Benchmarking Guide for Software Development and Maintenance Projects

Version control

Date	Reviewer(s)	Modifications / Additions
November 2016	See below	Version 1.0

Acknowledgements

Version 1.0 authors and reviewers 2016 (alphabetical order)		
Mauricio Aguiar TI Metricas, Brazil	Diana Baklizky TI Metricas, Brazil	Tom Cagley DCG Software Value, USA
Dácil Castelo Leda MC, Spain	Rafael de la Fuente Leda MC, Spain	Raúl Fernández Leda MC, Spain
Mike Harris DCG Software Value, USA		

Index

1 Introduction

It is very difficult to quantify and track the performance of any IT department in terms of productivity and quality. To that end, it is even more difficult to compare our data with our company's competitors in the market because the metrics used by each company may vary widely. Even if we successfully implement a program to track productivity and quality on a routine basis, it can still be difficult to answer questions like: *Are our software factories / teams more productive now than before? Are we within the market average? Are we paying more than what we should?*

The answers to these questions are neither reliable nor objective because the only available information is expressed in total costs, distribution costs, and cost rates.

Due to the lack of this information, IT management becomes a blind process, where decisions are often made without full awareness of their real consequences. For example, reducing the IT supplier's development price, without being aware of its impact on productivity, may increase company costs due to:

- Reduction in quality
- Reduction in productivity
- Increase in errors

The solution to this problem relies on two variables:

- Software production services must be measured (a standardized software development measurement methodology is recommended but not necessary).
- Information must be available for comparison.

The first point allows us to get normalized information in order to compare the results of our company both to the market and to ourselves (externally and internally). Concepts such as **productivity**, **cost per production unit**, **and errors per production unit** are impossible to obtain without software quantification. For this purpose, we can use standardized and non-standardized metrics. If we use standard metrics, information will be more reliable and we will be able to share information among different companies. In this sense, the ISO/IEC 14143 definition about functional size quantification is a big step to standardization and expansion of the functional size metrics. On the other hand, non-standard metrics can be less consistent and depend on the company's idiosyncrasies.

The second point enables us to better understand our position in terms of the market. Without this information, we could only produce internal comparisons and analyze our own evolution during different periods of time. Without an objective analysis of the real world, we can't compare whether our results are better than, equal to, or worse than the market's average level, and we don't even know if our services are really efficient. Of course, "averages" can be dangerous so it is important to understand the nature of the population over which the average has been taken. For this purpose, private companies and other organizations such as ISBSG provides data about the average and other related services based on information about the desired population for comparison.

2 Objective

The objective of this document is to explain the benchmarking process, its applications, the main guidelines for its execution, and the necessary processes that must be established in order to guarantee successful results. This is only intended for use in the software development field.

The present guide is not an exhaustive compilation of all the activities and required tasks but a general guide about what you have to take into account and consider from a practical business perspective.

2.1 Intended Audience

This document is primarily intended for those involved in IT management with a specific interest in optimizing value for money from their software development.

2.2 Scope

Only software development and maintenance projects are under the scope of this document, regardless of their development methodology (waterfall, Agile, etc.).

Therefore, any project not involving a software development process is excluded, for example, consulting projects, operating system upgrades, etc.

3 Practical Applications

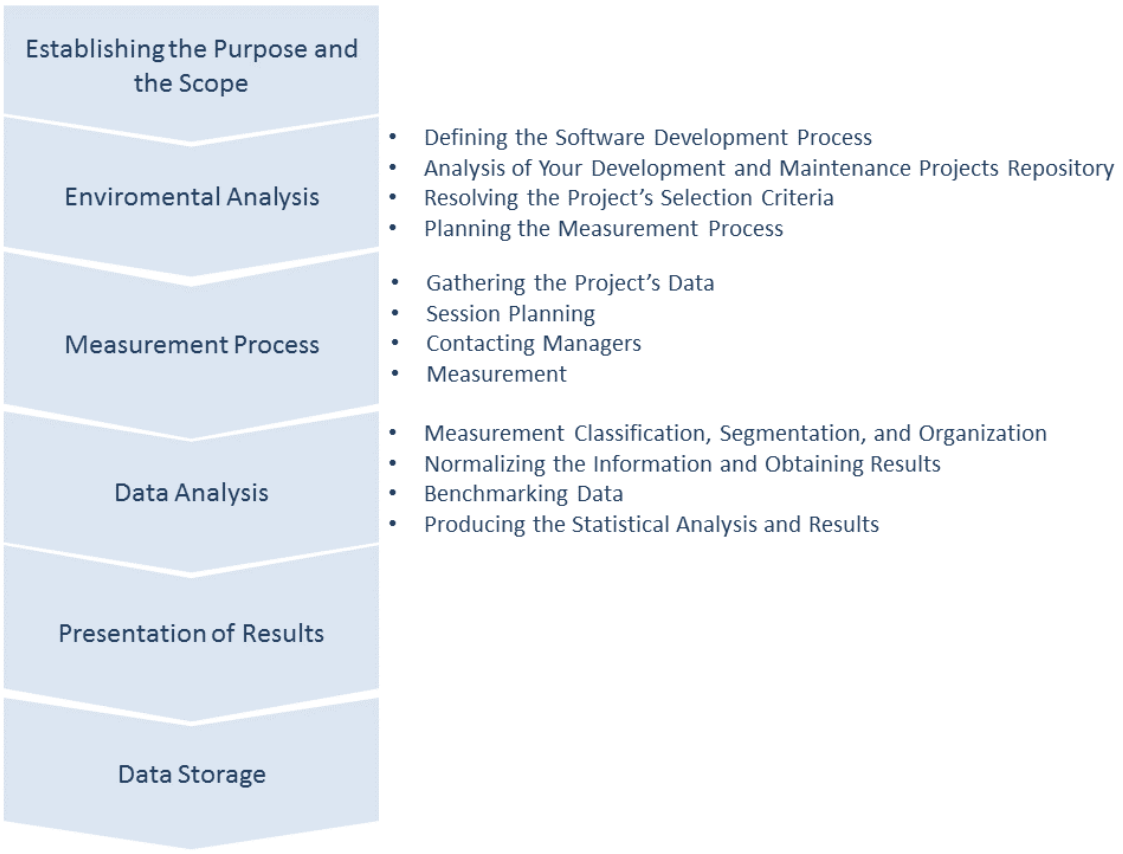
There are multiple applications for the information obtained from benchmarking in terms of productivity, quality, effort, rates or costs, including, but not limited to:

- Comparisons between our company and other companies:
  - My company versus the global market or some companies of the same type of industry..
  - Some specific technologies or business areas versus the market ones.
- Comparisons within my organization, between different departments, suppliers, technologies, periods, etc.
- Establishing reference values:
  - Reference values to make more objective project estimations.
  - Reference values to establish Service Level Agreements (SLA) for third parties.
  - Reference values to establish internal or third-party improvement objectives.
  - Reference values to produce control, continuous follow-up, and improvement of this same reference values.
- Decision making:
  - Evaluate the impact of any adopted decision: supplier modification, price variations, changes on the development process, changes on the implemented technology, etc.

4 Procedure

Benchmarking requires a series of activities summarized in this chapter. The objective of this section is to provide a general description of the activities involved in executing the benchmarking process.

Each of these activities is a step in a sequential process that must be followed to ensure the optimal quality of the results for the system's object of study.



Benchmarking Process Diagram

Note: As of June 2016, the ISO (International Organization for Standardization) and the IEC (International Electrotechnical Commission) are developing the standard ISO 29155 Systems and software engineering - Information technology project performance benchmarking framework, which is a framework for benchmarking software development projects within the IT field. This new ISO, which is still in process of development, could be in a future useful in order to standardize the software benchmarking process and ensure good practices.

4.1 Establishing the Purpose and the Scope

The benchmarking purpose will determine the objective of the study. According to the comparison options indicated in chapter 3, this could be:

- Comparison within our own company

- Comparison of our company with other organizations (competitors)

Depending upon the target of the study, the purpose may change:

- Productivity study, for instance, in order to know the effort per unit of production of my java developments
- Functional quality development study, for instance, in order to know the number of defects per unit of production of the sw delivered by my suppliers as detected by a Quality Assurance team
- Technical quality development study, for instance, in order to get the number of code violations per unit of production delivered by my suppliers

it is also necessary to determine the scope of the study. In this activity, the number of projects for the sample should be specified, as their size, characteristics (technology, sw lifecycle or any relevant characteristic for the study), involved areas or departments, etc., so there could be some restrictions regarding the scope:

- A study of specific technologies or development platforms
- A study of selected development suppliers
- A study to define the stages of lifecycle development

Some examples of the purpose and scope of the benchmarking exercise could be:

- Knowing the development productivity of java and cobol languages compared to the market for Banking Industry
- Knowing the development productivity of the implemented projects under the traditional waterfall model (or an Agile model) during the last year
- Making a comparison between the functional quality of the software developments delivered for my main suppliers during the last year.

4.2 Environmental Analysis

In order to execute a quality comparison of the data, the first step is to gather and study our company's information. The software development process is different for each customer; however, this process is the same among the different suppliers or domains within the same customer.

For this reason, we need to know the business environment where the development is executed.

4.2.1 Defining the Software Development Process

The development process could involve several technologies or domains depending on how IT in the company is structured. We have to focus solely on the activities and / or technologies defined within the scope.

At this point, the primary items to be analyzed are:

- Development methodology/methodologies used

- Activities or lifecycle stages (requirement analysis, functional design, technical design, implementation and unit tests, integration tests, etc.), main development technology or architecture. *This is a critical factor for success.*
- Business domain (banking, insurance, processing, logistics, human resources, etc.).
- The roles and persons participating in any of these stages, and which of these are (partially) outsourced, etc.
- The source of the model (internal, outsourced, etc.)

For instance, if our study tries to focus on aspects of third parties (suppliers), we need to know which tasks are executed in order to compare them individually and not confuse them with internally developed tasks.

4.2.2 Analysis of Your Development and Maintenance Projects Repository

Not all projects defined in the domain or within the organization can be used in the study. Certain projects, because of their nature or critical circumstances of their development, are so uncommon that they would create "noise" (could be considered as statistical outliers). For this reason, some development projects are excluded from the study because they are considered "uncommon"; for example: projects not completing the fully defined software lifecycle, exceptionally large projects or very small ones (related to the average domain or organization's size).

At this point it's very important to take into account the purpose and scope of the benchmark study, because the organization could include only "normal" projects and this could be wrong if, for instance, the purpose was to know the real productivity of the whole project portfolio.

We need to perform an evaluation of your projects repository. By doing this, we can distinguish which projects can be used in the study and which ones must be excluded because of their specific circumstances. In this evaluation, every project selected for the study should have the necessary information for the analysis (effort, technology, defects, documentation, etc.).

For each project, we need to verify the following information:

- The availability of all the required information to perform the study;
- That any specific or unique situation affecting the results of the study have been excluded;
- That each project is not an exception within the "normal" project's domain or the organization.

4.2.3 Resolving the Project's Selection Criteria

The projects meeting the selection criteria will be used as a sample for the benchmarking study. Depending on the purpose and scope, you could select a homogeneous sample or a heterogeneous one, but this decision has to be explained in detail for future reference.

Also, it is necessary to make a selection of a representative number of projects for any analysis segmentation. The number of projects selected is very important for the reliability results of the study.

For instance, to execute a benchmarking study between three different technologies, select a similar number of projects for each of them. Likewise, the selection criteria for these projects, regardless of the segment, should be as equivalent as possible.

Some of the criteria that should be used for obtaining the sample are:

- *Classification of Development*. If there are projects with both functional and non-functional requirements, and projects only describing functional requirements, one could prioritize those with only functional requirements based on the objective of the study.
- *Dedicated hours*. Sometimes it could be necessary to select projects with the same dedicated effort (or those within a specified range), but not if, for instance, it is required to compare productivity between large and small projects. Always look at the purpose and scope of the study.
- *Date of Development.* If you want your results to reflect the current state of your organization, consider only recently developed projects and ignore the older ones. On the other hand, if you want to compare the current productivity to the productivity from a past period, select projects from both periods.

Because of the nuances of the information you are collecting, it is imperative that the selection criteria is defined in the best possible way for repeated use.

4.2.4 Planning the Measurement Process

Once we have selected the projects for the study, we can begin the process of measuring each project's size. When scheduling this exercise it is important to consider the following information:

- When will the project documentation and information be available?
- When and by whom will the measurements be performed?
- When will the revision sessions start (if necessary)?

This schedule should be provided to those involved in the projects' development, in order to perform the measurement process and meet expected deadlines.

4.3 Measurement Process

During this step, we will measure the size of the selected projects according to the determined measurement method.

As stated in the introduction to this document, it is necessary to have a software quantification technique. For this reason, it is expressly recommended (but not necessary) to use functional metrics compliant with the standard ISO/IEC 14143-1:2007 *Information technology – Software measurement – Functional size measurement – Definition of concepts.* These metrics are (only a few are significantly used in the market, for example: IFPUG and COSMIC):

- NESMA FP
- IFPUG FP
- COSMIC FP

- FiSMA FP
- MK II FP

The metrics for functional size are based on the software quantification in terms of what is (requested to be) delivered. It is exactly for that reason that they constitute the ideal tool for analyzing productivity or economic efficiency. It also allows us to get a normalized metric for other variables of software management (effort, costs, defects, duration, etc.).

It is important to remember that the functional size results obtained by applying these methodologies are not exactly the same. Hence, you must use the same technique when sizing projects in order to compare results to those of other companies.

Organizations that provide benchmarking data indicate the metrics that were used in their submitted data. Note: NESMA and IFPUG apply similar principles and are equivalent within a limited variance.

Using a non-standard metric when measuring the size of a project will create a useful data comparison within our company, but we can't compare the obtained results to the market. We also have to be very careful when using internally developed or proprietary sizing methods created by organizations that promote those methods as being similar to the market's standard methodologies.

The use of non-standard metrics, without independent empiric evidence, might create other problems, such as measurement subjectivity or even its invalidity in meeting the objective financial efficiency and/or productivity (e.g., using Lines of Code to compute economic productivity is not recommended).

4.3.1 Gathering the Project's Data

The projects will be measured during personal meetings or by analyzing the functional documentation. Any required information and documents should be provided for supporting the projects' functional size measurement activity.

Required documents include:

- Requirement analysis reports
- Functional designs
- Technical designs
- User manuals
- User guides, etc.

These documents will clarify the functionality delivered to the users by the implemented systems. The most useful documentation for this purpose is the one explaining, in the clearest way possible, the user functionality delivered by the software (functional analysis or documents focused on the developed functionality).

In addition, it is necessary to compile the following project information:

- When developing a productivity benchmark, **the effort or cost broken down by relevant variables such as** lifecycle activity, internal/external parties, impacted application, or development supplier. This is often a problematic task.
- When performing a functional quality benchmarking, **the associated defects** (in pre-production or post-production).
- When producing a study about the delivery speed, **the milestones of the project development process** (start and end dates of the phases).

4.3.2 Session Planning

If the documentation requires any explanation by the project manager or any functional or business analyst expert, schedule a meeting session for discussion and clarification.

4.3.3 Contacting Managers

Once the sessions are agreed upon and scheduled, the experts will be contacted for the revision of any pending topics and / or related clarifications.

4.3.4 Measurement

With the documentation available and all questions answered, it will be time to size the projects. This will be accomplished by applying the selected method.

It is mandatory that the measurements are performed by (certified) metrics specialists, to increase certainty regarding the application of the method. In addition, the measurements should be well documented, including any observations and considerations, people interviewed, documentation used, etc., so that any other specialist will be able to review, understand, and audit the measurements in the future.

4.4 Data Analysis

4.4.1 Measurement Classification, Segmentation, and Organization

To obtain the required data, the projects need to be grouped according to the objective of the benchmarking process. So, when performing a study by technology, all projects of the same technology are grouped together. If the study is performed by technology and supplier, the projects are grouped together as a combination of technology-supplier.

It is important to recall that, as with any statistical study, the number of measurements in each group must be large enough to ensure the validity of the final results.

4.4.2 Normalizing the Information and Obtaining Results

In order to properly compare the results of the different groups internally (within our company) or with the market, it is necessary to make a comparison between the same concepts (“apples” to “apples” and not “apples” to “oranges”), so the data is aligned with the study objective. This task is the most sensitive activity in the process, and the way it is performed determines the validity of the results.

The normalization process ensures that the effort, cost, and any other project information considered in the study is of the same nature as the information gathered on the selected projects in the comparison group. This is independent of whether the comparison is to

competitor projects, internal projects, or to a previously performed benchmarking. It is necessary that the basis of comparison is clearly stated and meets the purpose of the study.

For instance, if itis required to compare the productivity of the software Design, Built & Unit Test activities, it is necessary to take only into account this effort and discard the effort of the other lifecycle activities.

Before performing the normalization process, the available data from both our organization and the market samples must be carefully analyzed. This analysis will determine which projects or groups need to go through the normalization process.

4.4.3 Benchmarking Data

As mentioned, the second basis for performing a good benchmarking is the existence of an external database that includes useful information about a variety of development projects, facilitating comparison. The access to this external data provides a general overview of the world's situation, the market's behaviour and its upcoming trends.

4.4.3.1 Private companies

Nowadays, the best option for benchmarking data analysis is to work with a private company. These companies offer localized data for certain markets, and they allow a more specific and precise comparison. They also offer benchmarking studies as a professional service with a wide experience, so organizations can rely on them in order to get a complete benchmarking.

Some of these companies include: LEDA-mc (www.leda-mc.com), TI Métricas (www.metricas.com.br), and DCG Software Value (www.softwarevalue.com).

4.4.3.2 ISBSG

ISBSG was founded in 1997 by a group of associations of software metrics. Its objective is promoting the use of IT industry data for software improvement.

It is a non-profit organization that keeps two main data repositories:

- One for development and maintenance projects including more than 6,000 projects developed in more than 30 different countries. This repository contains information on the size, phases, effort, technology, defects, etc. of projects.
- Another for maintenance and support applications including more than 1,000 applications (information about support activities, efforts, defects, hardware, etc.).

All ISBSG project data has been submitted by volunteers from several different organizations. Naturally, additional variation will be introduced by this diversity.

Anyone can register and purchase the mentioned data repositories and other available products and reports on the ISBSG website.

ISBSG doesn't offer benchmarking studies as a service. It only provides data and conclusions from its analysis.

4.4.4 Producing the Statistical Analysis and Results

With all the available data gathered, it will be time to execute the required statistical calculations to determine the values, the trends, and their accuracy. After this analysis, initial conclusions can be made.

For a reliable analysis, review the anomalous data in the set. The objective of this revision is to determine if these values are consistent with the real situation and the specific characteristics of the development process within the organization. On the other hand, if any project is considered as an exceptional one due to the nature of its development, it should be excluded from the study.

This analysis is not an easy task for inexperienced consultants. It requires statistical skills in order to explore and analyze all the data collected from different perspectives. Some of the required analysis includes data investigation and analysis, correlation analysis, analysis of variance, regression, and many more statistical methods. Of course, a statistical tool is necessary to perform this analysis, that should be performed by a professional statistician, or by someone with equivalent knowledge and experience.

The results must be statistically representative, and the experience of the consultant in this kind of analysis is another critical success factor. Those affected by the benchmarking study often try to refute these results (even more so if they are questionably created). As such, the study has to be rigorously defended by an experienced and qualified professional.

4.5 Presentation of Results

Once the study is performed, present the results to relevant stakeholders. Be prepared to defend the findings.

With the initial data obtained in the study, we can generate a preliminary version of the report. This report shall introduce clear information about:

- The objective of the study;
- The scope (projects, technologies, etc.);
- The sizing metrics (NESMA, IFPUG, COSMIC, etc.);
- Information on the data (source, # of projects, size range, industries, age and other descriptive information when available)
- The details about how the results were calculated;
- Information about the process, participants' collaboration, the quality of the reviewed documentation, etc.;
- Conclusions;
- Recommendations, based on the results.

It is important to review this report carefully, including the results and its conclusions. After this revision, draft a final report that you will present to the most relevant people involved in the project.

A very short, executive summary, should also be produced.

During this phase, it is very important to have quality data interpretation and communication skills. A highly experienced consultant is often imperative in order to most effectively report on and defend the results.

4.6 Data Storage

Knowing our positioning in the market not only satisfies our personal curiosity, but also provides very useful decision making information and reliable input data for potential changes in our organization's software development process.

In order to ensure that improvements have been made following any changes, perform a new study. This study will prove that the chosen actions have been effective and the proposed objectives have been accomplished, increasing the achievements of our company.

Consequently, it is important to select and store the data in such a way that allows us to effectively collect comparable data in future studies.

5 Critical Success Factors

In order to get the expected results of a benchmarking study, we must consider these critical success factors:

- Trained specialists:
  - In software measurement;
  - In software processes;
  - In software data analysis (activities, effort and cost, etc.);
  - In statistics.
- Global and local reference database access
- Experience, knowledge and skills to properly interpret the results, understanding its implications and transmitting them in an efficient manner.

6 Conclusions

Benchmarking is a necessary and powerful tool for quality IT management. It offers very useful information for making decisions and for improving and tracking the development process. On the whole, benchmarking serves as a navigational chart that allows us to take the helm and chart the right course.

The more benchmarking is understood and accepted, and the more it is required by companies, the more useful it will be and more reference information will be available.

Therefore, the existence and support of organizations either not-for-profit as the ISBSG, or for-profit private enterprises, offering an impartial analysis of this kind of information, is vital.

7 Related Publications.

[1] ISO 29155 Systems *and* software engineering - *Information technology project performance benchmarking framework.*

[2] Nesma, Netherlands Software Measurement user Association. “Definitions and counting guidelines for the application of function point analysis, a practical manual.” Version 2.1, 2004, www.nesma.org (ISO/IEC 24570)

[3] IFPUG, International Function Point User Group. Function Point Counting Practices Manual. 4.2. 2004. http://www.ifpug.org (ISO/IEC 20926)

[4] FiSMA, http://www.fisma.fi/in-english/ (ISO/IEC 29881)

[5] ISO/IEC19761:2003, Software Engineering “COSMIC, A Functional Size Measurement Method version 2.1, International Organization for Standardization” ISO, Geneva, 2003, www.cosmicon.com.(ISO/IEC 19761)

[6] ISO standard for Functional Size Measurement Methods, ISO/IEC 14143,

[7] Practical Software Project Estimation 3, Peter Hill, ISBSG,

[8] ISBSG repository “New Developments and Enhancements”, www.isbsg.org.