

IT Metrics and Benchmarking

Use Benchmark Data to Substantiate IT Resource Requirements

Benchmarking enables you to assess IT productivity and demonstrate potential resource savings to executive-level management. The data will make your assertions more compelling and defensible.

Don't Waste Valuable IT Resources on Benchmark Data

Benchmark data looks at where others have been in years past and has no relevance to what you are doing today. The data can easily be manipulated to serve anyone's particular interests.

"Many organizations today are attempting to realize the benefits of IT metrics and benchmarking — some successfully and some not so successfully"

— David Garmus, Guest Editor

Opening Statement

David Garmus

The Big Picture: Software Measurements in Large Corporations

Capers Jones

Extracting Real Value from Process Improvement

Thomas M. Cagley, Jr.

Hitting the Sweet Spot: Metrics Success at AT&T

John Cirone, Patricia Hinerman, and Patrick Rhodes

From Important to Vital: The Evolution of a Metrics Program from Internal to Outsourced Applications

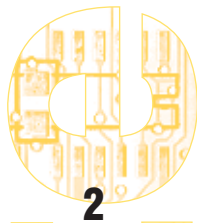
Barbara Beech

Benchmarking for the Rest of Us

Jim Brosseau

The Practical Collection, Acquisition, and Application of Software Metrics

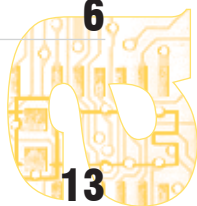
Peter R. Hill



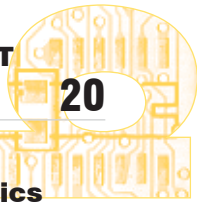
2



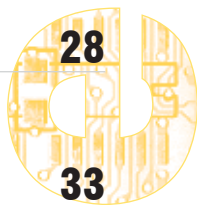
6



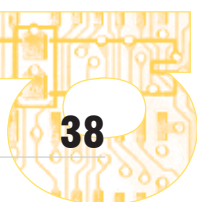
13



20



28



33



38

Benchmarking for the Rest of Us

by Jim Brosseau

To be effective, businesses of all sizes need to understand their own performance. While large, established organizations will typically have a solid infrastructure and a constant finger on their pulse, smaller, growing companies often struggle with fundamental issues. This is highly prevalent in technology organizations, where an entrepreneur with the Next Great Idea suddenly finds that organizational issues are consuming more and more of his or her precious time. It is wise for any organization to have a clear understanding of how well it is performing, either as an impetus to improve or as a basis for understanding how much work it should reasonably take on in the future.

What is the best structure for our organization? How productive should we expect to be? What should we be paying our staff? These and many other questions need to be resolved for small, growing companies to get beyond their nontechnical hurdles to success.

With all these questions to answer and so little time, there is often a rush to quickly find “the solution,” whether a general solution really exists for the industry as a whole or not. Among the frequently asked questions are the following:

What is the appropriate ratio of software testers to developers?

Companies want to use this number to mold the structure of the development organization, but there is no right answer here. I have worked on significant projects in which the developers successfully performed the bulk of the testing of the system, and I have worked with teams where dedicated testers outnumbered developers almost 2:1 and still could not keep up with the issues that were cropping up.

How productive should I expect my team to be (given a variety of factors)?

There is clear benchmarking data available indicating average ratios of function points to lines of code and productivity in terms of lines of code per day, given the type and criticality of the application being developed. There are a great deal of data points behind the scenes used to make the information statistically relevant, usually with extremely wide variations. This variability rarely makes it to the surface of the data presented, but it is a strong indicator that your mileage may vary — and probably by a large amount, even within your team.

How should I compensate my staff? Industry salary reports have dropped from their staggering

heights of a few years ago to reflect the changing times. Still, there is significant geographical variation to consider.

For a variety of reasons, many companies turn to externally generated benchmarking data to provide the answers they need. Unfortunately, there is a dark side to the quick and sometimes blind use of external information. Use of benchmarking data needs to be carefully tempered if it is to provide value for organizational improvement.

BENCHMARKING DATA IS ALLURING

For better or worse, most organizations refer to industry benchmarking data as a means of gauging their performance. Like most people in the industry, I’ve done my fair share of quoting statistics from the Standish Group’s 1994 *CHAOS Report* [6] and used the quarterly reports from the Software Engineering Institute (SEI) to describe industry performance in discussions with clients. A number of people and organizations have collected and disseminated a great deal of benchmarking data over the years, including the guest editor of this issue, David Garmus.

Collected benchmarking data is relatively easy to obtain as, for the most part, it is readily available, if for a price. It generally comes from well established, reputable sources, either published in books or trade journals or available for purchase from a number of organizations worldwide. It is usually well organized and indexed in a manner that will allow you to quickly arrive at the information you are looking for. Using data from reputable sources will help you to back up your assertions and can make your arguments much more compelling and defensible. It can be an indication that you have “done your homework.”

At times, however, the allure of benchmarking data comes from its external sterility. The data provided is based on other people’s performance, and it may provide a sanitized look at what the industry is doing. For some organizations, it can become a game to blithely quote industry performance figures while avoiding internal measurement, knowing that the truth can be a bitter pill to swallow.

BENCHMARKING DATA: CAVEAT EMPTOR

He uses statistics as a drunken man uses lamp-posts — for support rather than illumination.

— Andrew Lang

Imagine a situation in which you decide when it is best to leave for work in the morning by observing your neighbor’s departure patterns. Over the course of a month, his

average departure time is 7:15, give or take five minutes or so. That’s pretty consistent, so you decide that 7:15 must be appropriate for you as well. Unfortunately, your neighbor works about a mile away, while you have a cross-town trek. Worse yet, you may be on the afternoon shift, or you may work from home. Is that benchmarking data worthwhile?

Some organizations blithely quote industry performance figures while avoiding internal measurement, knowing that the truth can be a bitter pill to swallow.

There are a number of problems associated with using the industry data that we have all turned to on occasion. We need to be extremely careful to drill down past the superficial presentation of the data — usually a simplified table or graph — to determine if it is applicable to our situation at all. Quite often, the data will be presented in a form that may be visually compelling while obfuscating some important elements of information that would otherwise be helpful. With the general availability of spreadsheets and graphics packages, we often find ourselves interested in the superficial presentation rather than the intrinsic information.

Wide Variability, Hidden Bias

Beyond the simple data points presented in benchmark data, it is important to recognize that the

underlying data may have potential hidden biases or wide variations within the sample space. These attributes, if not clearly understood, can lead one to rely more heavily on the benchmark data than is reasonable.

Parametric estimation models, for example, are essentially the result of curve-fitting exercises based on a broad sample space of thousands of completed software projects, which can make the models compelling to use for early, whole-project estimates. The SLIM parametric estimation model is based on a large number of projects, divided into roughly a dozen different industry types [3], with the intent to provide a sample space that is relevant to your situation. As you drill deeper, though, you find that the variation within each of these industry types is very wide and that your performance may actually be closer to the median performance of an industry type that does not appear to be close to yours.

The COCOMO II parametric model [1] introduces a bias of another form. While the sample space is much smaller, it is important to note that many of the projects that have been used for curve-fitting the model are primarily in the defense and aerospace realm, where practices are such that there is a very low correlation with commercial software development or other development types.

Both the SLIM and COCOMO II models have been fit primarily with projects that are fairly large in terms of effort and scope. It would

be erroneous to assume that the models could be extrapolated down for use on small projects. To blindly use these models “out of the box” for small projects or projects that have not been calibrated appropriately would be to generate estimates that are falsely defensible. While the data behind the models has been validated, that does not mean that it cleanly maps to your situation.¹

Sparse, Slanted SEI Data

The SEI’s quarterly *Process Maturity Profile of the Software Community* may suffer from bias problems of its own. According to the August 2002 release [5], the report shows that 19.3% of reporting organizations are performing at Level 1 (the Initial level) of the SEI’s Capability Maturity Model (CMM) scale, which is quite a strong positive indicator for the industry as a whole. The fine print, however, indicates that this figure is “based on the most recent assessment, since 1998, of 1,124 organizations.”

There are a couple of points to note here. This sample space is extremely small considering the

number of software development organizations worldwide. In addition, it is biased not only toward organizations that are aware of the SEI, the CMM, and the suggested best practices they promote, but also toward organizations that have reported results to the SEI from formal assessments.

Historical benchmarking data will tell you where the industry has been, but it will not help you resolve how to best address your organizational needs in the future.

In most organizations that I have worked with in the past four years, the majority of the people were not aware of the SEI, and their practices and performance clearly placed them in the Initial level of the CMM. Among those organizations that claimed to have attained a higher level of maturity on the SEI’s scale (i.e., Levels 3-5), most, in my experience, were not able to perform in accordance with even those goals attributed to the Repeatable level (Level 2).

(Ir)Relevance of Annual Data

Often, benchmarking data is provided on an annual basis, which allows you to subscribe and remain current. One must be careful, however, to determine whether time-based trends would be relevant for the information provided. Clearly, annualized reports showing the

equivalent of the average results of 1,000 coin tosses will yield limited additional insight. While some benchmarking data will benefit from annualized updates (such as new data sectors or evolving trends), there are other classes of data that do not benefit to the same degree.

It’s Not a Divining Rod

There is danger in using benchmarking data to determine direction for your organization. Industry averages in IT spending, for example, can be extremely revealing if you are on the receiving end of that spending trend. They can also be used as one of the drivers for forecasting, especially if historical spending trends have tracked well with your performance in the past. If you are looking at how much your organization should be spending, historical benchmarking data will tell you where the industry has been, but it will not help you resolve how to best address your organizational needs in the future. Budgeting for future spending based on industry trends fails to address what is important for you.

BENCHMARKING DATA’S SILVER LINING

All models are wrong.
Some models are useful.

— George Box

All this is not to say that you should never use externally generated benchmarking data within your organization. There is a great deal of consideration and industry

¹Beyond the selection of a specific parametric model for estimation, there is the question of which estimation procedure to use. Many organizations will try to take a published procedure (such as that used by the NASA Software Engineering Laboratory [2]) and its embedded information (such as uncertainty, phases, and approaches) and call it their own. While there are industry-wide principles that an estimation procedure should embrace, there is not a one-size-fits-all solution.

research that has gone into much of the available benchmarking data, and it is important to understand how and whether the data appropriately applies to your situation.

Benchmarking data that is used as a basis for or result of certifications or qualifications — such as ISO quality standards, SEI maturity levels, or the Project Management Institute's Project Management Professional (PMP) designation — provides an indication that the organization or individual has passed a baseline level of performance or understanding. ISO-certified organizations have clearly identified their quality practices and demonstrated that they “practice what they preach” (although this is not a guarantee that their next project will be a success), and the certification can reasonably be used as part of the criteria in an acquisition process. Individuals with the PMP designation have been assessed to have knowledge of a base set of commonly accepted project management best practices and have performed a prescribed amount of work in the project management arena (but this is not a guarantee that they are effective project managers).

For much of the benchmarking data that is available, the underlying assumptions, variability of the data, and inherent biases can usually be identified with some digging. The information may be published along with the primary information that has been distilled, available

from the provided reference information, or obtained through deeper discussion with the data provider if one is so inclined (and diligent).

Some of your greatest insight will come as you track your own measurements over time and observe the variation and trends that are revealed.

THE PERSONALIZED SOLUTION: ANSWER YOUR OWN QUESTIONS FIRST

A reasonable approach to the use of metrics data is to balance external benchmarking data with internally derived data to help you understand whether or not you are achieving your organizational goals. As Peter Senge noted in *The Dance of Change*, we need to measure to learn rather than to merely report [4].

With an understanding that the first step is to identify our goals in the measurement process, we can lean on Vic Basili's GQM approach or extend and elaborate on that practice using techniques such as the balanced scorecard. Identifying these goals and the model we will use to validate the goals allows us to remove biases from the response and resist the temptation to use data simply because it is readily available. Our quests become tightly coupled with our culture and organizational needs.

Using this approach, we can then perform our own internal measurements and make comparisons against industry benchmarks where it makes sense. With the added internal diligence, we will have a more valuable understanding of the biases that are inherent in the data (and a comfort that the biases are more likely to be working for us than against us) and of the uncertainty or variability in the data set.

It is important to recognize the distinction between statistical variability across industry benchmarking data and individual performance variability that will arise in your own measured data. The former is an indication of the relative applicability of the information to your situation, while the latter is an expected artifact of the measurement approach that needs to be fully appreciated. You need to accept individual variation as a fact of life. Even if you use the information to cull the low-performing individuals (not a recommended practice), you will continue to have variability; by definition, 50% of the people will always fall below the median of your data set. It is dangerous to fall into the trap of using measures for segregating the team rather than for improving the organization.

Some of your greatest insight will come as you track your own measurements over time and observe the variation and trends that are revealed. This information is not something that can be gained from industry benchmarking data, but

you can see whether you are tracking toward or away from the industry data, which will provide a greater indication of the applicability of the benchmarking information to your situation.

For this tracking to be effective, you need to be consistent in your measurement approaches within your organization over time. One commonly hears concerns in the industry about inconsistency of measurement, whether it be for histogram categories to collect data or the approach used (such as the highly variable lines of code measure, for example). The bottom line here is that you should select a specific approach, identify that it is the standard, and stick with it in order to ensure that you are indeed making apples-to-apples comparisons.

Industry benchmarking data definitely has its place in your arsenal of information for making strategic business decisions. Still, it has limitations that must be overcome with a deep understanding of why you are measuring and balanced with data gathered internally with reasonable approaches. Taken with a grain of salt, benchmarking information can give us the perspective we need to better understand what our internal information is telling us.

REFERENCES

1. Boehm, Barry, Bradford Clark, Ellis Horowitz, Ray Madachy, Richard Shelby, and Chris Westland. "Cost Models for Future Software Life Cycle Processes: COCOMO 2.0." *Annals of Software Engineering* (1995) (<http://sunset.usc.edu/research/COCOMOII/index.html>).
2. National Aeronautics and Space Administration. *The Manager's Handbook for Software Development, Revision 1* (Software Engineering Laboratory Series SEL-84-101). NASA, 1990 (<http://sel.gsfc.nasa.gov/website/documents/online-doc/84-101.pdf>).
3. Putnam, Lawrence, and Ware Myers. *Measures for Excellence: Reliable Software on Time, Within Budget*. Yourdon Press Computing Series, Pearson Education POD, 1992.
4. Senge, Peter, Art Kleiner, Charlotte Roberts, George Roth, Rick Ross, and Bryan Smith. *The Dance of Change: The Challenges to Sustaining Momentum in Learning Organizations*. Currency/Doubleday, 1999.
5. Software Engineering Institute. *Process Maturity Profile of the Software Community 2002 Mid-Year Update*. SEI, August 2002.
6. The Standish Group. *The CHAOS Report*. The Standish Group, 1994 (www.standishgroup.com/sample_research/chaos_1994_1.php).

Jim Brosseau has 20 years' experience in the software industry in a wide variety of roles, application platforms, and domains. A common thread through his experience has been a drive to find a less painful approach to software development. Mr. Brosseau has worked in quality assurance at Canadian Marconi and was involved in the development and management of the test infrastructure used to support the Canadian Automated Air Traffic System. He is Principal of the Clarrus Consulting Group in Vancouver, Canada, and in the past four years, he has consulted with numerous organizations throughout North America, specifically to improve their development practices.

Mr. Brosseau publishes the Clarrus Compendium, a free weekly newsletter with a unique perspective on the software industry (www.clarrus.com/resources.htm). He has been published in PM Network magazine, the PMI GovSIG magazine, and the SEA Software Journal, and he has made presentations at Comdex West, PSQT North, the New Brunswick SPIN group, and several local associations.

Mr. Brosseau can be reached at Clarrus Consulting Group Inc., 7770 Elford Avenue, Burnaby, BC Canada V3N 4B7. Tel: +1 604 540 6718; Fax: +1 604 648 9534; E-mail: jim.brosseau@clarrus.com.

Cutter Benchmark Review

Analyzing IT Metrics for Informed Management Decisions

Cutter Benchmark Review helps you support — and even profit from — your organization's IT initiatives and programs.

SPECIAL OFFER!

Get **Cutter Benchmark Review** at the Charter Rate when you order by July 31st!

Simply complete and return the coupon below or call +1 781 648 8700, or fax +1 781 648 1950.

You can also order online at www.cutter.com or by sending e-mail to service@cutter.com.

Cutter Benchmark Review delivers the survey-based statistics you need to keep your IT goals on track.

SUBSCRIBE NOW AT THE CHARTER RATE!

Information and analysis are the foundation for performance management systems. However, few companies have built a solid performance measurement process that aligns measures throughout the organization. Entities that have implemented sound measurements and tie them to organizational strategy are enabling their managers and process owners to focus on strategic, rather than day-to-day, issues. In turn, better strategic decisions are being made based on those measures.

Cutter Consortium constantly conducts surveys and studies of IT and business practices worldwide, designed to deliver valuable insight and intelligence into how organizations are using IT to become more efficient and profitable. **Cutter Benchmark Review** is a succinct distillation of these findings that provides every member of your organization with the key data and analysis necessary to build and maintain a successful performance management system — and use the system to make more effective decisions.

The undiluted business intelligence provided by **Cutter Benchmark Review** is simply not available anywhere else. The cost in time and resources to complete this kind of research internally would be prohibitive — but you can get it today and for the next full year for about \$20 a month.

Take a few moments now to ensure that you have unfettered access to critical statistical analysis of important IT initiatives and programs by completing the order form below and becoming a subscriber to **Cutter Benchmark Review**. We're sure once you have this publication in your hands, you won't want to give it up!

YES! Please start my one-year subscription to **Cutter Benchmark Review** for the Charter Rate of \$195 (US \$255 outside N. America)

**CUTTER
CONSORTIUM**

Name/Company _____

Payment or P.O. enclosed.

Priority Code: 220*2CIT

Address/P.O. Box _____

Charge Mastercard, Visa, AmEx, Diners Club.
(Charge will appear as Cutter Consortium.)

City _____ State/Province _____

Card # _____

ZIP/Postal Code _____ Country _____

Expiration Date _____

Telephone _____ Fax _____

Signature _____

E-Mail _____

Mail to Cutter Consortium, 37 Broadway, Suite 1, Arlington, MA 02474-5552, USA, or send a fax to +1 781 648 1950.
Or call +1 781 648 8700 or send e-mail to service@cutter.com. Web site: www.cutter.com

Cutter IT Journal

Topic Index

- June 2003 IT Metrics and Benchmarking
- May 2003 Is Open Source Ready for Prime Time?
- April 2003 Project Portfolio Management:
Blueprint for Efficiency or Formula for Boondoggle?
- March 2003 Critical Chain Project Management:
Coming to a Radar Screen Near You!
- February 2003 XP and Culture Change: Part II
- January 2003 Ending "Garbage In, Garbage Out":
IT's Role in Improving Data Quality
- December 2002 Preventing IT Burnout
- November 2002 Globalization: Boon or Bane?
- October 2002 Whither Wireless?
- September 2002 XP and Culture Change
- August 2002 Plotting a Testing Course in the IT Universe
- July 2002 Confronting Complexity: Contemporary Software Testing
- June 2002 B2B Collaboration: Where to Start?



David Garmus, Guest Editor

David Garmus is a Principal in The David Consulting Group and an acknowledged authority in the sizing, measurement, and estimation of software application development and maintenance. He has helped numerous CIOs and CFOs successfully manage expectations in software development projects, using function point analysis to enable effective IT cost management and achieve a realistic return on investment. He is the coauthor of *Function Point Analysis: Measurement Practices for Successful Software Projects* and *Measuring the Software Process: A Practical Guide to Functional Measurements*. He has served as the President of the International Function Point Users Group (IFPUG).

Mr. Garmus has more than 30 years of experience managing software development and maintenance and has taught college-level courses in computing- and finance-related subjects. He is a member of Project Management Institute, the Quality Assurance Institute, and the IEEE Computer Society, and he holds a BS from the University of California-Los Angeles and an MBA from the Harvard Business School. He can be reached at dg@davidconsultinggroup.com.

Upcoming

Issue Themes

Enterprise Architecture Governance

The New CIO Agenda

Usability

Patterns

Killing IT Projects

<http://www.cutter.com/> or +1 800 964 5118

