

# **SOURCES OF SOFTWARE BENCHMARKS**

**Version 13**

**November 7, 2011**

**Capers Jones & Associates LLC.**

## **INTRODUCTION**

**Number of benchmark sources currently: 17**

**Number of projects in all benchmarks sources: 85,100**

Quantitative software benchmark data is valuable for measuring process improvement programs, for calibrating software estimating tools, and for improving software quality levels. It is also useful for studies of industry and company progress over time.

There are a number of organizations that gather and report quantitative benchmark information. However these organizations are independent and in fact some of them are competitors.

This catalog of software benchmark data sources is produced as a public service for the software community.

There are many different kinds of benchmarks including productivity and quality levels for specific projects; portfolio benchmarks for large numbers of projects, operational benchmarks for data center performance; security benchmarks, compensation and staffing benchmarks for human resource purposes; and software customer satisfaction benchmarks.

This catalog is intended to grow over time and to include all major sources of software benchmark information.

The information in this catalog is provided by the benchmark groups themselves and includes topics that the benchmark groups wish to be made available.

In this version the benchmark groups are listed in alphabetical order. In later versions the groups will be listed alphabetically for each type of benchmark.

## TABLE OF CONTENTS

Introduction to Software Benchmark Sources	3
1. Capers Jones & Associates LLC	4
2. CAST	9
3. COSMIC Consortium	12
4. David Consulting Group (DCG)	14
5. Galorath Incorporated	16
6. International Software Benchmark Standards Group (ISBSG)	18
7. Jerry Luftman (Stevens Institute of Technology)	20
8. Price Systems LLC	23
9. Process Fusion	26
10. Quantimetries	27
11. Quantitative Software Management (QSM)	30
12. Q/P Management Group	33
13. RCBS, Inc.	36
14. Reifer Consultants LLC	39
15. Software Benchmarking Organization	41
16. Software Improvement Group (SIG)	43
17. Test Maturity Model Integrated (TMMI) by Geoff Thompson	46
Appendix A: Survey of Software Benchmark Usage and Interest	48
Appendix B: A New Form of Software Benchmark	50

## INTRODUCTION TO SOFTWARE BENCHMARK SOURCES

The software industry does not have a good reputation for achieving acceptable levels of quality. Neither does the industry have a good reputation for schedule adherence, cost control, or achieving high quality levels.

One reason for these problems is a chronic shortage of solid empirical data about quality, productivity, schedules, costs, and how these results vary based on development methods, tools, and programming languages.

A number of companies, non-profit groups, and universities are attempting to collect quantitative benchmark data and make it available to clients or through publication. This catalog of benchmark sources has been created to alert software engineers, software managers, and executives to the kinds of benchmark data that is currently available.

The information in this catalog is provided by the benchmark groups themselves, and shows what they wish to make available to clients.

This catalog is not copyrighted and can be distributed or reproduced at will. If any organization that creates benchmark data would like to be included, please write a description of your benchmark data using a format similar to the formats already in the catalog. Please submit new benchmark information (or changes to current information) to Capers Jones & Associates LLC via email. The email address is [Capers.Jones3@gmail.com](mailto:Capers.Jones3@gmail.com).

The catalog is expected to grow as new sources of benchmark data provide inputs. Benchmark organizations from every country and every industry are invited to provide information about their benchmark data and services.

## **Capers Jones & Associates LLC**

Web site URL: Under construction

Email: [Capers.Jones3@gmail.com](mailto:Capers.Jones3@gmail.com)

**Sources of data:** Primarily on-site interviews of software projects. Much of the data is collected under non-disclosure agreements. Some self-reported data is included from Capers Jones studies while working at IBM and ITT corporations. Additional self-reported data from clients taught by Capers Jones and permitted to use assessment and benchmark questionnaires.

**Data metrics:** Productivity data is expressed in terms of function point metrics as defined by the International Function Point User's Group (IFPUG). Quality data is expressed in terms of defects per function point.

Also collected is data on defect potentials, defect removal efficiency, delivered defects, and customer defect reports at 90 day and 12 month intervals.

Long-range data over a period of years is collected from a small group of clients to study total cost of ownership (TCO) and cost of quality (COQ). Internal data from IBM also used for long-range studies due to author's 12 year period at IBM..

At the request of specific clients some data is converted into COSMIC function points, use-case points, story points, or other metrics.

**Data usage:** Data is used to create software estimating tools and predictive models for risk analysis. Data is also published in a number of books including The Economics of Software Quality, Software Engineering Best Practices, Applied Software Measurement, Estimating Software Costs and 12 others. Data has also been published in about 200 journal articles and monographs.

Data is provided to specific clients of assessment, baseline, and benchmark studies. These studies compare clients against similar companies in the same industry.

Data from Capers Jones is frequently cited in software litigation for breach of contract lawsuits or for suits alleging poor quality. Some data is also used in tax litigation dealing with the value of software assets.

**Data availability:** Data is provided to clients of assessment and benchmark studies. General data is published in books and journal articles. Samples of data and some reports are available upon request.

Some data and reports are made available through the library, Webinars, and seminars offered by the Information Technology Metrics and Productivity Institute (ITMPI.org).

**Kinds of data:** Software productivity levels and software quality levels for projects ranging from 10 to 200,000 function points. Data is primarily for individual software projects, but some portfolio data is also collected. Data also supports activity-based costing down to the level of 40 activities for development and 25 activities for maintenance. Agile data is collected for individual sprints. Unlike most Agile data collections function points are used for both productivity and quality.

Some data comes from commissioned studies such as an Air Force contract to evaluate the effectiveness of the CMMI and from an AT&T study to identify occupations employed within large software labs and development groups.

**Volume of data:** About 13,500 projects from 1978 through today. New data is added monthly. Old data is retained, which allows long-range studies at 5 and 10-year intervals. New data is received at between 5 and 10 projects per month from client interviews.

**Industry data:** Data from systems and embedded software, military software, commercial software, IT projects, civilian government projects, and outsourced projects.

Industries include banking, insurance, manufacturing, telecommunications, medical equipment, aerospace, defense, and government at both state and national levels.

Data is collected primarily from large organizations with more than 500 software personnel. Little data from small companies due to the fact that data collection is on-site and fee based.

Little or no data from the computer game industry or the entertainment industry. Little data from open-source organizations.

**Methodology data:** Data is collected for a variety of methodologies including

Agile, waterfall, Rational Unified Process (RUP), Team Software Process, (TSP), Extreme Programming (XP), and hybrid methods that combine features of several methods.

Some data is collected on the impact of Six Sigma, Quality Function Deployment (QFD), formal inspections, Joint Application Design (JAD), static analysis, and 40 kinds of testing.

Data is also collected for the five levels of the Capability Maturity Model Integrated (CMMI™) of the Software Engineering Institute.

**Language data:**

As is usual with large collections of data a variety of programming languages are included. The number of languages per application ranges from 1 to 15, with an Average of about 2.5. Most common combinations include COBOL and SQL and Java and HTML. Specific languages include Ada, Algol, APL, ASP Net, BLISS, C, C++, C#, CHILL, CORAL, Jovial, PL/I and many Derivatives, Objective-C, Jovial, and Visual Basic.. More than 150 languages out of a world total of 2,500 are included.

**Country data:**

About 80% of the data is from the U.S. Substantial data from Japan, United Kingdom, Germany, France, Norway, Denmark, Belgium, and other major European countries. Some data from Australia, South Korea, Thailand, Spain, and Malaysia.

Little or no data from Russia, South America, Central America, China, India, South East Asia, or the Middle East.

**Unique data:**

Due to special studies Capers Jones data includes information on more than 90 software occupation groups and more than 100 kinds of documents produced for large software projects. Also, the data supports activity-based cost studies down to the levels of 40 development activities and 25 maintenance tasks. Also included are data on the defect removal efficiency levels of 65 kinds of inspection, static analysis, and test stages.

Some of the test data on unit testing and desk checking came from volunteers who agreed to record information that is normally invisible and unreported. When working as a programmer Capers Jones was such a volunteer.

From longitudinal studies during development and after release the Jones data also shows the rate at which software requirements grow and change during development and after release. Monthly change rates exceed 1% per calendar month during development and more than 8% per year after release.

From working as an expert witness in 15 lawsuits, some special data is available on litigation costs for plaintiffs and defendants.

From on-site data collection and carrying out interviews with project teams and then comparing the results to Corporate resource tracking systems, it has been noted that “leakage” or missing data is endemic and approximates 50% of actual software effort. Unpaid overtime and the work of managers and part-time specialists are most common.

Quality data also leaks and omits more than 70% of internal defects. Most common omissions are those of desk checking, unit testing, static analysis, and all defect removal activities prior to release.

Leakage from both productivity and quality data bases inside Corporations makes it difficult to calibrate estimating tools and also causes alarm to higher executives when the gaps are revealed. The best solution for leakage is activity-based cost collection.

**Future data:**

There are several critical areas which lack good sources of quantitative data. These include studies of data quality, studies of intangible value, and studies of multi-national projects with geographically distributed development locations.

**Summary:**

Capers Jones has been collecting software data since working for IBM in 1978. In 1984 he founded Software Productivity Research and continued to collect data via SPR until 2000. Capers Jones & Associates LLC was formed in 2001.

He owns several proprietary data collection questionnaires that include both qualitative assessment information and quantitative data on productivity and quality. The majority of data comes from on-site interviews with software project teams but self-reported data is also included, especially from clients who have been trained and authorized to use the Jones questionnaires.

More recently remote data collection has been carried

out via Skype and telephone conference calls using shorter forms of the data collection questionnaires.

Some self-reported or client-reported benchmark data is included from companies taught by Capers Jones and from consortium members.

Some self-reported data is also included from internal studies carried out while at IBM and ITT, and also from clients such as AT&T, Siemens, NASA, the Navy, and the like.

## CAST

Web site URL: <http://www.castsoftware.com/Product/Appmarq.aspx>

Email: [info@castsoftware.com](mailto:info@castsoftware.com)

**Sources of data:** Appmarq is a repository of structural quality data for custom software applications in business IT. Data is collected via automated analyses with the CAST Application Intelligence Platform (AIP), which performs a thorough structural quality analysis at the code and whole-application level. Metrics from the application-level database are fed into the central Appmarq repository. All data is made anonymous and normalized before entering the central benchmarking database.

The AIP data are combined with application “demographics,” which are the qualitative application characteristics such as age, business function, industry, and sourcing paradigm. These demographics are collected directly from the customer via survey instrument and provide a means to identify peer applications when benchmarking.

**Data metrics:** The data represents software structural quality metrics, which, at their highest level, include:

- Business risk exposure (performance, security, robustness)
- Cost efficiency (transferability, changeability, maintainability)
- Methodology maturity (architecture, documentation, programming standards)
- Application Size (KLOC, backfired function points)
- Application Complexity (cyclomatic complexity, SQL complexity)
- Rule level details (specific rules being violated)
- Demographics (industry, functional domain, extent of in-house/outsource, extent of onshore/offshore, age of application, number of releases, methodology and certifications)

**Data usage:** The data collected is used for research on trends in the structural quality of business applications, as well as best practices and standards development. A detailed research paper with industry-relevant findings is published each year. Some initial discussions are starting with academia to use Appmarq data for scholarly research.

Data is also provided to specific clients in the form of customized reports, which from a measurement baseline, benchmarks the

structural quality of their applications against those of their industry peers and of the same technology.

**Data availability:** Data is provided to clients of assessment and benchmark studies. General data is published in a yearly report. A summary of key findings is made available and distributed across a large number of organizations.

**Volume of data:** Data has been collected over a period of 4 years. The dataset currently stands at more than 800 distinct applications. Data is continually added to the dataset, as new benchmarks are conducted and/or data automatically extracted from the CAST AIP repository.

**Industry data:** Data is collected primarily from large IT-intensive companies in both private and public sectors. Industries include Finance, Insurance, Telecommunications, Manufacturing, Transportation, Retail, Utilities, Pharmaceuticals and Public Administration.

**Methodology data:** Data is collected from a variety of methodologies including agile/iterative, waterfall and hybrid methods that combine several methods. Data is also collected for the five levels of the Capability Maturity Model Integrated (CMMI™) of the Software Engineering Institute (SEI).

**Technology data:** Data is collected from a wide range of technologies, including Java EE, .NET, C/C++, SAP, Oracle, Mainframe technologies and other technologies commonly found in IT.

**Country data:** A quarter of the data comes from North America, the remainder is distributed across UK, France, Belgium, Italy, Germany, Spain, India and several other countries.

**Summary:** Starting in 2007, CAST has been collecting metrics and structural characteristics from custom applications deployed by large IT-focused enterprises across North America, Europe and India. Appmarq is a structural quality benchmarking repository with the purpose of analyzing the trends in the structural quality of business applications regarding characteristics such as robustness, performance, security, and changeability.

Unlike other benchmarking datasets, Appmarq is built on automated analysis conducted and normalized by machine; hence the data are an apples-to-apples comparison across different environments. Also unique is that Appmarq puts a singular focus on the characteristics of the software product, rather than the

process, project and performance level metrics that feed into the actual product.

## **COSMIC (the Common Software Measurement International Consortium)**

Web site URL: [www.cosmicon.com](http://www.cosmicon.com)

Email: [harold.van.heeringen@sogeti.nl](mailto:harold.van.heeringen@sogeti.nl); [luca.santillo@gmail.com](mailto:luca.santillo@gmail.com);

**Sources of data:** Voluntary submission of software project data. All COSMIC-measured project data should be submitted to the ISBSG repository (see the ISBSG entry in this list of benchmark sources). – [www.isbsg.org](http://www.isbsg.org) – which provides a service to all COSMIC method users. The data is collected under an agreement of anonymity.

**Data metrics:** Productivity, speed of delivery and quality data where COSMIC-measured software sizes are used as the measure of project work-output.

**Data usage:** Data has been used to develop COSMIC method benchmarks and has been analyzed in many ways to enable performance comparisons and for project estimating. A report entitled ‘The Performance of Real-Time, Business Application and Component Software Projects’ is available from [www.isbsg.org](http://www.isbsg.org). A version showing the report contents, but not the actual numbers, is available for free download from [www.cosmicon.com](http://www.cosmicon.com).

**Data availability:** The ISBSG data is ‘open’ and can be licensed by anyone in its ‘raw’ form at a moderate cost.

Data is provided to academics for use in approved research work.

**Kinds of data:** Data on both Development and Enhancement projects, for business application, real-time and component software projects. Projects range in size from 8 to 1,800 COSMIC Function Points.

**Volume of data:** Most data is for projects completed within the last five years. The number of projects is expected to pass 500 in the course of 2011.

**Industry data:** Business application software project data comes from Banking, Government and Public Administration, Insurance, Engineering, Medical & Healthcare, Retail & Wholesale industries.

Real-time software project data comes from Automotive, Telecommunications, Computer and Software industries.

**Methodology data:** See the entry for the ISBSG in this list of benchmarking services

**Country data:** See the entry for the ISBSG in this list of benchmarking services. The country of origin of any project data is not revealed.

**Summary:**

The COSMIC Functional Size Measurement method is the only such method developed by an international group of software metrics experts, designed to conform to fundamental software engineering principles, and to be applicable to business application, real-time and infrastructure software. It can be used in any layer of a software architecture and at any level of decomposition. Use of the COSMIC method for performance measurement, benchmarking and estimating is therefore expected to enable much higher degrees of precision than when using older software size measures. For more, see [www.cosmicon.com](http://www.cosmicon.com) .

## **David Consulting Group (DCG)**

Web site URL: [www.davidconsultinggroup.com](http://www.davidconsultinggroup.com)

Email: [info@davidconsultinggroup.com](mailto:info@davidconsultinggroup.com)

**Sources of data:** As part of many of DCG's software measurement and software sizing assignments, we collect non-proprietary data for project size, effort, duration, resources, defects and cost. All data collection activities are quality verified by DCG Measurement Specialists. All DCG counts are performed by IFPUG Certified Function Point Specialists (CFPS).

**Data metrics:** All collected data is stored in an internal, proprietary, confidential database. All data is expressed (normalized) in terms of function point metrics as defined by the International Function Point User's Group (IFPUG).

DCG Measurement Specialists extract selected Data Views of four key metrics:

1. Size (Function Points)
2. Quality (Defects)
3. Duration (Project Months)
4. Productivity (Development Hours per Function Point)

For each metric view, DCG provides an equation (formula) based upon a statistical analysis and the number of data points represented across the dimensions and attributes of the benchmark.

**Data usage:** The primary use of the DCG Industry Database is to support our benchmarking consulting engagements. The Database allows us to work with customers to analyze the data to give them realistic comparisons with their own internal performance.

**Data availability:** DCG data is not available for sale in its raw state. Aggregate DCG Data Views are available for sale as part of a client benchmarking consulting engagements.

**Kinds of data:** DCG collects size, effort, duration and quality data. At least two of these attributes are available for all projects in the database. Sizing data is expressed in IFPUG function points.

**Volume of data:** DCG's repository contains several thousand projects. Most client benchmarks use a subset of all available projects because of the specificity of the comparison requested. The nature of the use of our data requires DCG to prioritize the quality of the data and sources of the data ahead of the volume of data.

**Industry data:** Data is collected from DCG records of software development or application maintenance project activity. Projects are typically performed for the software development or application maintenance divisions of major corporations and for software product companies of all sizes.

Industries include telecommunication; insurance; finance; manufacturing; technology; Government; and public sector.

**Methodology data:** Data is attributed along four key dimensions with sub-attributes within each dimension. The dimensions include Technology, Platform, SDLC and Application Development Type (e.g., new development or enhancement).

**Country data:** A majority of the data comes from North American companies with the remainder, in order, from Europe, India and Asia. However, country origin data is not stored in the database.

**Summary:** The primary use of the DCG Industry Database is to support our benchmarking consulting engagements. The Database allows us to work with customers to analyze the data to give them realistic comparisons with their own internal performance.

The nature of the use of our data requires DCG to prioritize the quality of the data and sources of the data ahead of the volume of data.

## **Galorath Incorporated**

Web site URL: [www.galorath.com](http://www.galorath.com)

Email: [info@galorath.com](mailto:info@galorath.com)

- Sources of data:** Repositories obtained through product customers, industry sources, and public domain and consulting. Galorath also maintains a partnership with ISBSG and offers the ISBSG data in SEER Historical Database (SEER-HD) format to subscribers.
- Data metrics:** Productivity expressed in function points and source lines of code. Most data has language, platform and application descriptors. Some of the data also contains, dates, defects delivered, time phased staffing, documentation pages, detailed SEER parameter settings.
- Data usage:** Data is used to calibrate SEER knowledge bases, determine language factors, generate trend lines, benchmark organizations, and validate model relationships.
- Data availability:** Data is delivered in the form of knowledge bases, language factors, trend lines and estimating relationships to SEER customers. Where permitted data is provided in source form so SEER users can see their estimates compared with individual data points. SEER customers who submit data to Galorath receive their data in the SEER Historical Database format that can be used directly by the application.
- Kinds of data:** Records of completed projects at the program and project level. Size ranges from extremely small (less than one month duration) to very large (hundreds of staff, many years' duration.) Additionally metadata is available from some sources, showing items such as productivity ranges, size ranges, etc. where management is not willing to release raw data.
- Volume of data:** Over 20,000 records with nearly 13,000 total records containing at least size, in addition to other metrics of which approximately 8500 records contain completed effort and a lesser number, project duration. Volume increases regularly.
- Industry data:** Data from IT/business systems, industrial systems and embedded software, military, and commercial enterprises.
- Methodology data:** Project processes in collected data range from traditional waterfall through incremental and agile development.

**Country data:** Military and aerospace data is US and European. Commercial data is worldwide.

**Summary:** Galorath maintains an active data collection program and has done so for the last 20 years. It makes data available to users of SEER for Software in several forms such as its ProjectMiner data mining tool and Metrics and Benchmarking data visualization tool.

## **ISBSG Limited**

Web site URL: [www.isbsg.org](http://www.isbsg.org)

Email: [peter.r.hill@isbsg.org](mailto:peter.r.hill@isbsg.org)

- Sources of data:** Voluntary submission of software project data. The data is collected under an agreement of anonymity. Data is submitted either directly or by consultants with the permission of their customers.
- Data metrics:** Productivity data is expressed in terms of function point metrics as defined by the International Function Point User's Group (IFPUG); NESMA; COSMIC; and FiSMA. Quality data is expressed in terms of defects per function point or functional unit.
- LOC sizes are stored if provided but not validated and not used for any analysis. Likewise for other sizing methods, e.g. Feature Points, Use Case Points etc.
- Also collected is data on the report of defect occurrence during the project and within the first 30 days of operation.
- Data has come from more 25 countries and 20 major organization types
- Data usage:** Data is used with some basic ISBSG software estimating tools; used by commercial estimation tool vendors; and for benchmarking services. The data is also analyzed and published in a number of analysis reports and books including *Practical Software Project Estimation*.
- Data is provided to academics for use in approved research work.
- Data availability:** The ISBSG data is 'open' and can be licensed by anyone in its 'raw' form at a moderate cost. Web portal access to the data will be available from January 2012.
- General data and analysis results are published in books, analysis reports and journal articles. Samples of data and some reports are available upon request.
- Kinds of data:** Data on both Development and Enhancement projects. Projects ranging from 10 to 10,000 function points in size, but with an average size around 300FP.
- Data is primarily for individual software projects, but some portfolio data is also collected.

**Volume of data:** About 6,000 projects from 1989 through today. New data is added monthly. Old data is retained, allowing studies over time.

**Industry data:** Data from MIS commercial software; real-time systems; IT projects; government projects; outsourced projects; and Off-shore projects.

Industries include banking; Insurance; Legal; Manufacturing; Telecommunications; Accounting; Sales; Transport; Government; and Public Sector.

Data is submitted from a wide range of organizations of differing sizes. There is no data from the military.

**Methodology data:** Data is collected for a variety of methodologies including Agile, Waterfall, Joint Application Design (JAD), Rational Unified Process (RUP), Team Software Process, (TSP), and hybrid methods that combine features of several methods.

Some data is collected on the impact of being compliant with CMM and CMMI and relevant ISO standards.

**Country data:** About 30% of the data is from the U.S. 16% from Japan; 16% Australia; 10% Finland; 8% Netherlands; 6% India; 5% Canada; 2% China. Also data from 12 other countries.

**Summary:** The ISBSG's formation in 1997 was built upon several years of previous cooperation by a group of national software metrics associations with the common aim of developing and promoting the use of IT industry history data to improve software processes and products, for the benefit of both businesses and governments worldwide.

Current members of the ISBSG represent IT and Metrics Associations based in the following countries: Australia, China, Finland, Germany, India, Italy, Japan, Korea, Netherlands, Spain, Switzerland, UK, USA.

Two independent repositories of IT industry data:

1. Software Development and Enhancement – over 5,700 projects
2. Software Maintenance and Support – ~500 applications

**Jerry Luftman Ph.D. Stevens Institute of Technology; & Luftman LLC**

Web site URL: <http://howe.stevens.edu/index.php?id=14>

Email: [luftman@hotmail.com](mailto:luftman@hotmail.com)

Analytic: IT Business Alignment Maturity

**Sources of data:** The repository is comprised of data from over one-third of the Global 1,000 companies; albeit the assessment has been used by companies of all sizes. The data is captured primarily via on-site interviews of IT and business executives; in cases where a larger sample size is requested, online survey is available. The data is collected under non-disclosure agreements. While most of the data has been collected via Luftman, consulting companies and advanced graduate students from around the globe have participated; hence permission to use the assessment and benchmark is available.

A similar maturity assessment is also available to benchmark the alignment maturity of clients and service providers.

**Data metrics:** The maturity of the relationship of IT and business organizations is based on an assessment of 6 components: Communications, Value Analytics, Governance, Partnership, Infrastructure, and Human Resources. A 5 point Likert scale (much like CMMI) is applied to the different elements comprising the 6 components. While an overall alignment maturity score is derived, each of the components and elements also have scores that are benchmarked.

**Data usage:** Data is used to provide not just a description of where an organization is strong/weak/disagreement, but the model provides a prescriptive roadmap to identify specific initiatives to help improve the relationship among the participating organizations.

A report is provided that describes the identified strengths and opportunities for improvement, specific recommendations for addressing them, and a benchmark comparing the company to things like how they compare with other companies from the same industry and geography, as well as comparing how their C level executives compare to their peers.

Data from Luftman's research is frequently cited in academic and business publications.

**Data availability:** Data is provided to clients of assessment and benchmark studies. General data is published in books and journal articles. Samples of data and some reports are available upon request.

- Kinds of data:** Data pertaining to the respective elements under each of the six components is extremely insightful. Identifying the relationship of alignment maturity and firm performance, as well as identifying appropriate IT organizational structure and IT reporting is fundamental.
- Volume of data:** With over one-third of the global 1,000 companies reported in the repository, the opportunities for analysis are endless.
- Industry data:** Data has been collected from all industries from around the globe. Hence, the opportunity to compare industries from different geographies becomes another value analytic.
- Methodology data:** The alignment assessment focuses on a “higher level” of relationship across an organization; hence, while methods applied for governance, value, and service levels are collected, the focus is on identifying opportunities for improving the relationship among IT and business organizations.
- Language data:** As indicated, the assessment is at a higher level than programming language; albeit the tool has been translated in to Spanish. Portuguese, Chinese...
- Country data:** The geographic spread of the data is: 50% – U.S.; 18% – Europe; 12% – Latin America; 12% – Asia; 8% – Australia.
- Unique data:** Of the over one-third of the Global 1,000 organizations that are included in the benchmarking repository, the participant spread is 1,200+ - IT executives; 1,330+ - Business executives.
- Important insights include appropriate IT reporting structure, IT organizational structure, the need to assess ALL six of the components, and the strong correlation of IT business alignment and firm performance across industry.
- Future data:** Data continues to be collect and benchmarked from around the globe. Different industry benchmarks continue to provide important lessons. Additionally, performing subsequent assessments after several years has indicated continuous improvement. Additionally, over time, while maturity scores are

improving, the overall average is still in the low 3's; hence there are still opportunities for improvement.

**Summary:**

Jerry Luftman has been collecting alignment maturity data since 2001. Jerry Luftman's career includes strategic positions in management (Information Technology and consulting), management consulting, Information Systems, and education. Dr. Luftman's experience combines the strengths of practitioner, consultant, and academic. His proficiency in business-IT alignment, eighteen books, published research, consulting, mentoring, and teaching/speaking engagements further exemplify Dr. Luftman's expertise and leadership in his field.

After a notable twenty-two year career with IBM, he is currently a Distinguished Professor at Stevens Institute of Technology School of Management. Besides the traditional responsibilities of a Professor, teaching Masters and Ph.D. students, Dr. Luftman is the founder and Executive Director of the Stevens' Information Systems programs; one of the largest in the world. In this capacity, Dr. Luftman manages all aspects of the program including course building, research, marketing, administrative processes, and executive programs. Dr. Luftman's project experience ranges from senior management issues through tactical implementation. His current initiatives include creating business and information technology strategies and plans, facilitating IT governance processes, and developing and deploying applications leveraging emerging technologies. Dr. Luftman most recently pioneered the vehicle for assessing the maturity of IT-business alignment; where he has a repository of over one-third of the Global 1,000 companies. He also serves on the executive board of several companies, organizations, and publications.

Dr. Luftman's last responsibility at IBM was a key speaker at their Customer Executive Conference Center in Palisades, N.Y. While responsible for management research, he played a significant role in defining and introducing the IT Strategy and Planning, and Business Process Reengineering practice methods for the IBM Management Consulting Group. His framework for applying the strategic alignment model is fundamental in helping clients understand, define, and scope the appropriate strategic planning approach to pursue. Dr. Luftman's annual international IT trends survey sponsored by SIM, CIONet, and other CIO associations is recognized internationally as an industry barometer.

**PRICE Systems, LLC**

www.pricesystems.com

Email: arlene.minkiewicz@pricesystems.com

**Sources of data:**

Data comes from many sources. Proprietary data is collected from clients through consulting engagements, research engagements and training. Publically available data is purchased from external sources. Public data is also mined from sources published technical reports and articles as well as reports in the media.

**Data metrics:**

Data collected include software size (in Source Lines of Code (SLOC), Function Points or Use Cases), project effort, project cost, various technical parameters about the nature of the product being developed, the development team, and the developing organization.

Schedule duration for projects is collected as well.

**Data usage:**

Data is used to for creation, verification and validation of estimating relationships for cost, effort and schedule in the TruePlanning for Software product. Data is used to update knowledge bases which provide user guidance through various means within the product.

Data collected is also used to support research efforts, internally and externally funded, focused on specific areas of software development, technology, or process.

**Data availability:**

Raw data is generally unavailable due to its proprietary nature, summary data is available upon request

**Kinds of data:**

Software project data for projects (or project components) ranging from 1000 – 3,000,000 SLOC and projects (or project components) ranging from 25 – 1,000,000 Function Points. Data is primarily for individual software projects – some data is provided down to the component level, others are higher level project data.

**Volume of data:**

Data collected from more than 3500 projects from 1977 through today. New data is added monthly. Old data is retained.

**Industry data:**

Data from systems and embedded software, military software, real time software, commercial software, IT projects, MIS systems, civilian government projects, and outsourced projects.

Industries include banking, insurance, manufacturing, accounting, transportation, telecommunications, aerospace, defense, and government at both state and national levels.

Data is collected primarily from large organizations or agencies.

**Methodology data:**

Data is collected for a variety of methodologies and development process models.

**Language data:**

Data collected covers a wide range of programming languages.

**Country data:**

About 85% of the data comes from sources in the US. Additional data is provided from mostly European sources.

**Unique data:**

In addition to data that we have collected from our clients and other sources, PRICE also has been collecting data from our own software development efforts over the last 30+ years we have been developing software applications. Having been an agile shop for much of the last decade, this data includes many of the common productivity metrics associated with agile development

**Summary:**

PRICE Systems has been collecting software data since the mid 70's during the research and development of the PRICE Software model, which has evolved into the TruePlanning for Software model. Data collection continues to support regular upgrades to the software cost, schedule and effort estimating ability in this model.

PRICE Systems is a global leader of cost management solution in Aerospace, Defense, Space and Information Systems, serving over 250 customers worldwide.

**Process-Fusion.net**

**URL:** [www.Process-fusion.net](http://www.Process-fusion.net)

**Email:** [ggack@process-fusion.net](mailto:ggack@process-fusion.net)

- Sources of data:** Approximately 75% of this data set comes from formal inspections training event workshops in which participants inspect actual work products prepared prior to the workshop. Approximately 25% of the data are from post-training inspections of work products.
- Data metrics:** This data set includes approximately 1,600 individual inspections conducted by team of 3 or 4 in most cases. All of these inspections are consistent with IEEE Std. 1028-2008. Data elements collected for each inspection include work product type (requirements, design, code), size in lines, orientation hours, preparation hours, team meeting hours, number of major and minor defects identified, and post-correction review hours.
- Data usage:** Summaries of these data may be used as a basis for estimating the number of defects likely to be found in the several work product types identified, the number of hours per defect (or defects per hour) likely to be required when using formal inspections consistent with the process defined by IEEE Std. 1028-2008 (a.k.a. “Fagan-style Inspections”). Standard deviation and distribution data are also available.
- Data availability:** Summary data are available from the author upon request.
- Volume of data:** Approximately 1,600 inspections.
- Industry data:** Data from systems and embedded software, military software, commercial software, IT projects and outsourced projects.
- Data is collected primarily from large organizations with more than 500 software personnel.
- Language data:** A variety of programming languages are included.
- Country data:** About 80% of the data is from the U.S.

## **QuantiMetrics**

Web site URL: [www.quantimetrix.net](http://www.quantimetrix.net)

Email: [bram@quantimetrix.net](mailto:bram@quantimetrix.net)

- Sources of data:** Organisations (public and private) who subscribe to the software benchmarking services of QuantiMetrics. Data submitted is subject to rigorous consistency, validity and reasonableness checks; measurement practices of subscribers are pre-audited to avoid potential data problems. In many instances QuantiMetrics carry out the size measurements using certified staff; where size measures are provided by subscribers, these are audited by QuantiMetrics.
- Data metrics:** Project data: size measures; resource and time use by development phase, resource use by activity type; costs; quality measures by quality step; methods, languages/tools and platforms used; project team staffing and experience; qualitative factors deemed to have affected performance. Measures reflecting plans at significant mile-stones, are also collected.
- Applications data: size measures; annual resource used by activity type; volumes of changes, change and support requests; costs; application support team experience and ratings on maintainability and other factors; languages/tools and platforms used; application age; application instances/versions; user population; support window
- Data usage:** Applications project and support performance benchmarking and estimating; at least a third of instances this data has been used in the context of outsourcing contracts
- Data availability:** Data is deemed to be proprietary – participants are provided with like-for-like benchmarks for all measures they provide for their projects and applications. Anonymity is preserved at all times.
- Kinds of data:** Project and Application support for business applications
- Volume of data:** Data accumulated goes back to mid 1980s; early data largely based on line-of-code measures of size, while use of IFPUG functional size measurement dominant from 1990s and onward. The number of projects accumulated in total is about 10,000, yet benchmarks are only based on the most recent 3,000 or so projects, of which a large proportion are post Y2K. Applications measured number about 1,000.

<b>Industry data:</b>	Industry participants are largely from the private sector, with financial institutions the most dominant, then telecoms businesses although the applicability of the benchmarks is appropriate for all business type projects.
<b>Methodology data:</b>	Project data includes information on approaches and methodology (proprietary or generic type). In the case of Agile-based methods ratings of the degree of adoption of practices also obtained.
<b>Language data:</b>	No restriction is placed on the development languages (tools) used; where statistically reliable data is available benchmarks are language specific, otherwise they are based on type of language/tool. Early submissions were dominated by third generation languages, followed by a period when a variety 4 <sup>th</sup> generation tools was the norm, while now object oriented languages and “visual” tools are more the norm. Projects also include those that configure and roll out business applications software.
<b>Country data:</b>	Largest and dominant contributions are from UK, Germany, South Africa and India
<b>Unique data:</b>	QuantiMetrics use statistical techniques to model and normalize performance, particularly for the effects of: size and time in the case of projects; size, age and user population for applications support data.
<b>Future data:</b>	Expect to see more use of COSMIC as a size measure
<b>Summary:</b>	QuantiMetrics have worked to ensure the database contains good-quality data and that the most statistically reliable like-for-like benchmarks are provided. Reports and presentations are aimed to identify systemic cause-effect relationships and, in particular, to suggest best practices and opportunities for improvement.

## **Quantitative Software Management, Inc. (QSM)**

Web site URL: <http://www.qsm.com>

Email: [info@qsm.com](mailto:info@qsm.com)

**Sources of data:** Since 1978, QSM has collected completed project data from licensed SLIM-Suite® users and trained QSM consulting staff. Consulting data is also collected by permission during productivity assessment, benchmark, software estimation, project audit, and cost-to-complete engagements. Many projects in our database are subject to non-disclosure agreements but regardless of whether formal agreements are in place, it is our policy to guard the confidentiality and identity of clients who contribute project data. For this reason, QSM releases industry data only in summary form to preclude identification of individual projects/companies or disclosure of sensitive business information.

**Data metrics:** Our basic metric set focuses on size, time, effort, and defects (SEI Core Metrics) for the Feasibility, Requirements/Design, Code/Test, and Maintenance phases. These core measurements are supplemented by nearly 300 other quantitative and qualitative metrics. Approximately 98% of our projects have time and effort data for the Code and Test phase and 70% have time/effort data for both the R&D and C&T phases.

Productivity is captured via the following metrics:

- QSM Productivity Index (PI)
- Cost per SLOC or Function Point
- SLOC or Function Points per month
- SLOC or Function Points per Effort Unit (Months, Hours, Days, Weeks, Years)

Quality data is captured via the following metrics:

- Total Errors (System Integration test – Delivery)
- MTTD
- Errors/Month
- Errors/Effort Unit (various)
- Errors/KESLOC or Function Point
- Post release defects at 1, 3, 6, 9, 12, and 18 months

<b>Data usage:</b>	Data from our database is used to: <ul style="list-style-type: none"> <li>• Keep our tools and algorithms in sync with the latest industry practices and technologies.</li> <li>• Create the nearly 600 industry trend lines included with SLIM-Suite applications.</li> <li>• Support our basic measurement, benchmarking, estimate validation and verification and other consulting services</li> <li>• Conduct ongoing research available via books, papers, articles and blog posts available via our web site.</li> </ul>
<b>Data availability:</b>	To protect our clients, we release project data in summary form only. QSM consulting and tools clients may receive graphs and summaries that allow them to compare their projects against industry trend lines and actual projects that are similar in size, application type, and complexity. Summary statistics and information about the Putnam Model are also available via books like <u>Five Core Metrics</u> , <u>Measures for Excellence</u> and <u>Industrial Strength Software</u> , webinars, QSM IT Almanacs, blog posts, technical reports and whitepapers, conference presentations, and journal articles.
<b>Kinds of data:</b>	Individual project duration, effort, size, productivity, and quality data is available for projects ranging from 2-57,000 function points and 100 - over 10 million SLOC. Our core metric set is supported by nearly 300 other quantitative and qualitative measures.
<b>Volume of data:</b>	About 10,000 projects from 1978 through the present. To facilitate longitudinal studies of the industry, older project data is retained.
<b>Industry data:</b>	QSM data is stratified into 9 major application domains (Avionics, IT, Command & Control, Microcode, Process Control, Real Time, Scientific, System Software, and Telecom) and 45 sub-domains. Software projects predominate, but we have a growing number of hardware and infrastructure (non-software call center) projects as well. <p>Data contributors include DoD; civilian commercial firms; and national, state and local government entities. In addition to domain complexity bins, our data is also broken out by major industry and industry sector. Major industries include the financial sector,</p>

banking, insurance, manufacturing, telecommunications, systems integration, medical, aerospace, utilities, defense, and government.

**Methodology data:** The QSM database includes a variety of lifecycle and development methodologies (Incremental, Agile, RUP, Spiral, Waterfall, Object Oriented) and standards (CMM/CMMI, DoD, ISO).

**Language data:** Approximately 600 languages are represented with most projects recording multiple languages. Common primary languages are JAVA, COBOL, C, C++, C#, VISUAL BASIC, .NET, IEF / COOLGEN, PL/1, ABAP, SQL, ORACLE, POWERBUILDER, SABRETALK, JAVA SCRIPT, DATASTAGE, HTML. Frequently used secondary languages include JCL, SQL, JAVA, COBOL, ASSEMBLER, C++, HTML, VISUAL BASIC, XML, ASP.NET, and JSP.

**Country data:** QSM has collected and analyzed software projects from North America, Europe, Asia, Australia, and Africa. About 50% of our data is from the U.S. Another 35-40% is from India, Japan, the Netherlands, the United Kingdom, Germany, France, and other major European countries.

**Future data:** We are currently expanding our data collection to include non-software and hardware projects as well as institutional or portfolio level performance measures.

**Summary:** As pioneers and innovators in the software estimation field since 1978, Quantitative Software Management is more than just a company name; it's the philosophy that defines everything we do. Our goal is to help our clients become world-class software developers so they can deliver world-class software projects. For over 30 years we've provided superior estimation products and consulting services that enable our clients to estimate, track and benchmark projects with confidence. Our tools and methods can help you negotiate achievable goals, set realistic expectations and communicate more effectively with your colleagues and customers.

At the helm of QSM is Larry Putnam, who founded the company in 1978. Larry possesses more than 30 years of experience as a software management consultant and has worked on more than 400 client engagements. Considered to be among the top problem solvers in the software estimation and measurement field, Larry has authored four books and hundreds of articles. He is sought internationally as a speaker and management consultant.

Headquartered in Washington, DC, QSM also has regional distributors in Massachusetts, England, and the Netherlands. QSM has established itself as the leading total solution provider of choice for software developers in high-performance, mission-critical environments. For more information about QSM research, benchmarks, tools, and services please visit our website at <http://www.qsm.com>.

Q/P Management Group, Inc.  
Web site URL: [www.qpmg.com](http://www.qpmg.com)  
Email: [moreinfo@qpmg.com](mailto:moreinfo@qpmg.com)

**Sources of data:** Q/P Management Group's benchmark database is comprised of over twenty thousand (20,000) projects and applications from major corporations, commercial developers, and government agencies. The database contains statistics on a broad range of tools and techniques utilized by these organizations. Data is collected through a number of vehicles including client benchmark studies, outsourcing baseline assessments, ongoing performance measurement engagements, and through helping companies establish measurement and estimating programs. The data is collected under confidential agreements with the clients contributing to the database. Project and application data are added to the database after rigorous analysis and verification to ensure the highest degree of data integrity in the industry.

**Data metrics:** Productivity for projects is expressed in function points developed per hour (FP/Hour) or hours per function point (Hours/FP). All function point data is manually counted, utilizing the IFPUG method. Project Cost is expressed in cost per function point (Cost/FP). Project Quality is expressed in defects reported per developed function point (Defect/FP). Project Schedule is expressed in project calendar days required to deliver the project from initiation to first implementation. Maintenance Productivity is expressed in application function points maintained per full time equivalent employee (FP/FTE) or function points maintained per hour (FP/Hour). Maintenance Quality is expressed in annual defects reported per application function point (Annual Defects/FP). Maintenance Cost is expressed in annual maintenance cost per application function point (Annual Cost/FP). Other benchmark data is available including Percent Lifecycle Effort by Phase, Estimating Accuracy, Maintenance Activity Breakdown and Defects by Severity Level.

**Data usage:** Q/P and their clients utilize the data in various ways, including comparing performance of internal or vendor resources against industry benchmarks, as a means to identify and measure process improvements and to determine pricing for commercial software products. The data is also used for estimating software development projects' productivity, cost, schedule, and staffing.

**Data availability:** Benchmarks are utilized by clients that engage Q/P consultants to conduct benchmark studies. Benchmark data can also be licensed through two primary mechanisms. The first allows an organization

to license the data for use in estimating and reporting through Q/P's Software Measurement, Reporting and Estimating (SMRe) product. Alternatively, organizations can license summarized segments of the database for internal use to evaluate performance. Samples of summary data tables are available from Q/P on request. Contact Q/P for pricing options.

**Kinds of data:** Data exists on both software development projects and application maintenance. Projects include both new development and enhancement and range in size from less than 5 to over 20,000 function points. Application age ranges from 1 to 40+ years and application size ranges from less than 100 to over 30,000 function points. The software classifications and functions in the database include:

Data Warehouses	Middle-ware
Telemarketing	Customer Care
Cellular Services	Telco Billing
Equipment Procurement	Plant and Facilities
Network Management	Credit Card Processing
Funds Transfers	Securities Fraud
ATM Networks	P.O.S. Networks
Human Resources	Finance and Accounting
Telco Provisioning	Distribution
Stocks and Bonds	Account Management
Contract Management	Weather Forecasting

**Volume of data:** Since 1993 there have been over 15,000 project data points and over 6,000 application data points collected. The database is updated annually utilizing data collected during the previous year. Older data is retained within the database for historical trend analysis purposes but is excluded from current performance reporting.

- Industry data:** Data is collected from a variety of industries including: financial services, government, healthcare, telecommunications, manufacturing, pharmaceuticals, public utilities, retail, transportation and software developers.
- Methodology data:** Data is collected on a variety of development methodologies including, but not limited to: Agile, Waterfall and Rational Unified Process (RUP), as well as hybrid techniques utilizing portions of several disparate methods. Data included in the data base are aligned with the CMMI level the organization had achieved at the time the data was collected.
- Country data:** Over 60 percent of the data is from North American sources. The remainder of the data is primarily obtained through European and South American sources. Offshore (India) development and maintenance is well represented in both project and application data.
- Summary:** Q/P Management Group (Q/P) was one of the first organizations to offer performance benchmark assessments using function point based benchmarks. The benchmarking methodologies used by Q/P are comprehensive and based on industry accepted standards. Q/P's function point based benchmark data base is the largest and most wide-ranging available in the industry today. Numerous companies and government agencies utilize this database to benchmark software quality, productivity and cost.

**RBCS, Inc.**

Web site URL: [www.rbc-us.com](http://www.rbc-us.com)

Email: [info@rbc-us.com](mailto:info@rbc-us.com)

**Sources of data:** On-site interviews of software testing and quality assurance teams and groups within larger organizations that develop software and/or systems for internal use or for sale to customers, or which provide testing and quality assurance services. Discussions with clients in public and private training sessions. The data is collected under non-disclosure agreements, and thus only anonymized and collective data is available for comparative purposes.

**Data metrics:** The metrics are some two dozen testing and quality related process, product, and project metrics. Examples include: defect detection effectiveness, cost per defect found, percentage coverage of risks and/or requirements, percentage of invalid defect reports, percentage of re-opened defect reports, consistency of test case documentation, accuracy of test estimates, percentage overhead of test environment setup/maintenance, and various others.

**Data usage:** Data is used to assess the capability and process improvement potential of organizations involved in creation of software or systems for their own use or for use by clients/customers, and of organizations engaged in providing testing services. Data and metrics are mentioned in some of Rex Black's book, such as *Managing the Testing Process*, *Advanced Software Testing: Volume 2*, and *The Expert Test Manager* (scheduled for publication in 2012). Data is also used in many of Rex Black's presentations and articles.

Data is provided to specific clients of assessment, baseline, and benchmark studies. These studies compare clients against similar companies.

Data is discussed in various RBCS training courses, especially test management courses.

Data is also discussed in various RBCS free webinars on topics related to test management.

**Data availability:** Data is provided to clients of assessment and benchmark studies. General data is published in books and journal articles.

**Kinds of data:** Software and system quality and testing capability measurements focused on project, process, and product. People-focused metrics

are discussed in consultancy but not provided for general use, due to the potential for problems with such metrics.

Data comes from work with clients on assessments of their projects, processes, and products, as well as work with training clients.

**Volume of data:** RBCS has over 200 clients and has been in the software testing and quality assurance training, consulting, and outsource business since 1994.

**Industry data:** Data from systems and software projects spanning almost all applications of software and hardware/software systems, with no particular emphasis in terms of industry. Both private and government clients are included.

Data is collected from organizations ranging in size from small companies with less than a dozen employees to some of the largest organizations in the world (e.g., Sony, CA, etc.)

Due to the fact that assessments are paid consulting services, open source data is not included.

**Methodology data:** Data is collected for a variety of methodologies including Agile, waterfall, Rational Unified Process (RUP), Scrum, Extreme Programming (XP), and hybrid methods that combine features of several methods.

Data covers all types of testing, including functional testing, performance testing, usability testing, internationalization testing, manual testing, automated testing, regression testing, and so forth.

**Language data:** Many programming languages are included, but some of the older languages such as COBOL and FORTRAN are under-represented.

**Country data:** RBCS clients are distributed widely around the world, and include many multinationals. Emerging software testing countries are under-represented in some cases, but RBCS is heavily involved in software testing and quality training and consulting with emerging economies such as India, Bangladesh, Kenya, Nigeria, South Africa, Malaysia, and China.

**Unique data:** RBCS data is unique in the range of test-related metrics and data available to clients. While other consultancies provide proprietary test assessment and/or “maturity” services, many of these are

almost entirely questionnaire-based and/or interview based, and thus suffer from the Rashomon Effect. RBCS assessments, being evidence-focused and data-focused, provide deeper insights to clients. In addition, RBCS assessments are based on an open-source framework published in Rex Black's book *Critical Testing Processes*.

**Future data:**

RBCS intends to continue to provide quality and testing assessment services to clients around the world, and is working to expand the depth of the data through more assessments. RBCS is also considering developing an online assessment process using its proven e-learning platform.

**Summary:**

RBCS is unique in test consultancies in being focused on an evidence-based and metrics-based approach to assessment, which provides clients with reliability benchmarks and assessment results. RBCS has been in the quality and testing consulting business since 1994, and founder Rex Black's background in software engineering extends back to 1983.

Reifer Consultants LLC  
Web site URL: [www.reifer.com](http://www.reifer.com)  
Email: [info@reifer.com](mailto:info@reifer.com)

- Sources of data: Data sharing agreements with firms, government agencies and universities. The data is collected and protected under a data sharing agreement through collaborators, clients and via public sources.
- Data metrics: Productivity is expressed in equivalent source line of code (SLOC) per staff-month, cost in \$/SLOC, duration in months and quality in defects/thousand (K) SLOC. Data are segmented into three sectors (industry, defense and government) and twenty-three applications domains. Information is provided on paradigm utilization (agile, waterfall, etc.), percentage work in-source/out-sourced and personnel turnover rates by sector.
- Unique information about maintenance shops is presented. This is based on data collected during the past three years on hundreds of projects.
- Data usage: Data is used to create benchmarks for use in competitive positioning, comparative analysis, improvement studies and should-cost studies. Benchmarks are often used to validate estimates submitted for software development and maintenance efforts. These data are used as the basis of benchmarking reports developed for clients.
- Data availability: Benchmarks are provided clients on a subscription of contract basis. While analysis results can be shared, data are proprietary and are not sold.
- Kinds of data: Records of completed projects at the program and project levels. Data employed comes from 49 commercial firms (699 projects), 27 defense firms (225 projects) and 12 government organizations (76 projects). Size average for commercial firms is 675 KSLOC (range is between 20 to 4,180 KSLOC), for defense is 487 KSLOC (range is between 18 to 5,200 KSLOC) and for government organizations is 591 KSLOC (range is between 25 to 1,825 KSLOC). The average duration for projects in the database averages between 12 to 18 months. Metadata is reported by sector and applications domains.
- Volume of data: Data is taken from a database of over 10,000 records. Data used for benchmarks comes from 1,000 current projects selected to be representative of both development and maintenance efforts, none

of which is over 7 years old. Demographics for data on size, types of firms and applications domain characteristics supplied for 23 applications types.

Industry data:

Applications domains used to capture benchmarks include:

Industry Benchmarks (13): Automation, command and control, Info Technology – banking, Info Technology – general, Info Technology – insurance, medical, process control, scientific systems, software tools, telecommunications, test systems, training/simulation and web business.

Defense Benchmarks (7): airborne, ground, Info Technology, medical, missile, space and trainers.

Government Benchmarks (3): Info Technology, infrastructure systems and sensitive systems.

Methodology data:

Project processes include from agile development to traditional waterfall by sector and applications domain.

Some data collected for CMM, CMMI and ISO utilization.

Country data:

All data in database for United States. Separate Asian and European databases are maintained because data for the three are statistically not consistent. We believe it to be a mistake to mix such data because their attributes are not highly correlated.

Summary:

Reifer Consultants LLC has primarily developed and used benchmarks in its own consulting practice for over 30 years. We trust them because the data in them is highly correlated and statistically sound. Our clients trust them because the benchmarks have proven to be accurate predictors of productivity, cost, duration and quality. Reifer Consultants LLC has offered these benchmarks to its clients who have used them beneficially to make selective improvements and validate bids and estimates.

## Software Benchmarking Organization

Web site URL: [www.sw-benchmarking.org](http://www.sw-benchmarking.org)

Email: [hans.sassenburg@sw-benchmarking.org](mailto:hans.sassenburg@sw-benchmarking.org)

- Sources of data:** On-site benchmark studies and assessments of software projects. Data is always collected under non-disclosure agreements.
- Data metrics:** Data is collected for the following metrics:
- Schedule (calendar months)
  - Effort (person months)
  - Productivity (function points per hour using IFPUG definition)
  - Cost of Quality (effort distribution over 4 areas)
  - Deferral rate (ratio of deferred baselined features)
  - Feature size (function points)
  - Technical size (KLOC)
  - Re-use level (ratio)
  - Complexity (using McCabe definition)
  - Test coverage (unit, integration, system testing)
  - Defect density (defects per function point at release time)
  - Defect removal efficiency (ratio of defects removed before releasing)
- Data usage:** Data is used:
- to benchmark “engineering capability” against industry averages and best-in-class figures in same industry, using data from SBO assessments as well as published data from other sources (like Capers Jones)
  - to assess the feasibility of initial and running projects
- Data availability:** Data is provided to clients of studies and assessments
- Kinds of data:** Data is primarily for individual software projects
- Volume of data:** About 150 projects from 2001 through today. New data is added frequently.
- Industry data:** Data from systems and embedded software, military software, commercial software, IT projects, civilian government projects, and outsourced projects.
- Industries include banking, insurance, manufacturing, telecommunications, medical equipment, aerospace, defense, and government at both state and national levels.
- Data is collected primarily from projects with 10 – 75 software engineers

**Methodology data:** Data is collected for a variety of methodologies including Agile, waterfall, Rational Unified Process (RUP), and hybrid methods that combine features of several methods.

Data is also collected for the five levels of the Capability Maturity Model Integrated (CMMI™) of the Software Engineering Institute.

**Language data:** As is usual with large collections of data a variety of programming languages are included. The number of languages per application ranges from 1 to 5, with an average of about 2. Most common combinations include Assembler, C, C++, C# and Java.

**Country data:** Most of the data is from Western Europe (including The Netherlands, Germany, France, Switzerland, United Kingdom) with a strong focus on the embedded software industry.

Limited data from other countries/regions.

**Unique data:** The collected data is used to compute 16 KPIs, arranged in four different categories. This enables an organization to identify root causes of underperformance and estimate the effect of corrective measures.

**Future data:** Currently, limited data is available regarding complexity and test coverage. This will receive more attention.

**Summary:** SBO has been collecting software data since working 2001 using benchmarking and assessment studies. SBO uses a network of accredited partners.

In addition, SBO delivers metrics workshops as well as supporting technology for metrics collection and analysis.

## Software Improvement Group

Web site: [www.sig.eu](http://www.sig.eu)

Email: [info@sig.eu](mailto:info@sig.eu)

**Sources of data:** SIG obtains source code and other data on software systems from its customers and from various open source projects. SIG is equipped with an ISO/IEC 17025 certified software evaluation laboratory, in which source code is processed and analyzed according to a standardized evaluation procedure. Analysis results are stored in the benchmark repository, called the Software Analysis Warehouse (SAW).

**Data Metrics:** The SAW stores diverse types of metrics calculated automatically from source code or other software system-related artifacts. For most systems, these metrics are available for multiple snapshots. Metrics stored in the SAW include:

- Traditional code metrics such as coupling, complexity, and size. These metrics are available at the most detailed level of methods and functions, as well as aggregated into quality profiles at the level of files, packages, modules, and entire systems.
- Churn metrics, i.e. added, deleted, and modified lines of code at the level of files.
- Maintainability ratings according to the *SIG/TÜViT Evaluation Criteria* at the level modules and systems.
- Programming languages used in projects and their respective volumes.
- Issue handling performance metrics, including resolution time, resolution speed ratings, and issue handling productivity.

**Data Usage:** The data collected is used for two main purposes. First, the data is used to support various consulting services. For example, the benchmark data set is used for annual re-calibration of the SIG maintainability model, which is applied in SIG's services for software risk analysis, software quality and productivity monitoring, and software product certification. Yearly re-calibration ensures that the model remains a faithful reflection of the state of the art of software quality.

The second use of the data is for research. SIG is actively engaged in many research initiatives internationally in various research themes such as mining software repositories, software architecture, and software testing. The data is used for validating theories or beliefs in software engineering. Results of such validations have been published in international journals and conferences.

- Data availability:** Data obtained from customers is subject to non-disclosure agreements, and is therefore not available.
- Data presented in public SIG Analysis Reports or in scientific publications is typically available (consult the respective publications), or otherwise may be provided upon request.
- Kinds of data:** Systems stored in the SAW have either one or multiple snapshots. Systems with only one snapshot are obtained from customers that requested a one-off quality assessment. Systems with multiple snapshots are those monitored by SIG throughout a certain period of time. Apart from proprietary system, the SAW contains data for a range of open source systems that were analyzed for research purposes.
- Volume of data:** At the time of writing, data in the SAW regards over 400 unique systems and around 7,500 snapshots. The total lines of code reach about 74 million lines of code. The size of systems ranges up to 3M source lines of code.
- Customers upload new system snapshots on a regular basis. Depending on the agreement the upload frequency can be weekly, monthly, or quarterly. Over 100 new systems are added to the SAW yearly.
- Industry data:** Systems in the SAW come from various domains, which include transportation, insurance, banking, and telecommunication.
- System owners are either medium to large size enterprises or governmental organizations.
- Methodology data:** Data is collected from systems developed using a variety of methodologies. Iterative/agile methodologies seem to be the most frequent methodology followed. Apart from systems that are under development or were recently developed, data is collected on systems that are under maintenance. These include legacy systems that were once developed using non-iterative, waterfall methodologies.
- Technology data:** The systems of which data is currently stored in the SAW are built using over 70 different technologies, which include Java, .NET, C/C++, PL/SQL, BPEL, various dialects of COBOL, and Pascal.
- Country data:** The majority of the data is obtained from customers in Europe, particularly from Western Europe and Scandinavian countries. Data is also increasingly coming from other parts of Europe and

other regions such as UK, Greece, and Turkey. IT suppliers of the customers come from a more diverse region such as Eastern Europe and Asia.

**Unique data:** Multiple snapshots of system code give unique opportunities to perform various data analyses. With hundreds of systems, in which each of them has multiple snapshots, SIG has performed trend analysis for numerous metrics and has developed several prediction models. For example, SIG has developed an economic model to estimate renovation efforts based on churn analyses of multiple system snapshots.

**Future data:** SIG is constantly collecting more data from industrial systems. This ensures that new systems, especially those developed using new technologies, are represented in the benchmark data set. Furthermore, SIG has started to collect data on software specifications and models.

**Summary:** SIG is an Amsterdam-based consultancy firm that provides consulting services in the area of software quality, software risk, and software economics. These services are supported by a standardized software evaluation methodology, carried out by SIG's certified software evaluation lab. In partnership with TÜV Informationstechnik (TÜViT), SIG provides certification of software maintainability.

The Software Analysis Warehouse (SAW) is the main benchmark database of SIG. The SAW contains detailed information on numerous snapshots of a steadily increasing number of software systems (currently more than 400). The benchmark is used in consultancy services as well as in software engineering research.

Well-established tooling and procedures allow SIG to generate a large volume of software engineering data from both commercial and open source systems in a consistent and controlled manner. Data collection is highly curated by technical and client-facing consultants, ensuring a reliable source of data with very strong traceability characteristics.

## Test Maturity Model integrated (TMMi) survey

Survey Web site URL: [www.experimentus.com](http://www.experimentus.com)

Email: [geoff.thompson@experimentus.com](mailto:geoff.thompson@experimentus.com)

**Sources of data:** Submissions invited via multitude of social networking sites (LinkedIn, Twitter and Facebook etc.) and also from organisations who have been assessed using the TMMi model. The data is collected under an agreement of anonymity. Data is submitted directly via a weblink on the Experimentus website.

**Data metrics:** A questionnaire is used to collect responses aimed at identifying company and individual test process maturity, based upon the contents of the TMMi (Test Maturity Model – see [www.tmmifoundation.org](http://www.tmmifoundation.org)) model.

The data collected is pooled and is used to identify, for each of the 10 Key Process Areas in TMMi levels 2 and 3, what level maturity each companies test process is at, as well as an overall general perspective on the maturity of the test industry.

The results are published every 18 months, with the data being collected during the previous 18 months.

Data is collected from over 100 companies over 5 continents.

**Data usage:** The data is used to interpret the current maturity of the overall test processes in use across the world, as well as specifically within each company.

**Data availability:** The ‘raw’ data is not available for public viewing; however the report is freely downloadable from the Experimentus website.

**Kinds of data:** The data is in the form of responses to specific statements specific to the differing components of the TMMi test process, in four dimensions – well-established, exists but is not deployed successfully, embryonic stage and no process exists.

In addition to the data collected from these 16 questions, all data collected from all formal assessment carried out by Experimentus in the 18 months between reports is also summarised and included.

**Volume of data:** Each 18 months, approximately 100 different responses are collected to 16 specific statements.

- Industry data:** Industries responding include IT Support Services, Banking, Finance, Legal, Manufacturing, Retail, Local Government and Education.
- Data is submitted from a wide range of organizations of differing sizes. There is no data from the military.
- Methodology data:** Data is collected for a variety of methodologies including Agile, SCRUM, Waterfall, Joint Application Design (JAD), Rational Unified Process (RUP), Team Software Process, (TSP), and hybrid methods that combine features of several methods.
- Country data:** About 30% of the data is from the UK, 20% from the U.S. 10% from Japan; 10% Australia; 10% Netherlands, 7% India. Also the data from 7 other countries is included.
- Summary:** Experimentus was established in 2004 to provide true Consultancy in the international IT Quality arena, helping companies to deliver better quality software quickly and cheaply. Its principal measurement of client maturity is to assess against the TMMi model. They were one of the first companies to have their TMMi assessment method accredited by the TMMi Foundation and were the first company in the world to have accredited assessors. They carry out around 14 formal TMMi assessments each year; in a variety of countries, most notably UK, Spain and Malaysia. The data from these assessments is included within the report.
- The TMMi Foundation was formally launched in early 2007; it has over 600 members from all across the world. It owns and maintains the TMMi Model, as well as accreditation of assessment methods and assessors. More details can be found at [www.tmmifoundation.com](http://www.tmmifoundation.com).

# APPENDIX A: SURVEY OF BENCHMARK USAGE AND INTERESTS

## SOFTWARE BENCHMARK USAGE SURVEY

Version 2.0 3/7/2010

	No Interest "0"	Used Today "1"	Would Use if Available "2"
1. Portfolio benchmarks			
2. Industry benchmarks (banks, insurance, defense, etc.)			
3. International benchmarks (US, UK, Japan, China, etc.)			
4. Application class bench marks (embedded, systems, IT, etc.)			
5. Application size benchmarks (1, 10, 100, 1000, function points etc.)			
6. Requirements creep benchmarks (monthly rates of change)			
7. Data center and operations benchmarks (availability, MTTF, etc.)			
8. Data quality benchmarks			
9. Data base volume benchmarks			
10. Staffing and specialization benchmarks			
11. Staff turnover and attrition benchmarks			
12. Staff compensation benchmarks			
13. Organization structure benchmarks (matrix, small team, Agile, etc.)			
14. Development productivity benchmarks			
15. Software quality benchmarks			
16. Software security benchmarks (cost of prevention, recovery, etc.)			
17. Maintenance and support benchmarks			
18. Legacy renovation benchmarks			
19. Total cost of ownership (TCO) benchmarks			

20. Cost of quality (COQ) benchmarks
21. Customer satisfaction benchmarks
22. Methodology benchmarks (Agile, RUP, TSP, etc.)
23. Tool usage benchmarks (Project management, static analysis, etc.)
24. Reusability benchmarks (volumes of various reusable deliverables)
25. Software usage benchmarks (by occupation, by function)
26. Outsource results benchmarks (domestic)
27. Outsource results benchmarks (international)
28. Schedule slip benchmarks
29. Cost overrun benchmarks
30. Project failure benchmarks (from litigation records)
31. Litigation cost benchmarks - breach of contract
32. Litigation cost benchmarks - taxable value of software
33. Litigation cost benchmarks - non competition violations
34. Litigation cost benchmarks - damages due to poor quality
35. Litigation cost benchmarks - intellectual property

## **APPENDIX B: A NEW FORM OF SOFTWARE BENCHMARK**

### **Introduction**

Normally software assessments and benchmarks are provided to specific companies and compare selected applications against similar applications from other companies in the same industry. This is useful information, but it does not provide any context or any information about the industry itself.

What would be both useful and salable would be a new kind of benchmark that would consolidate information about specific industries, the major companies within the industries, the software used by those companies, and also productivity and quality ranges derived from assessment and benchmark studies.

Using “banking” as an industry example here are some 55 kinds of assessment and benchmark information that would be provided:

#### **Table 1: Fifty Five Data Topics for Industry-Specific Software Benchmarks**

1. Number of large companies (source = Hoover’s Guides to business)
2. Number of medium companies (source = Hoover’s Guides to business)
3. Number of small companies (source = Hoover’s Guides to business)
4. Regulatory agencies that control business sectors
5. Standards and safety regulations for the industry
6. Supply chains and related industries
7. Current government investigations involving the industry
8. Mergers and acquisitions within the industry
9. Start-up companies within the industry
10. Business failures, government takeovers, or bankruptcies within the industry
11. Recent patents and intellectual property filed by industry members
12. Ranges of industry profitability and economic health
13. Current litigation involving the industry
14. Domestic competitive situation of the major players within the industry
15. Global competitive situation of the major players within the industry
16. Professional associations that serve the industry
17. Major kinds of hardware platforms used within the industry
18. Major kinds of data utilized within the industry
19. Major kinds of software applications utilized by the industry
20. Major kinds of ERP applications utilized by the industry
21. Major COTS vendors that provide packages to the industry
22. Major open-source applications utilized by the industry
23. Major outsource vendors that service the industry
24. Major sources of reusable components serving the industry
25. Ranges of portfolio and application sizes within the industry

26. Ranges of data base and data warehouse sizes within the industry
27. Numbers of software users within the industry
28. Number of customer organizations served by software
29. Number of actual clients or users served by software
30. Numbers of software developers within the industry
31. Numbers of maintenance personnel within the industry
32. Numbers of technical specialists within the industry (quality, testing, etc.)
33. Rates of change for software personnel (expanding, shrinking, stable)
34. Software security issues within the industry
35. Statistics on data theft, denial of service, and other security breaches
36. Security policies, standards, and best practices for the industry
37. Software development productivity benchmarks with function points
38. Software maintenance/enhancement productivity benchmarks (function points)
39. Software total cost of ownership (TCO) benchmarks with function points
40. Software quality benchmarks with function points
41. Software cancelled project benchmarks by size in function points and type
42. Software costs and schedule overruns within the industry
43. Legacy application replacement strategy within the industry
44. Distribution of CMMI levels within the industry
45. Distribution of TickIt scores or maintainability scores within the industry
46. Major development methods used within the industry
47. Major maintenance methods used within the industry
48. Typical tool suites used within the industry
49. Best practices utilized by the industry
50. Average practices utilized by the industry
51. Worst practices utilized by the industry
52. Major quality control methods used within the industry
53. Future industry technology trends (cloud computing, SOA, etc.)
54. Future industry software trends
55. Major sources of industry data (web sites; periodicals, etc.)

For a bank to assemble all of this information by itself it would be necessary to gather data from about a dozen industry and government sources plus probably commissioning benchmarks on a sample of 10 to more than 25 applications. Competitive information from other banks would not be accessible. Essentially this kind of information would not be gathered by individual banks because of a lack of organizational focus, plus the rather high costs involved.

### **Probable Clients for Software Benchmarks by Industry**

As of 2011 the software benchmark business sector is divided into two subsets. One form of benchmark uses fairly simple questionnaires with the data being self-reported by clients.

Because self-reported benchmarks have no fees for providing information and consolidated benchmark reports are available for low cost, this form of benchmarking is

widespread and popular. The International Software Benchmark Standards Group (ISBSG) is the major provider of self-reported benchmarks. The ISBSG clients consist mainly of project managers and some CIO's.

The second form of software benchmark uses more complicated questionnaires and also includes on-site data collection in order to gather and validate quantitative and qualitative information from samples of 10 to more than 25 projects at the same time.

These on-site benchmarks usually include non-disclosure agreements for data collection and distribution so the data is delivered to specific companies.

Because the costs for collecting and analyzing the data ranges from \$25,000 to more than \$50,000 these benchmarks require approval and funding from the level of a CIO or a CTO. The reports back to clients are of course used by first-line and project managers, but the funding is usually provided by a higher level of management.

Software assessments are also on-site consulting studies. For assessments using the model of the Software Engineering Institute (SEI) certified assessors are used. For other forms of assessment such as the author's, trained consultants are used. Here too the costs are fairly expensive and in the \$25,000 to \$75,000 range.

The data that would be assembled for the new kind of benchmarks discussed in this report would include a combination of self-reported data, on-site data collection, and subscriptions to a number of industry information providers such as Hoover Business Guides, Gartner Group, the Department of Commerce, plus access to other government statistics as well. At a nominal charge of \$10,000 for such a benchmark report, funding approval would probably be at the CIO and CTO level.

While the costs of these benchmarks are less than the cost of today's on-site benchmarks and assessments for individual clients, these new kinds of benchmarks could be offered to dozens or even hundreds of clients so they would generate much greater revenues and profits than conventional single-client benchmarks.

In order to be useful, the benchmark reports would consolidate data from at least 10 companies and 100 projects within an industry, and then use extrapolation to cover other companies within the same industry. Of course larger samples would be desirable. Remote data might be gathered from 100 banks or more, while on-site data might be gathered from 20 banks or more.

The on-site data collection would probably be derived from conventional fee-based studies that provide information to specific clients. However once the data is sanitized and aggregated, it would be marketed to multiple clients.

Because of the richness of the data provided, these new benchmarks would attract a much wider and more diverse set of clients than normal self-reported or on-site software benchmarks. Obviously the information would be of use to CIO's and CTO's, but

because of the in-depth industry coverage the information would also be of use to CEO's and to client executives as well.

For example these new benchmarks would be of use to VP's of marketing, sales, manufacturing, human resources, and research and development. In addition the information would no doubt be acquired by major consulting companies, by law firms that specialize in software litigation, by outsource vendors, and by other kinds of information providers such as journals and web sites.

In addition the new form of benchmark would also be useful to many related companies that provide services or products to banking clients: outsource vendors, software vendors, consulting companies, equipment manufacturers, personnel companies, major law firms, and government regulators. In fact sales of the new kind of benchmark to these ancillary companies would probably exceed sales to the banking community itself. For each benchmark study acquired by a bank, probably at least three studies would be acquired by banking service and product providers. This is a new and previously untapped market for benchmark studies.

The basic idea of the new form of benchmark is to elevate the value of benchmark information from data that is "useful but not essential" to the level of "we must have this information to stay competitive." A second goal is to elevate the target audience of the benchmark information from project managers, CIO's, and CTO's up the level of CEO's and senior operating unit executives.

Once this new form of benchmark is launched, it will probably lead to a significant increase in other forms of benchmarking.

It is obvious that the initial launch within an industry such as banking needs to attract a fairly significant number of clients. Therefore the new form of benchmark should start with an industry where such information is already perceived as valuable; i.e. banks, defense, insurance, health care, medical equipment, and the like.

Once launched in the United States, these new benchmarks would certainly lead to an increase in overseas benchmarks using the same formats and data collection methods. However to facilitate overseas benchmark data collection, local subcontractors would probably be a desirable method of proceeding.

In addition, some overseas data might be gathered via on-line methods such as Skype, remote surveys, and perhaps Wiki sites. In fact a virtual benchmark environment using the same technology as Second Life is technically possible. In such an environment avatars of consultants and clients might have conversations and discuss data gathering methods without actual travel.

It is apparent that once the expanded benchmarks start being created, continuous collection of data and continuous updates will become part of the benchmark and assessment process.

## **Expanding the Sources of Benchmark Data**

Currently the actual personnel who provide data for both assessments and benchmarks are primarily software engineers and technical workers, project managers, some higher-level managers, and occasionally executives at the level of CIO or CTO.

Once the value of the expanded benchmarks becomes apparent, it can be anticipated that additional information might be collected from a much wider variety of stakeholders, executives, software personnel, and actual users of software:

### **Table 2: Executive Sources of Benchmark Information**

1. Corporate CEO's
2. Boards of directors or advisory boards
3. Government executives (state and Federal CIO's, agency chiefs, etc.)
4. Operating unit VP's (manufacturing, finance, etc.)
5. Agile embedded stakeholders
6. CIO's for information systems
7. CTO's for systems software
8. CTO's for embedded applications and hybrid systems
9. Customer executives who approve and acquire software
10. Outsource executives
11. Corporate attorneys
12. User association executives

In addition, technical information might be acquired from a larger spectrum of software technical personnel than is normal for today's assessments and benchmarks:

### **Table 3: Technical Sources of Benchmark Information**

1. Architects
2. Business analysts
3. Data base analysts
4. Software quality assurance (SQA)
5. Project Office personnel
6. Test personnel
7. Scrum Masters
8. Integration and configuration control personnel
9. Scrum masters
10. Embedded stakeholders
11. Software users and clients
12. Customer support personnel

A combination of remote interviews using simple questionnaire, more detailed questionnaires for on-site use, and conference calls would be used to gather the expanded

forms of information from an expanded set of software stakeholders and software production personnel.

The basic idea is to consolidate information on what stakeholders, top executives, development personnel, maintenance personnel, quality assurance personnel, and actual clients or users think about software applications and the methods used to create them.

The expansion of information sources will obviously cause some additional effort in gathering information. But after a few months of trials and tuning, hopefully the additional effort will not cause more than about a 25% increase in total data collection effort.

### **International Software Benchmarks by Industry**

As of 2011 there are benchmark studies and assessments performed overseas, and there are a few reports that compare performance by both country and industry. Of the ones that attempt to do so, sample sizes are small and the results are somewhat marginal in terms of economic breadth and reliability.

There is a very strong demand for reliable international benchmarks that would show the comparative performance of the countries where software production is a major economic topic. The probable revenues for international benchmarks would be proportional to the size of the software industries within the countries. Due to the logistical issues of carrying out on-site international benchmarks, the global market for such benchmarks would probably be three to five years behind the U.S. market.

That being said, the revenues from international software benchmarks by industry would probably be sufficient to fund expansion throughout the world. The more countries that provide data, the more valuable the overall collection of data will become.

Following are a few hypothetical examples of potential annual benchmark revenues by about 2020, assuming U.S. benchmarks as described in this report begin in 2012:

**Table 4: Probable Software Benchmark Revenues by Country Circa 2020**

China	\$45,000,000
India	\$45,000,000
Japan	\$35,000,000
Russia	\$30,000,000
Brazil	\$25,000,000
U.K.	\$20,000,000
Germany	\$20,000,000
France	\$20,000,000
Italy	\$15,000,000
Ukraine	\$10,000,000
Spain	\$10,000,000

Scandinavia	\$10,000,000
Australia	\$10,000,000
Mexico	\$10,000,000
South Korea	\$10,000,000
Canada	\$7,000,000
Taiwan	\$7,000,000
Israel	\$5,000,000
Netherlands	\$3,000,000
Belgium	\$3,000,000
<b>TOTAL</b>	<b>\$340,000,000</b>

Of course all of these countries could not be studied at the same time, but eventually the value of expanded global and industry software benchmarks has the potential to create a significant body of knowledge about software in every industrialized country and every major industry.

Over and above the countries shown in table 2 benchmarks might also be provided for many other countries in Asia, Central and South America, Africa, and the Pacific regions such as Singapore and Malaysia.

### **Effort Needed to Create Benchmarks by Industry**

Assembling the initial information needed to produce this example of an industry benchmark for banking would probably require about 90 days of full-time effort by at least one capable researcher. Two would be better in order to provide backup in case of illness.

The initial data collection would also require fee-based subscriptions to a variety of data sources such as Hoover, Gartner Group, the ISBSG, and other industrial information sources. However some of these sources of data provide short-term trials, so the initial expense would be fairly low for commercial data sources.

The probable initial cost of such an industry benchmark in a single industry such as banking would probably be in the range of \$150,000. This includes data collection from clients, commercial information providers, analysis of the data, and production of an initial report.

Later expenses would include development of a web site, marketing materials, and other collateral materials that are not absolutely needed for the initial benchmark report.

Because the data in such a benchmark is dynamic and changes rapidly, continuous updates would be needed to keep the information current. Probably 36 days per year would be needed to refresh the information (i.e. 3 days per month per industry). Monthly or quarterly updates would be provided to clients.

The value of this kind of benchmark compendium would be high enough so that the benchmark report might be marketed at perhaps \$10,000 for the initial report and annual subscription fees for updates of perhaps \$2,500. Extracts and subsets from the report could be marketed individually for costs in the range of \$500. These would appeal to smaller companies within the industry.

Information would also be made available via the web, with some samples of the data being provided for free, but extended data being fee-based. The fees could either be on a per-use basis or an annual subscription basis.

Another possible deliverable would be a daily web site entitled “Software Daily News” that resembles the very useful and popular current website entitled “Science Daily News.” The science web site covers a wide range of scientific disciplines ranging from archeology through astrophysics and includes both news summaries and full-length articles.

For a major industry such as banking, a benchmark report of this nature might attract about 250 domestic banks and 75 overseas banks. It would also go to government agencies and major software vendors.

Each industry benchmark report would probably generate about \$2,500,000 in the U.S. and about \$750,000 abroad, or \$2,750,000. Recurring revenues would amount to perhaps \$400,000 per year per industry per year. If there were 10 industries supported the revenues would ascend to more than \$27,500,000 for initial subscriptions and more than \$4,000,000 per year in recurring revenues.

Eventually on a global level these new benchmarks might have more than 5,000 major corporate clients, more than 1,000 government clients, and a large but unpredictable set of ancillary clients such as law firms, other consulting groups, universities, and the like.

Obviously such benchmarks would be most useful for industries that have a large amount of effort in the area of software and data processing. These industries include but are not limited to:

**Table 5: Industry Candidates for Software Benchmarks**

1. Aerospace
2. Agriculture
3. Airlines
4. Automotive
5. Banking
6. Chemicals
7. Computers and peripheral equipment
8. Cruise lines
9. Defense
10. Education – university

11. Education – primary and secondary
12. Entertainment
13. Energy and oil
14. Governments – state
15. Governments – Federal
16. Government - municipal
17. Health care
18. Hotels
19. Insurance
20. Manufacturing
21. Open-source
22. Process control
23. Pharmaceuticals
24. Public Utilities
25. Publishing
26. Retail trade
27. Software
28. Telecommunications
29. Transportation
30. Wholesale trade

A full suite of such benchmarks for major industries would probably generate in the range of \$30,000,000 to \$50,000,000 per year from U.S. clients. A team of perhaps 10 researchers and 5 logistical support personnel, plus licenses and subscriptions for external sources of data.

If the company producing the benchmark reports also collected benchmark and assessment data itself, probably another 10 consultants would be needed. The data collection would probably generate about \$5,000,000 to \$10,000,000 per year.

Data collection could also be subcontracted to existing benchmark groups such as ISBSG, Software Productivity Research (SPR), the Davids Consulting Group, and the like. Marketing and sales personnel plus a small executive contingent would be needed. The total size would probably be close to 35 total personnel. However subcontracts for collecting benchmark data might be issued to more than 25 companies in more than 25 countries.

It is conceivable that within 10 years of the initial launch, the new form of benchmark might involve more than 300 subcontract personnel in more than 25 countries.

The annual cost for operating such the core group would probably be in the range of \$4,000,000 per year. However except for the \$150,000 investment for the initial report, the organization should be self-sustaining and profitable because no heavy capital investments are needed.

The organization might generate revenues for benchmark subscriptions from eventually more than 500 domestic companies and perhaps 300 overseas companies and perhaps 100 government agencies. Assuming three to five subscriptions per company revenues from benchmark subscriptions might be in the range of \$10,000,000 to \$15,000,000 per year, plus the initial cost for each subscription.

The consulting work of collecting data on a fee basis would probably bring in revenues of perhaps \$8,000,000 per year. Total revenues from all sources might total \$30,000,000 to \$50,000,000 per year.

By contrast as of 2011 software benchmarks are a niche industry with perhaps 20 U.S. companies collecting data, with combined annual revenues that probably are below \$20,000,000. This is because benchmarks are perceived by clients as useful but not essential.

The business idea behind the new form of benchmark is to elevate the importance from being perceived as useful to being perceived as essential. To achieve this goal, more and better information needs to be provided to clients than is currently made available.

As useful as today's benchmarks are for productivity and quality studies, the lack of context information limits the utility of benchmarks and restricts the potential audience.

### **Additional Tools and Services in Addition to Benchmarks and Assessments**

In addition to marketing benchmarks and performing assessment and benchmark consulting studies, the organization would also be positioned to market and perhaps develop several kinds of tools and ancillary products and services:

- Benchmark tools for data collection that clients might use to facilitate their own data collection and analysis.
- Predictive tools for estimating software schedules, efforts, costs, quality, reliability, maintenance, and enhancements.
- Special studies that evaluate important topics. For example large-scale studies that compared a number of development methods such as waterfall, iterative, object-oriented, Agile, RUP, TSP, XP, and others would be both useful to the industry and highly salable.
- Training in skills which the benchmarks demonstrate need improvement within an industry such as quality control, test case design, change control, legacy migration, and a number of others.
- Portfolio analyses are seldom performed because of the difficulty of sizing and analyzing as many as 3,000 applications that might total more than 7,500,000 function points. Further, between 25% and more than 50% of typical portfolios

are in the form of COTS packages or open-source software which cannot be sized using standard methods because the vendors do not provide the needed inputs. It is possible to size a portfolio using one or more of the new high-speed function point sizing methods. Portfolio studies would no doubt be covered by non-disclosure agreements and hence marketed only to specific companies.

- Litigation support depends upon accurate data to determine industry averages for quality, productivity, and other benchmarks. The kinds of data discussed herein would probably be widely used as background information in breach of contract litigation between clients and outsource vendors. It would also be used in litigation against software companies for poor quality or damages.
- Outsource contracts frequently include clauses dealing with quality, schedule, productivity, reliability, change control, and the like. Without accurate benchmark data some contracts contain clauses that are probably impossible to achieve. Accurate benchmarks would be very useful for developing outsource contracts that are mutually agreeable.
- Mergers and acquisitions occur almost daily in many industries.. The data in this new form of benchmark would be of significant interest to business brokers and companies considering either acquisition or being acquired. It would also be of interest to companies that are merely seeking partnerships, distributors, subcontractors, or potential clients.
- Venture-backed start-up businesses have been in decline due to the recession, but are starting to ramp up again. The data contained in this new form of benchmark report should be of some interest to both entrepreneurs and venture capitalists considering starting new businesses.

Tool and special study revenues would be over and above the revenues already discussed. They are currently unpredictable because the suite of tools is not fully defined. However annual revenues in excess of \$10,000,000 per tool would not be uncommon. Special studies could easily top \$ 35,000,000 per year.

### **Benchmark Company Exit Strategies**

Normally companies have four possible end games: 1) they go public; 2) they are acquired by larger companies; 3) they go out of business 4) they continue indefinitely with only marginal growth..

Options 1 and 2 are the most likely end games for such a benchmark organization as the one described here. While such benchmarks might be created by a non-profit group or perhaps by a university, it is more likely that a for-profit organization would be the best choice.

A for-profit company is most likely because if the idea of the new form of benchmark expands and becomes successful, the benchmark production group would be an attractive acquisition candidate for large data providers such as Gartner Group, Accenture, Google, or similar large corporations where information is a valuable commodity..

### **Software Benchmarks Circa 2011**

As of early 2011 none of the normal benchmark sources or consulting companies provide this kind of information so there is no real competition. Gartner Group and the International Data Group (IDG) provide subsets of the kinds of information discussed, but not actual benchmarks.

- The International Software Benchmark Standards Group (ISBSG) provides useful benchmarks, but does not have any data on large applications > 10,000 function points in size.
- The Software Engineering Institute (SEI) provides assessments, but is sparse with benchmark data and provides little or no information about companies, industries, data, and other key topics.
- The Information Technology Metrics and Productivity Institute (ITMPI) provides many useful reports and webinars on specific topics, but does not provide assessment and benchmark information or any context information about various industry segments. Some of the CAI tools might well be useful for collecting benchmark data.
- The government-sponsored Data Analysis Center for Software (DACs) provides useful information in a government and defense context, but no benchmarks or assessments.
- The Standish Group publishes interesting statistics on software failures, but does not provide conventional benchmarks and assessments.
- A number of privately held consulting companies such as the Davids' Consulting Group, Quantitative Software Management (QSM), Software Productivity Research (SPR), and the Software Improvement Group (SIG) in Amsterdam and several others provide benchmarks and assessments for individual clients. These groups occasionally publish studies using data from multiple clients, but sample sizes are fairly small.
- Universities tend to provide small-scale studies on specific topics but are not funded or equipped to produce large-scale industry-wide studies.
- The software metrics associations such as IFPUG and COSMIC provide the current rules for counting functional metrics, but seldom produce benchmarks and they don't do assessments at all.

The bottom line is that the new kind of benchmark discussed in this report has little competition circa 2011. Gartner Group is the best positioned to compete, but to date has not attempted specific software benchmarks or assessments. Gartner Group aims at CEO's and top executives, but does not get down to the level of assessments and benchmarks.

It is apparent that the overall benchmark that contains all forms of data shown in this report might be subset into special purpose reports that might be marketed separately or offered via subscription to specialized communities. Examples of such special reports might include, but are not limited to:

**Table 6: Special Benchmark Reports**

1. Software Quality Assurance (SQA)
2. Software Estimation Tool Analysis
3. Software Development
4. Software Maintenance and Enhancement
5. Software Security and Safety
6. Software Failure Rates by Industry
7. Software Project Management
8. Software Project Offices
9. Data Quality by Industry
10. Data Base Development
11. Web Site Development
12. Software Process Improvement
13. Software Education and Training
14. Analysis of Cloud Computing
15. Best Practice Analysis Based on Empirical Results

These separate subsets would clearly generate additional revenues over and above those discussed for the entire assessment and benchmark report. However it is premature to attempt to quantify the numbers of subscribers and revenues for these subsets of information.

**Summary and Conclusions about Software Benchmarks**

The software industry circa 2011 is characterized by many project failures, by frequent cost and schedule overruns, by poor quality, and by lack of reliable information as to what the phrase “best practices” really means in terms of results.

The software industry is also subject to frequent fads and fallacies as new development methods surface, are touted as panaceas, and then gradually demonstrate only marginal improvements if any over alternate development methods. Poor measurement practices and inadequate benchmarks are what makes these fads and fallacies endemic problems for the software community.

Benchmarks are very useful methods for minimizing these common problems, but less than 30% of large U.S. software companies have either commissioned benchmarks or use benchmark data. For small and mid-sized U.S. companies, less than 10% utilize benchmark data. In fact many corporations not only fail to use benchmarks and assessments, they have never even heard of the SEI, ISBSG, ITMPI, and the other major sources of assessment and benchmark information.

The new form of benchmark discussed in this paper is aimed at expanding the information contained in software benchmarks from basic productivity and quality levels up to a complete condensation of critical industry topics where software is part of the equation. The additional kinds of data and information will hopefully elevate benchmarks and assessments from useful but optional studies to mandatory business practices that are demanded by a majority of CEO's and top operational executives.

Once such information is published for a specific industry such as banking, it is expected that demands from other industries will drive the business growth for similar benchmarks in other industries.

The goal for the new form of benchmark is to reach close to 100% of major corporations, more than 50% of medium corporations, and perhaps 25% of smaller corporations. It is difficult to predict government penetration at all levels, but no doubt all 50 states would want to subscribe to this new form of benchmark if state government data were included. A number of Federal agencies would also want to have access to the kinds of data provided by the new benchmarks.

It is hoped that the kinds of information included in these new benchmarks will not only lead to a profitable business, but will assist the software industry in overcoming its traditional problems of cost and schedule overruns combined with poor quality.

