



The QSM Database

An In-Depth Look

An In-Depth Look at the QSM Database

The QSM Database is the cornerstone of our tools and services. We use validated quality and performance data (size, time, effort, defects, and productivity) collected from over 14,400 completed software projects to keep our products current with the latest tools and methods, support our benchmarking business, inform our customers as they move into new areas, and develop better predictive algorithms. Often, our clients and prospects ask for more information regarding the data and the types of projects represented. This whitepaper addresses some frequently asked questions.

Data Sources

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 formats that preclude the identification of individual projects or companies and the disclosure of sensitive business information.

In 1994, QSM began collecting project data continuously, typically updating the database every 2-3 years. Over the last 5 years, we have added an average of 300 validated projects each year.

Project Metrics

Our core metric set focuses on size, time, effort, and defects for the Concept Definition, Requirements & Design, Code & Test, and Maintenance phases. These measurements are supplemented by nearly 300 other quantitative and qualitative metrics. Almost all (99.9%) of our projects have time and effort data for the Code and Test phase. while 72% provided time/effort data for both the R&D and C&T phases. By contrast, only 37% of projects report time and effort data for the Concept Definition and Maintenance phases.

Delivered size data is recorded for 97% of projects in our database, with defect data recorded for 44% of projects. 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 Defects (System Integration test – Delivery)
- [MTTD](#)
- Defects /Month
- Defects /Effort Unit (various)
- Defects /KESLOC or Function Point
- Post release defects at 1, 3, 6, 9, 12, and 18 months

Data Usage

Data from our database is used to:

- Keep our tools and algorithms aligned with the latest industry practices and technologies.
- Create 800+ industry trend lines and single metric statistics included with SLIM-Suite applications.
- Support our basic measurement, benchmarking, estimate validation and verification and other consulting services.
- Conduct ongoing research available via books, papers, articles, and blog posts offered on our web site.

Data Availability

To protect our data contributors, we release project data in summary form only. QSM consulting and tools clients may receive graphs and summaries that compare their projects against industry trend lines as well as charts showing actual projects that are similar in size, application type, and complexity. Individual project data is not disclosed. Summary statistics and information about the Putnam Model are also available in books like [Five Core Metrics](#), [Measures for Excellence](#) and [Industrial Strength Software](#), [QSM webinars](#), [QSM IT Almanacs](#), [blog posts](#), [technical reports and whitepapers](#), conference presentations, the QSM [function point language table](#), QSM [performance benchmarking table](#), and journal articles.

Kinds of Data

Actual project schedule, effort, size, productivity, and quality data is available for projects ranging from 1-454,000 total (new, modified, and unmodified) function points and less than 100 - over 24 million total SLOC. Our core metric set is supported by nearly 300 other quantitative and qualitative measures. Custom metrics and user defined variables allow QSM clients to capture measures specific to their organization.

Size Data

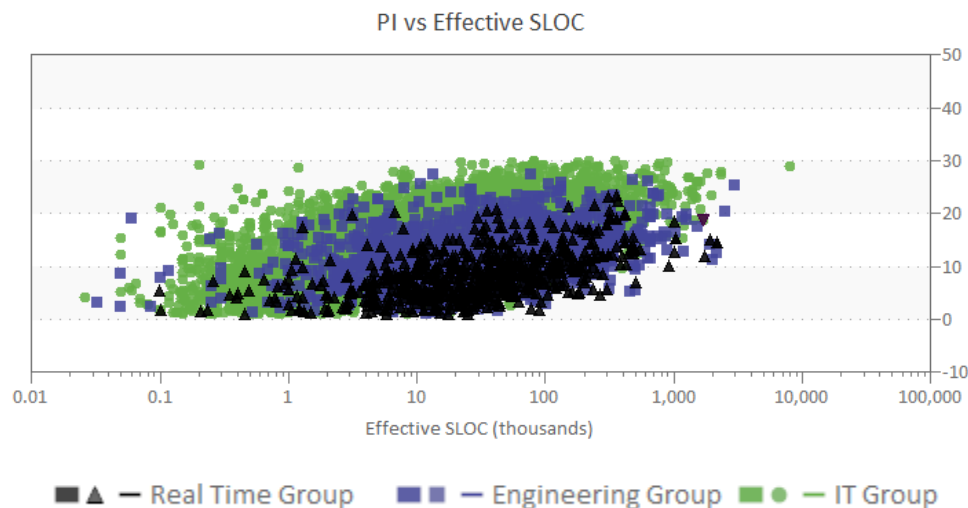
QSM's innovative approach to capturing reused, modified, and newly developed functionality allows teams to express project size in a wide variety of sizing components and methods such as logical lines of code, function points, Agile size measures (stories, epics, story points), features, requirements, RICEF objects, and use cases. Projects can report just newly developed functionality or capture new, modified, and even pre-existing (reused) functionality if a more granular view of the work is required.

Volume of Data

Over 14,400 projects from 1978 through the present. To facilitate longitudinal studies of the industry, older project data is retained.

Industry Data

QSM data is stratified into three high-level domains (Software, Hardware, and Infrastructure), twelve application types (Avionics, IT, Command & Control, Microcode, Process Control, Real Time, Scientific, System Software, Telecom, ASIC, Electronic, and Systems) and 46 sub-types. Software projects predominate, but we have a growing number of hardware and [IT infrastructure](#) projects as well. Real Time, Avionics, and Microcode projects typically exhibit the lowest productivity, taking longer to develop than Business/IT projects that automate well established and/or less algorithmically complex processes. Over time, this productivity gap has narrowed somewhat due to the increasing complexity of the target platforms used by modern IT programs.



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 the major industry and industry sector for which the project was developed. Major industries include the financial sector, banking, insurance, manufacturing, telecommunications, transportation, systems integration, medical, aerospace, utilities, defense, and government.

Industry Trends

QSM industry trend lines are available for nine high level application domains, six application subgroups, and three application super groups. The nine application domains are:

Business	Command & Control	Scientific
System Software	Telecommunications	Process Control
Avionics	Microcode/Firmware	Real-time Embedded

Stratifying the data by application type reduces variability at each size range and allows for more accurate curve fitting. One application domain, Business IT projects, has been further stratified into several sub-groupings:

Business Agile	IT Cloud Migration
Business Government	Web Systems
Business Financial	Package Implementation

Three application super groups are also available to benchmark projects of mixed or unknown application domains:

Real Time Group	Engineering Group
All Systems	

Lifecycle/Development 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).

Programming Languages

More than 700 languages are represented in the database. About 40% of projects reported just a single language. 38% reported multiple languages. Common primary languages are JAVA, COBOL, C, C++, C#, VISUAL BASIC, .NET, IEF / COOLGEN, PL/I, ABAP, SQL, PL/SQL, POWERBUILDER, JAVASCRIPT, PHP, and COLD FUSION. Frequently reported secondary languages include COBOL, JCL, SQL, PL/SQL, SCL, JAVA, ASSEMBLER, C++, EASYTRIEVE, PHP, INFORMIX, XML, WEBLOGIC, VISUAL BASIC, XML, ASP.NET, and JSP.

Country Data

QSM has collected and analyzed software projects from North America, South/Central America, Europe, Asia, Australia, and Africa. About 54% of our data is from North America. Another 34% is from Europe, with the remainder developed in South America, Asia, and Australia/New Zealand.

Future Data

We are currently expanding our data collection to include non-software and hardware projects, various cloud migration strategies, and institutional or portfolio level performance measures.

Summary

As pioneers and innovators in the software estimation field since 1978, Quantitative Software Management is more than our company name - it's the philosophy that defines everything we do. Our goal is to help clients use proven business analytics to deliver world-class software solutions on time and within budget. QSM provides superior [estimation products](#) and [consulting](#) that enable our clients to [estimate](#), [track](#) and [benchmark projects](#) with confidence. QSM tools and methods can help you negotiate achievable goals, set realistic expectations, and communicate more effectively with your colleagues and customers.

Headquartered in Washington, DC with offices across the U.S. and partners around the world, QSM has established itself as the leading total solution provider of business analytics tools, consulting, research, and education for software developers in high-performance, mission-critical environments.

Kate Armel is the Director of Testing, Training, & Technical Support at QSM. She has 24 years of experience providing technical and consultative support for software estimation, project tracking and forecasting, and industry benchmarking. She also oversees data collection, validation, and data analysis for the QSM database; development of over 900 industry regression trends and statistics; delivery of client, internal, and technical support services; software testing and quality assurance; training, documentation and online help for SLIM-Suite and SLIM-Collaborate applications, APIs, and utilities; and technical writing, research, and analysis to support product development, research, and consulting services. Ms. Armel was the Chief Editor and analyst/co-author of the 2006 QSM IT Software Almanac and has authored numerous published articles.