systems, acquirers of those systems also have a vested interest in the quality of the data and information produced by development organizations. DoD and other government acquirers have long recognized the need for visibility into the software development process. Measures can relate to product status, project performance, or process capability. Over time, directives and guidance have been promulgated,

including MIL-HDBK-881A for the measurement of development tasks in the work breakdown structure; contracting guidance on the specification of technical performance measures; requirements for standard cost and schedule data associated with the Earned Value Measurement System (EVMS); and the Software Resources Data Report (SRDR) that captures the estimated versus actual characteristics of new software developments or upgrades in terms of measures of requirements, size, effort, functionality, and quality (defects) [DoD 2005, DCARC 2008].

Although the Data Warehousing Institute estimates that data quality problems cost U.S. businesses more than \$600 billion per year, 61% of respondents to a 2004 TDWI-Forrester Technology Survey said they do not use any method or tool to assess data quality in their organizations [TDWI 2008]. Furthermore, little attention has been paid to which infrastructure requirements enable the production of high quality data and information and how to use that information to produce higher quality decisions within organizations.

3 Common Errors in Measurement and Analysis

Measurement errors come from several sources within a development organization. These errors undermine the quality of the data and its potential contribution to organizational performance. Some of the most common measurement errors are discussed in this section.

3.1 Missing or Unclear Measurement Goals

When underlying measurement goals and rationale are not shared, other activities can steal priority and undermine effective implementation of the measurement process. Organizations should not assume that those who collect and report measurement data understand the goals that are associated with those measures. In the 2006 State of the Measurement Practice Survey, there were significant differences between staff and management responses when asked how often they understood the purpose for the data they collected or reported. While approximately 75% of the program managers and executives reported that they frequently understood why data was being collected, only 52% of the programmers and 60% of the engineers chose that option [Kasunic 2006]. Measurement goals must be documented and reviewed on a regular basis to promote a common and shared understanding across the organization.

3.2 Lack of Resources and Training

A significant amount of measurement error is introduced by the use of inappropriate data analysis techniques. Similarly, opportunities for discovery are missed when individuals are unaware of analysis methods that can transform the data into useful information that supports decision-making. Organizations must be willing to invest in the development of a cadre of personnel with the appropriate competencies to support implementation of a measurement program.

3.3 Differing Operational Definitions

When measurement is used for simple comparison or benchmarking, the measures to be compared *must* be commonly defined. However, in the software development world, measurement definitions have not been standardized. In fact, only 36% of the participants in the SEI's 2006 State of the Measurement Practice Survey strongly agreed or agreed with this statement: "The definitions of measures that are used in my organization are commonly understood and consistent" [Kasunic 2006].

Consider the case of four different software projects that have similar characteristics. Each project measures productivity and uses the common term "productivity" to refer to the measure. However, the actual definitions that have been assigned to the term are different [Kasunic 2008]. Some other terms that commonly cause problems when an operational definition is not provided are *project duration*, *defect severity or type*, *lines of code (LOC)*, and *milestone completion*. If operational definitions are not agreed upon at the start of a project and used consistently across the organization, the data collected will be unreliable when combined or compared.

3.4 The Measurement Method Itself

How to measure something is a choice. The choice involves measurement scales, technology, and methods. For instance, if you are measuring a room, you could use a ruler to do this and mark the end of it with your finger, put it down again where your finger is, and continue on till you reach the other wall. Alternatively, you could use a tape measure that is longer than the room. The former is likely to have more measurement error than the latter simply because of the method and technology used. It might be improved by making a mark on the floor and carefully aligning the ruler with the mark.

As in any process, a measurement process can be subject to both common and special cause variation. Control is attained only when the sources of variation are identified and mitigated to the extent possible and affordable. When human activity is integral to the process, similar measurement error as described above can occur, particularly when tasks such as identification, rating, or ranking are involved. Without a good sense of whether such activities are repeatable and reproducible—within and between personnel—management cannot form a valid opinion as to the accuracy or precision of measurements and therefore the analyses and information produced from the data.

3.5 Lack of Rigor in the Measurement Process

A lack of rigor in data collection can lead to sloppy, unreliable data. Three common behaviors related to a lack of rigor follow.

1. **Guessing rather than measuring.** When asked to report data, organizations or individuals who do not have a measurement plan in place might instead use estimation or guessing to provide the information. For example, employees might estimate how long they spent on a task at the end of the month when filling out effort or time sheets instead of tracking the actual time spent each day.
2. **Biased data collection.** Organizations or individuals sometimes focus data collection on areas that are already successful and ignore problem areas to make their data appear more positive. For example, a business might intentionally or unintentionally survey only happy or returning customers and exclude unhappy ones.
3. **Shifting the purpose associated with a measurement.** This often happens when measures are collected without a link to an explicit purpose. For instance, lines of code (LoC) might be collected and used as a measure of product size, project progress, or even product complexity. Understanding the purpose and intended use should influence the measurement method and process. When these are not adjusted, the measure itself is likely to have greater error with respect to the new purpose and greater interpretation ambiguity. Problems also occur if data is collected for a specific purpose, then used for something else. For example, if effort data is collected for accounting and employees are only allowed to work 40 hours a week, the data they submit will not include any overtime they spent on the project. If project managers use that data to estimate how long a similar project might take, their estimates will be inaccurate.

3.6 Dysfunctional Incentives

Measurement programs linked to dysfunctional incentives can cause the data reported to be skewed. For example, if employees are rewarded for high productivity as measured by lines of code per hour, they will be motivated to increase the number of lines they write. However, if they use three lines to write what they previously would have written in one, there has not actually been an increase in productivity.

3.7 Lack of Priority or Interest in the Measures and Their Analysis

While organizations generally indicate that they want data, they sometimes place a low priority on its collection. If there are no visible uses of the data and no sustained management sponsorship for measurement and analysis activities, people have little motivation to spend their time collecting it. A lack of consequences associated with poor data collection or measurement also indicates a low priority for an organization.

3.8 Constrained Precision of Measurement

Accuracy is important when recording data. Although it might not seem important that the three hours of effort expended on a task was recorded as 0.5 days, the cumulative effect of such deviations from the true value in multiple entries can make data unreliable.

3.9 Failure to Analyze for Variation

Often, averages are used to characterize performance measurements. In processes, this can be particularly misleading without an associated measure of variability. Control of process performance outcomes relies on stability, which is effectively measured by identifying the underlying distribution of the data and addressing root causes of abnormal or undesirable variation.

3.10 Data Entry Errors

Data entry errors pose an ongoing threat to data quality. Some examples include numbers or letters transposed, incorrect numbers entered, decimal points within numbers misplaced or missing, the wrong data field selected on the data entry form, and the use of incorrect formats.

In some cases, missing data is reported as the number “0” instead of “missing.” This is a potentially subtle error. In some databases, data fields are set at 0 and it is not possible to distinguish between an activity being performed and finding no defects and an activity not being performed at all. The latter case should have a “missing data” or “not yet performed” indicator rather than a 0 to avoid problems when performing calculations.

4 Creating a Measurement and Infrastructure Diagnostic

4.1 Defining Information Quality

The words “quality” and “high quality” conjure up a number of images: expensive cars, expensive clothes, expensive high-definition televisions ... expensive things. However, it is a mistake to always associate quality with cost. Quality is not about luxury or superiority—it is about perceptions and expectations. A hand-crafted tortoiseshell fountain pen with an engraved gold nib, costing \$400, is a quality writing instrument. However, a mass-produced plastic ballpoint pen that does not leak or smudge, operates until the ink is exhausted, and costs 99¢ is also a quality writing instrument.

Philip Crosby provides a useful definition of quality wherein he defines meeting expectations as *conformance to requirements* [Crosby 1979]. Quality can only be measured in terms of requirements and conformance to requirements. Quality is not about perfection; it is concerned with the task of meeting clearly specified and achievable levels of performance. These levels of performance can be applied equally well to manufactured goods or to the delivery of services such as health care, the provision of training, and even to the provision of measurement information.

As we extend Crosby’s definition of quality to information quality, one must identify the “customer” of data – the knowledge worker who requires data to make effective decisions. Information quality exists when information enables knowledge workers to accomplish their enterprise, organizational, and project or team goals. In order to meet the needs of knowledge workers, measurement information products must have high quality. Larry English defines three components of information quality:

- **Data definition quality**—the quality of the specification of data, including a clear, precise, and complete definition. Data definition is to a measurement specification what product specifications are to a manufactured product such as an automobile or an airplane. Data definition represents integrity criteria for data content. Without data definition quality, there is no clear communication that information producers can use to determine what data they are expected to capture. Nor are there clear criteria to measure data content quality.
- **Data content quality**—the correctness of data values. It includes completeness, non-duplication, conformance to the defined data definitions, and the accuracy of the data.
- **Data presentation quality**—the quality of the information product as delivered to the knowledge worker, transforming the data material into a useful information product that supports decision making. It includes timeliness (i.e., you have the information when you need it) or accessibility (i.e., you can get the information when you need it) of the data, as well as a presentation format that easily communicates the significance of the data in a way that is fit and effective for the knowledge worker’s use [English 2004].

4.2 Measurement Information is the Product of a Process

Measurement information is the product of a process that creates and/or updates it. The information potentially becomes the supply material or “input” for many other business processes. The

supply chain of measurement information as a product is illustrated in Figure 1 [Chrissis 2006]. As the product of a process, the same principles of quality improvement that Deming, Crosby, and Juran applied within the manufacturing domain can and should be applied to the measurement process in order to improve information quality [Deming 1986, Crosby 1979, Juran 1998].

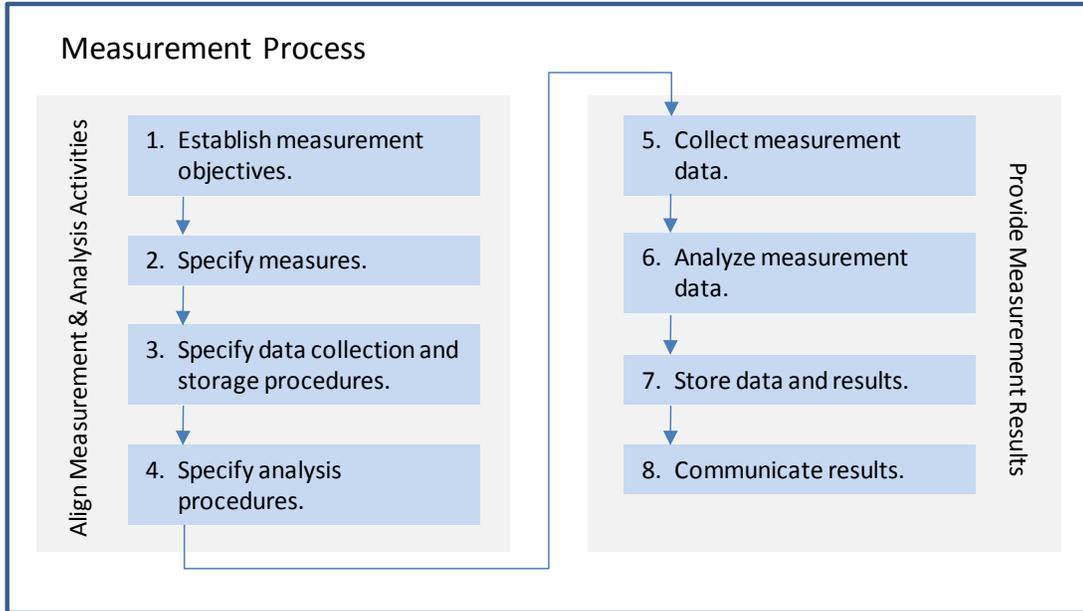


Figure 1: Simplified Representation of the Measurement Process

4.3 Improving Information Quality Begins with Understanding Where You Are

As we consider improvement of the measurement process and data quality, the SEI IDEALSM model comes to mind [SEI 2008]. This model is illustrated in Figure 2.

The IDEAL model is an organizational improvement model that serves as a roadmap for initiating, planning, and implementing improvement actions. The IDEAL model is named for the five phases it describes: initiating, diagnosing, establishing, acting, and learning.

Note that the IDEAL model emphasizes that a complete understanding of the current state of a process is essential before improvement changes are implemented. This understanding is achieved during the diagnosing phase. During the diagnosing phase, two characterizations of the organization are developed: the current state of the organization and the desired future state. These organizational states are used to develop a practical and effective plan for improving the process.

SM IDEAL is a service mark of Carnegie Mellon University.

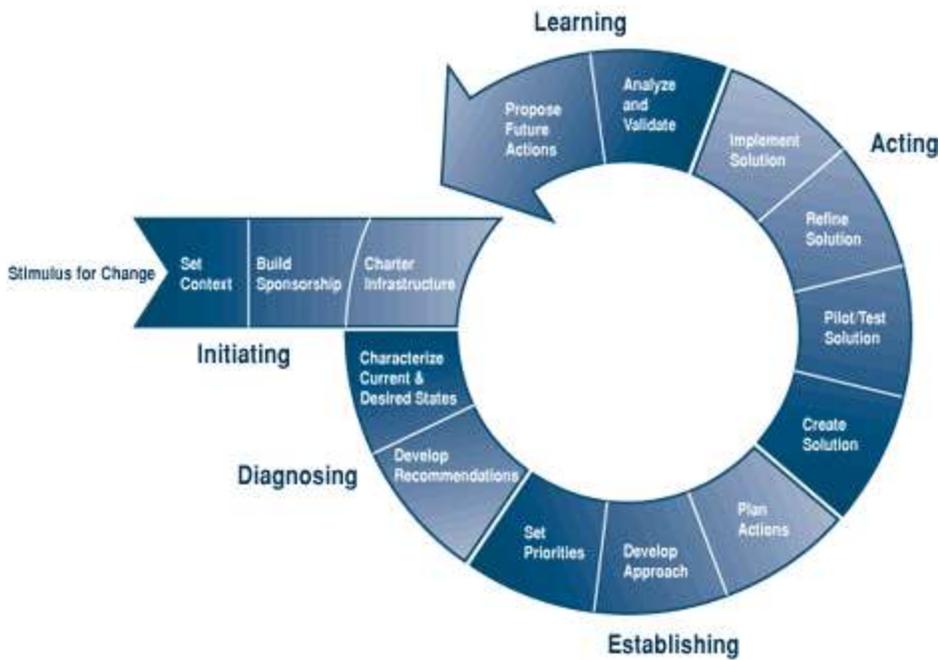


Figure 2: SEI IDEAL Model

Characterizing the current and desired states of the measurement process is similar to identifying the origin and destination of a journey. Characterizing these two states can be done more easily and effectively using a criterion-based reference that catalogues established effective measurement and analysis practices. Herein lies the rationale and motivation for the development of a method for diagnosing the health of an organization’s measurement program.

4.4 A Diagnosis Method for the Measurement Process

Ensuring information quality is a challenge for most organizations—partly because they may not be fully aware of their own data quality levels. Without this information, they cannot know the full business impact of poor or unknown data quality or how to begin addressing it.

The SEI is embarking on an effort to develop the Measurement and Analysis Infrastructure Diagnostic (MAID) method. The intention is to develop a criterion-based assessment method that will allow an organization to evaluate key characteristics of their measurement program (see Figure 3).

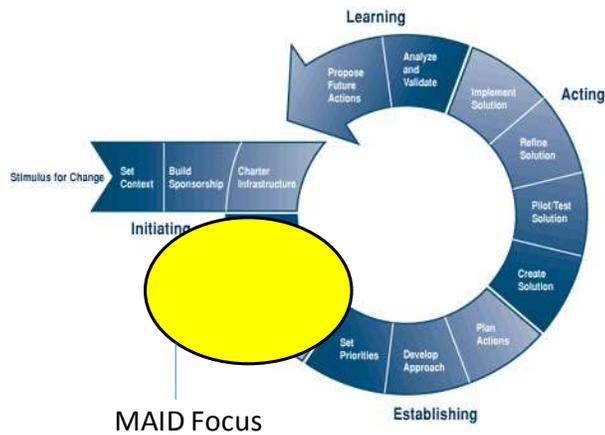


Figure 3: MAID Focuses on the Diagnosing Phase of the IDEAL Improvement Framework

MAID will draw upon, synthesize, and organize well-established best measurement practice guidance that is already embodied in existing models, standards, methods, and applicable best practices (see Figure 4).

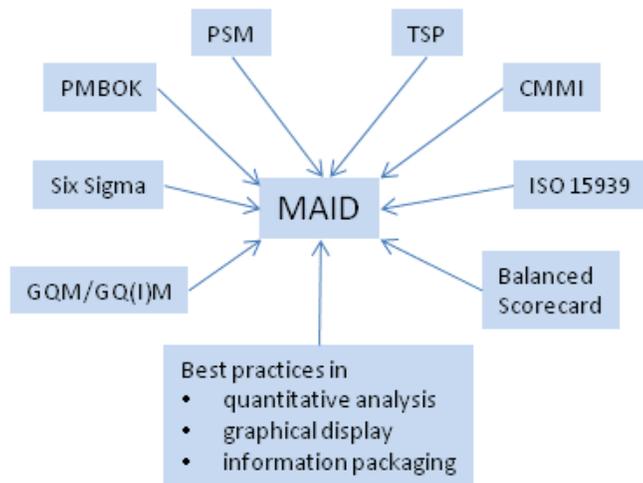


Figure 4: Inputs to be Considered for Development of MAID

Table 1 provides a list of references for the various inputs that will contribute to the MAID assessment criteria.

Table 1: Resources Used as Guidance to Develop MAID

Resource	Acronym [Reference]
The Carnegie Mellon® SEI CMMI® framework	CMMI® [SEI 2006]
Software Engineering - Software Measurement Process	ISO/IEC 15939 [ISO 2002]
Balanced Scorecard Measures	Balanced Scorecard [Kaplan 1992]
Goal-Driven Measurement	GQM/GQ(I)M [Basili 1994, Park 1996]
Six Sigma	Six Sigma [Pyzdek 2003, Breyfogle 2003]
Team Software Process	TSP (see http://www.sei.cmu.edu/tsp/)
Practical Software Measurement	PSM [McGarry 2001]
Project Management Body of Knowledge	PMBOK [PMI 2004]

4.5 Elements of MAID

MAID assessment criteria will be developed to address

- (a) the measurement process (see Figure 1)
- (b) information quality

4.5.1 Assessing the Measurement Process

A set of detailed assessment criteria will be developed to evaluate each stage of the measurement process including

1. measurement planning
2. data collection and storage
3. data processing, information analysis, and interpretation of analysis results
4. measurement information reporting
5. impacts of measurement information

4.5.2 Assessing Information Quality

In addition to measurement process assessment, MAID will also diagnose issues associated with information products that are generated by the measurement process.

It is difficult to manage information quality without understanding the attributes that define it. Wang has described a hierarchy of data quality dimensions [Wang 1995]. This hierarchy is illustrated in Figure 5 and was motivated by the factors in the list that follows.

1. The knowledge worker must be able to get to the data, which means that the data must be *accessible*. That is, the user must have the means and privilege to get the data. In order to be accessible to the user, the data must be available.

® CMM and CMMI are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

2. The knowledge worker must be able to *interpret* the data. That is, the user must understand the *syntax* and *semantics* of the data.
3. The data must be *useful*, meaning the data can be used as an input to the user’s decision-making process. To be useful, the data must be relevant (i.e., fits requirements for making the decision), and *timely*. Timeliness, in turn, can be characterized by currency (when the data item was stored in the database) and *volatility* (how long the item remains valid).
4. The data must be *believable* to the user, to the extent that the user can use the data as a decision input. To be believable, the user may consider, among other factors, that the data be *complete*, *consistent*, *credible*, and *accurate*.

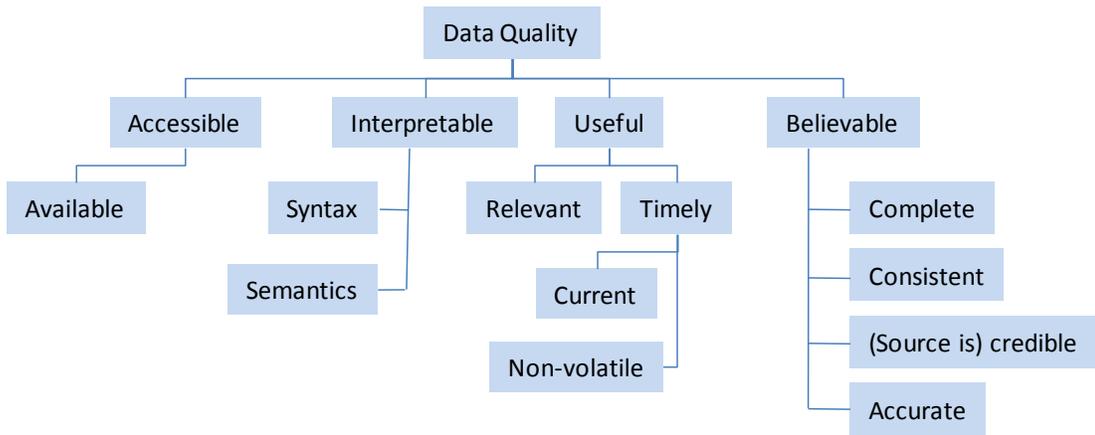


Figure 5: A Hierarchy of Data Quality Dimensions

Information quality is a broad, umbrella term for the accuracy, completeness, consistency, conformity, and timeliness of a particular piece or set of data and for how data enters and flows through the enterprise. Different organizations have different definitions and requirements for data quality, but it ultimately boils down to data that is “fit for purpose.”

4.6 Application Scope – Conducting a MAID Assessment

Figure 6 is a simple depiction of an enterprise that is composed of organizations. These organizations instigate programs or projects and teams to accomplish the business objectives of the enterprise.

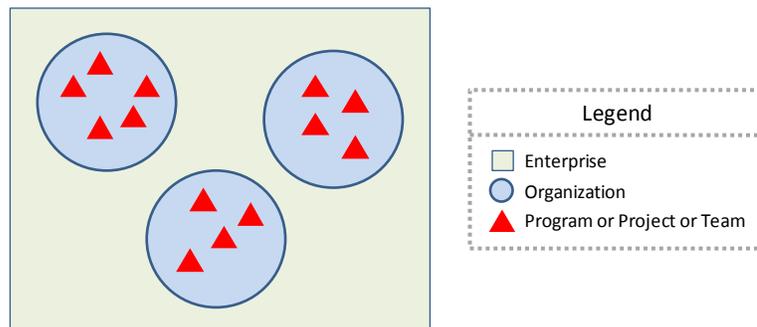
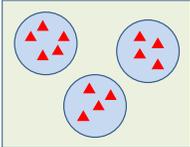


Figure 6: Representation of Enterprise-Based Entities

Several different types of MAID assessments will be available depending on the scope of what is to be assessed. The scope of a MAID assessment is determined during the planning phase of MAID and is based on the desire of the management stakeholders. As illustrated in Table 2, MAID assessments could assess the entire enterprise, one or more organizations within the enterprise, or one or more projects within an organization.

Table 2: Scope of MAID is Determined by Management Stakeholders

MAID Scope	Description
	MAID is conducted to evaluate the measurement program across the enterprise.
	MAID can be scoped to evaluate one or more organizations within the enterprise. When programs, projects, or teams contribute measurement information to the organizational measurements, they become part of the assessment.
	MAID can also be scoped to assess the measurement practices of one or more projects within an organization.

4.7 Looking Ahead

There is a growing need for organizational confidence in measurement data collection, processing, analysis, storage, and interpretation to support decision making. Organizations need to be able to answer the following questions:

- How good is our data?
- Are we doing the right things in terms of measurement and analysis? Are we doing them well?
- How good is the information we generate?
- Are we providing value to the organization and stakeholders?
- How can we improve our measurement processes and improve information quality?
- How can we meet the CMMI-based goals associated with measurement?

To assist organizations with these questions, the SEI is developing MAID. If your organization is interested in piloting the MAID method with the SEI, please contact: customer-relations@sei.cmu.edu.

References

URLs are valid as of the publication date of this document.

[Basili 1994]

Basili, V., Caldiera, G., & Rombach, H. D. "The Goal Question Metric Approach," 528-53. *Encyclopedia of Software Engineering*, John Wiley & Sons, Inc., 1994.

[Breyfogle 2003]

Breyfogle, Forrest W., III. *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*, 2nd ed. John Wiley & Sons, 2003.

[Chrissis 2006]

Chrissis, M. B., Konrad, M., & Shrum, S. *CMMI: Guidelines for Process Integration and Product Improvement*, 2nd ed. New York: Addison-Wesley, 2006.

[Crosby 1979]

Crosby, Philip B. *Quality Is Free*. McGraw-Hill. 1979 (ISBN 0-07-014512-1).

[Deming 1986]

Deming, W. Edwards. *Out of the Crisis*. MIT Center for Advanced Engineering Study, 1986 (ISBN 0-911379-01-0).

[DCARC 2008]

Defense Cost and Resource Center. *Software Resources Data Report*. <http://dcarc.pae.osd.mil/Policy/srdr/index.aspx> (2008).

[DoD 2005]

Department of Defense. *Handbook: Work Breakdown Structures for Defense Materiel Items*. http://www.acq.osd.mil/pm/currentpolicy/wbs/MIL_HDBK-881A/MILHDBK881A/WebHelp3/MILHDBK881A.htm (2005).

[English 2004]

English, Larry P. "Information Quality: Meeting Customer Needs." *Information Impact Newsletter* (June 2004).

[Friedman 2005]

Friedman, Ted, Gassman, Bill, & Newman, David. *Predicts 2006: Emerging Data Management Drivers and Strategic Imperatives*. <http://www.gartner.com/AnalystBiography?authorId=18810> (November 2005). Requires registration and purchase of document.

[ISO 2002]

International Organization for Standardization. *ISO/IEC 15939:2002 -- Software engineering -- Software measurement process*. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=29572 (2002).

[ISO 2005]

International Organization for Standardization. *ISO/IEC TR 19759:2005 -- Software Engineering—Guide to the Software Engineering Body of Knowledge (SWEBOK)*. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33897 (2005).

[Juran 1998]

Juran, Joseph M., co-ed. & Godfrey, Edward G., co-ed. *Juran's Quality Handbook*, 5th ed. McGraw-Hill, 1998.

[Kaplan 1992]

Kaplan, R. S. & Norton D. P. "The Balanced Scorecard Measures That Drive Performance." *Harvard Business Review* (Jan-Feb): 71-80.

[Kasunic 2006]

Kasunic, Mark. *The State of Software Measurement Practice: Results of 2006 Survey*. (CMU/SEI-2006-TR-009). Software Engineering Institute, Carnegie Mellon University, 2006. <http://www.sei.cmu.edu/pub/documents/06.reports/pdf/06tr009.pdf>

[Kasunic 2008]

Kasunic, Mark. *A Data Specification for Software Project Performance Measures: Results of a Collaboration on Performance Measurement* (CMU/SEI-2008-TR-012). Software Engineering Institute, Carnegie Mellon University, 2008. <http://www.sei.cmu.edu/pub/documents/08.reports/08tr012.pdf>

[Krill 2001]

Krill, Paul. "CRM Plagued by Data Quality Issues." <http://www.infoworld.com/articles/fe/xml/01/10/08/011008feedge.html> (2001).

[McGarry 2001]

McGarry John, Card David, Jones Cheryl, Layman Beth, Clark Elizabeth, Dean Joseph & Hall, Fred. *Practical Software Measurement: Objective Information for Decision Makers*. Addison-Wesley Professional, 2001.

[Park 1996]

Park, Robert E., Goethert, Wolfhart B., Florac, William A. *Goal-Driven Software Measurement – A Guidebook*. (CMU/SEI-96-HB-002). Software Engineering Institute, Carnegie Mellon University, 1996. <http://www.sei.cmu.edu/pub/documents/96.reports/pdf/hb002.96.pdf>

[PMI 2004]

Project Management Institute. *A Guide to the Project Management Body of Knowledge, Third Edition*. Project Management Institute, 2004 (ISBN-10: 193069945X, ISBN-13: 978-1930699458).

[Pyzdek 2003]

Pyzdek, Thomas. *The Six Sigma Handbook: The Complete Guide for Greenbelts, Blackbelts, and Managers at All Levels, Revised and Expanded Edition*, 2nd ed. McGraw-Hill, 2003.

[SEI 2006]

SEI, Software Engineering Institute. “Capability Maturity Model Integration (CMMI).” <http://www.sei.cmu.edu/cmmi/> (2006).

[SEI 2008]

SEI, Software Engineering Institute. “The IDEAL Model.” <http://www.sei.cmu.edu/ideal/> (2008).

[TDWI 2008]

TDWI Partners. *TDWI's Data Quality Report*. <http://www.tdwi.org/research/display.aspx?ID=6049> (1998-2008).

[Wang 1995]

Wang, Richard Y., Reddy, M.P., Kon, Henry B. “Toward Quality Data: An Attribute-Based Approach.” *Decision Support Systems*. 13, (1995): 349-373.

REPORT DOCUMENTATION PAGE			<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE November 2008	3. REPORT TYPE AND DATES COVERED Final		
4. TITLE AND SUBTITLE Can You Trust Your Data? Establishing the Need for a Measurement and Analysis Infrastructure Diagnostic		5. FUNDING NUMBERS FA8721-05-C-0003		
6. AUTHOR(S) David Zubrow, James McCurley, & Mark Kasunic				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Software Engineering Institute Carnegie Mellon University Pittsburgh, PA 15213		8. PERFORMING ORGANIZATION REPORT NUMBER CMU/SEI-2008-TN-028		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) HQ ESC/XPK 5 Eglin Street Hanscom AFB, MA 01731-2116		10. SPONSORING/MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES				
12A DISTRIBUTION/AVAILABILITY STATEMENT Unclassified/Unlimited, DTIC, NTIS		12B DISTRIBUTION CODE		
13. ABSTRACT (MAXIMUM 200 WORDS) An organization's measurement and analysis infrastructure directly impacts the quality of the decisions made by people at all organizational levels. Ensuring information quality is a challenge for most organizations—partly because they might not be fully aware of their own data quality levels. Without this information, they cannot know the full business impact of poor or unknown data quality or determine how to begin improving their data. This report describes common errors in measurement and analysis and the need for a criterion-based assessment method that will allow organizations to evaluate key characteristics of their measurement programs.				
14. SUBJECT TERMS Data quality, infrastructure, measurement and analysis		15. NUMBER OF PAGES 31		
16. PRICE CODE				
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	