TURKEY SOFTWARE QUALITY REPORT

Test Data Management
2016–17

CONTENT

FOREWORD

Turkish Testing Board (TTB - turkishtestingboard.org) is pleased to bring you the 2016-17 edition of the Turkey Software Quality Report (TSQR). TSQR 2016's focus is on test data management. Apart from traditional testing surveys which solely focus on the technical side of testing, we have tried to put emphasis also on the business side of testing. You will find tips and trends in the test data management area.

The report is designed to help organizations to make paradigm shifts in their mindsets. It not only draws a clear picture of the current situation in the Turkish market but also sets the de-facto standards and trends for future information technology (IT) projects. We hope this report will be a reference point for all decision makers.

With the help of TSQR, we are trying to lay down the foundations of a healthy discussion platform for improvement in Turkish IT market. TSQR will be presented at the opening ceremony speech of TestIstanbul 2016 (testistanbul.org) on April 26th initiating a series of keynotes, presentations and discussions.

Turkish Testing Board

EXECUTIVE SUMMARY

IT organizations are inevitably focusing on the collection and organization of data for their testing processes. The ability to control process and use test data has become the key competitive advantage for IT organizations, because the benefits gained by the implementation of these mechanisms will worth the investment.

Testing is a very critical part of high quality software development; nevertheless test data management gets only minimum attention in organizations. This report mainly provides practical recommendations (test data ownership, test data security, test data generation techniques, etc.) for test data management initiatives. The further aim of this report is to highlight the most essential needs of test data management process so as to be applicable to any IT organization attempting to build or improve its test data management methodologies.

As IT organizations deeply rely on data-driven test activities; accurateness and effectiveness of data is more critical than ever. Any misusage of them carries the risk of reputational damage, such as the loss of customer trust, therefore it raises the level of data confidentiality requirements in many industries.
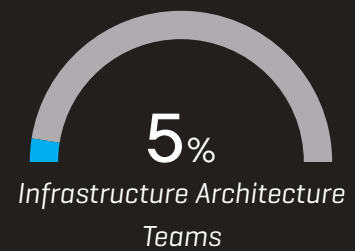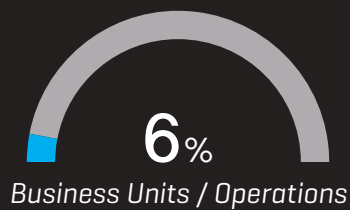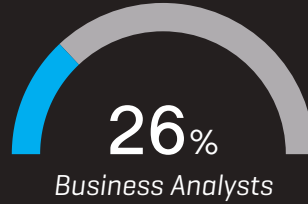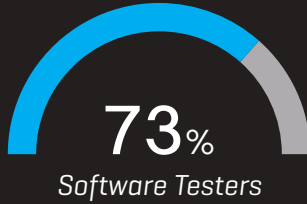
No matter which approach is chosen to overcome unique challenges of test data management in any IT organization, basic prerequisites to be successful are; combination of good test cases and test data, along with the proper usage of tools to help automate extraction, cleansing and management of the data being used.

You can access the softcopy of this, and previous reports at turkishtestingboard.org and hope to see you at TestIstanbul 2016 Conference on April 26th to discuss our further findings.

Turkish Testing Board

IN YOUR ORGANIZATION, **WHO IS RESPONSIBLE** FOR TEST DATA GENERATION?

* multiple selection was allowed

73%
Software Testers

26%
Business Analysts

22%
Software Developers

6%
Business Units / Operations

5%
Database Administrators

5%
Infrastructure Architecture Teams

4%
Third Parties

ANALYSIS OF THE CURRENT SITUATION

According to the results, software testers have a significantly larger proportion of data generation responsibility in their organizations. It is fairly predictable to see software testers' share as 73%, mainly because this role is also responsible for design and execution of test scenarios. On the other hand, test data generation is an integrated part of the test design process since specific scenarios reveal required data sets.

Business analysts and software developers seem to share a similar ratio of responsibility. However, it is worth to mention that their area of expertise would be different in data generation tasks. We may assume that business analysts are more familiar with the preconditions, requirements and the scope of user scenarios (and determine the required data specifications) whereas developers are better equipped to find quick and effective methods to actually generate the test data.

FUTURE PREDICTIONS

It would be fair to assume that similar ratios in test. With the involvement of more and advanced automation in test data management, testers with more coding and technical skills like technical testers and test architects will take place in the process.

2 HOW DO YOU GENERATE TEST DATA?

multiple selection was allowed

64% Manually populated data

40% Subset of production data

28% Full copy of production data

17% Database virtualization

ANALYSIS OF THE CURRENT SITUATION

Manually creating necessary test data seems to be the most widespread method chosen with a 64% ratio. Even though all manual processes take longer time than others, they may be considered as quick-and-dirty methods, since test data are reviewed by a real person and usually created from scratch. It may also be easier to immediately start generating the test data along with the test scenarios and even just before test execution. However, what usually prevents organizations from using some form of automated procedures, tools or virtualization for test data generation is the lack of expertise or resources within the teams.
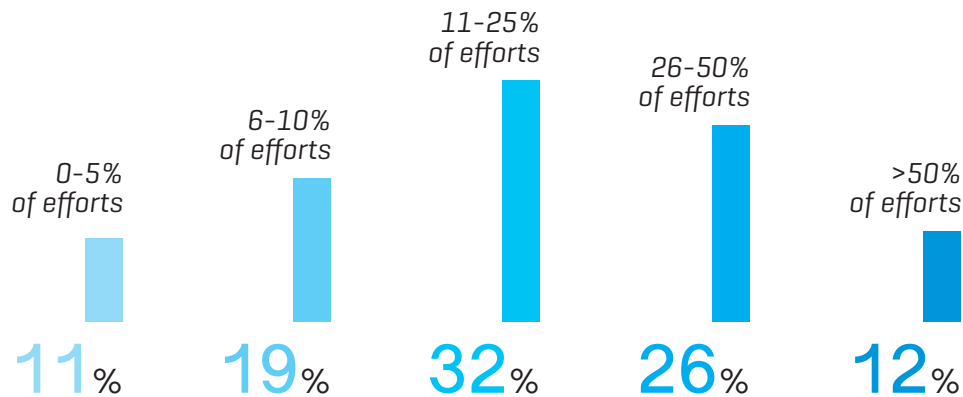
Taking a subset of production data is also popular with a 40%. When production data is used as test data, (whether it is a subset or a full copy) it would have to be masked or somehow protected. As long as security is the primary concern in many organizations, this method is definitely faster and the representation of datasets more reliable.

FUTURE PREDICTIONS

Database virtualization enables the test data to be separately prepared and maintained ready for test execution. In the future, we may expect this ratio of 17% to increase solidly, since it is more effective and efficient. If software testers acquire more technical skillsets in time, they will be attempting similar test data preparation efforts but with faster and more dominant control over their test data sets.

WHAT PERCENTAGE OF YOUR TEST EFFORTS ARE **CONSUMED** FOR TEST DATA MANAGEMENT?

*0-5%
of efforts*

*6-10%
of efforts*

*11-25%
of efforts*

*26-50%
of efforts*

*>50%
of efforts*

11% **19**% **32**% **26**% **12**%

ANALYSIS OF THE CURRENT SITUATION

Quality of any software product heavily depends on the quality of software testing. Testing on the other hand, depends on the test data that is being used. Several researches reveal that lots of testing projects are cancelled/failed due to poor test data quality and these projects are more costly than successful projects of the same size and type.

According to the results, more than half of the participants are spending around 30% of their testing efforts for test data management activities. Besides, around 10% of the respondents are even dedicating more than half of their testing efforts to test data management. Ultimately, we may easily conclude that test data management are amongst the most important testing activities and test teams seem to be unaware of the fact that they are spending considerable amounts of time on these tasks.
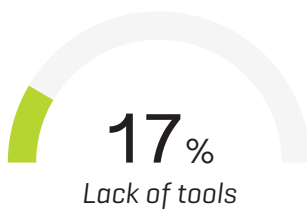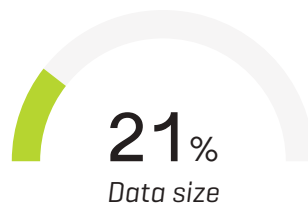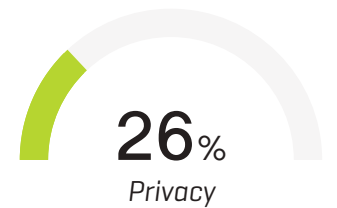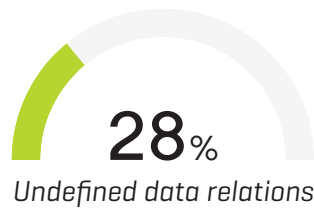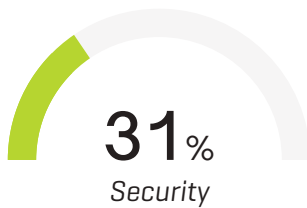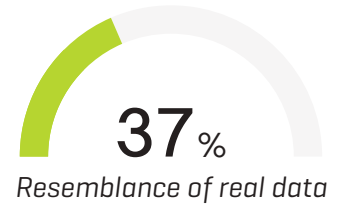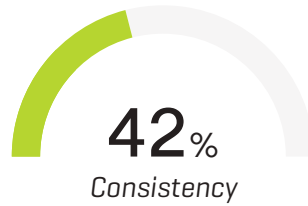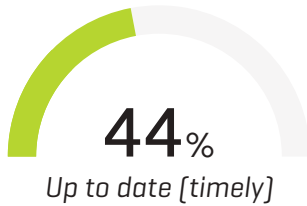
FUTURE PREDICTIONS

Awareness on using high quality test data may shift test efforts to more value added and automated test data management activities rather than unreliable and expensive manual solutions. Organizations, that has more time-to-market focus and that are operating in regulatory environments will be more willing to invest in intelligent test data management solutions.

WHAT ARE YOUR TOP CHALLENGES WHILE **CREATING** TEST DATA?

* multiple selection was allowed

44%
Up to date (timely)

42%
Consistency

37%
Resemblance of real data

31%
Security

28%
Undefined data relations

26%
Privacy

22%
Skilled staff

21%
Data size

18%
Conformance to regulatory compliance

17%
Lack of tools

15%
Bureaucratic permission processes
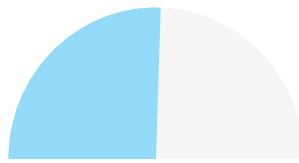
ANALYSIS OF THE CURRENT SITUATION

Lack of production-like data, in several circumstances, may prevent test teams to achieve reliable test results. Unsurprisingly, the most critical issue is still not having the most up-to-date production-like data for testing purposes and sustaining the consistency within those data. Almost half of the participants agree that using up-to-date data is the top challenge they are currently facing with. Other challenges can be listed as the timeliness and security of the test data and the capability of the test data to reflect live data characterictics.

FUTURE PREDICTIONS

Lack of production-like data will probably continue to be the most important issue for upcoming years. By investing on powerful test data management tools and establishing solid processes, test teams may lower these proportions. In addition, software testers working closely with DBAs (Database Administrators), may lower the effects of inconsistent and undefined data relations but will not eliminate the challenges 100%.

5 HOW DO YOU PROTECT SENSITIVE DATA IN TEST ENVIRONMENTS?

* multiple selection was allowed

51%
Masking data

27%
Generating synthetic data

22%
Shuffling data

19%
Relying on personal integrity

18%
We don't protect sensitive test data

ANALYSIS OF THE CURRENT SITUATION

Survey results indicate that, majority of the organizations are trying to secure their test data systematically. Particularly in finance, banking and insurance sectors, legal and regulatory obligations force enterprises to protect their sensitive data.

In order to achieve the best protection, most of the organizations are masking their test data (51%), they are generating synthetic test data (27%) and they shuffle sensitive data (22%) to make anonymous test data sets. On the other hand, around 20% of the organizations still rely on personal integrity on protecting their sensitive data, and around 18% of respondents indicate that there exist no defined procedures for protecting their sensitive data. This means more risks are carried for no reward.
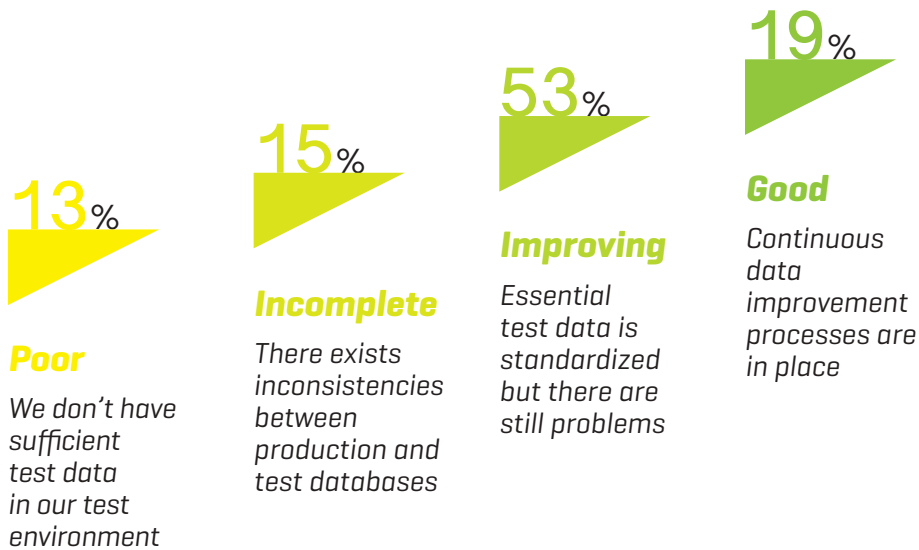
FUTURE PREDICTIONS

Trend is promising. Day by day, test data management is seen as an independent activity. With the newest advancements in Big Data, organizations will focus more on protecting their test data. Power of analytics will make organizations more able to foresee deterministic behaviors of their users, and predict non-deterministic outcomes.

6 WHICH ONE **BEST DESCRIBES** YOUR TEST DATA QUALITY?

13%
Poor
We don't have sufficient test data in our test environment

15%
Incomplete
There exists inconsistencies between production and test databases

53%
Improving
Essential test data is standardized but there are still problems

19%
Good
Continuous data improvement processes are in place

ANALYSIS OF THE CURRENT SITUATION

Test Data Management is the new hot topic of software testing industry without any doubt. Many organizations have started to invest in this area and even some have observed satisfactory results. More than two thirds of the participants indicate that, they have standardized their test data processes with few problems. On the contrary, some organizations still suffer from data inconsistencies and shortages in test environments.

Although there might be a plenty of different causes, problems seem rooted in architectural deficiencies. These problems are distinctly rare among well-defined engineering processes.

FUTURE PREDICTIONS

Distance covered so far, especially in the lastest few years, points that upcoming improvements in sense of test data management will happen in a short time. With the development of new software engineering patterns, such as micro-services, obstacles impeding can be removed and processes can be defined.

It should be always remembered that better testing increases software development quality, while proper data boosts test quality.

WHAT PERCENTAGE OF YOUR TEST DATA MANAGEMENT EFFORTS ARE **AUTOMATED**?

0-20%
of efforts

21-50%
of efforts

51-75%
of efforts
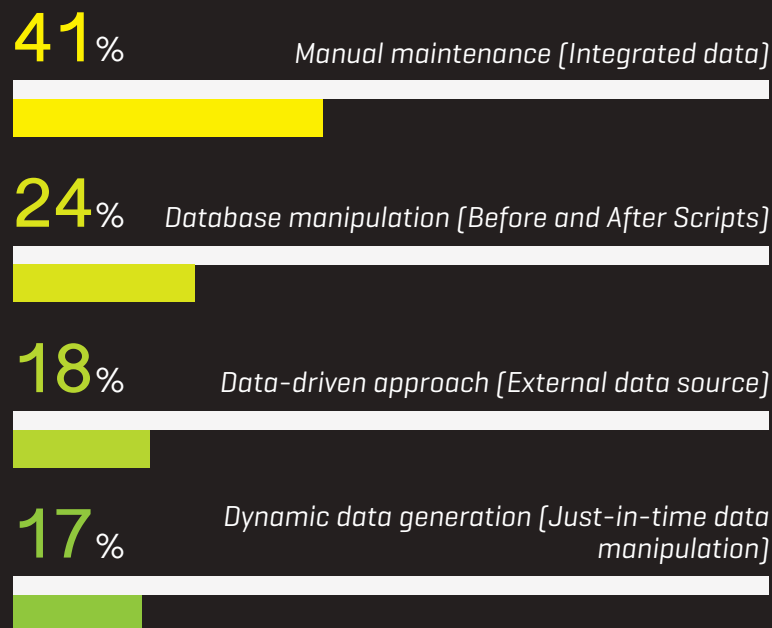
76-100%
of efforts

59% **20**% **17**% **4**%
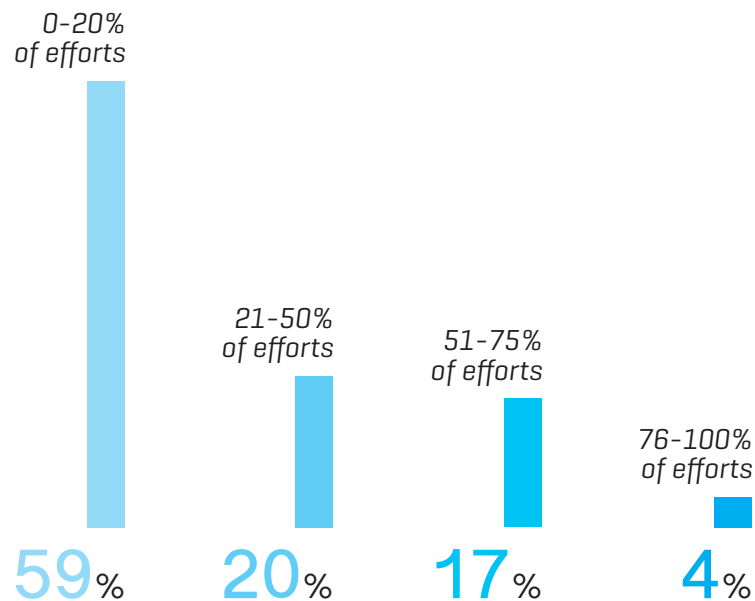
ANALYSIS OF THE CURRENT SITUATION

Survey shows that many organizations do not have any automated processes for test data management. Instead, they still have manual processes on-board to handle test data management needs. As the survey indicates, only 4% of the organizations have fully automated their test data management processes which are supported by a commercial tool or at least the data scripts are in place. When there's lack of automation or tooling in test data management processes, testers focus more on preparing test data instead of executing tests. Besides, testers working with insufficient test data, focus more on "happy-path" scenarios where the benefits of negative testing is not realized.

FUTURE PREDICTIONS

In today's world, the number of IT projects grow very rapidly and the need of testing is higher than it was before. In order to conduct high quality testing, test data quality seems to be an important driver. To be able to testing the applications rigorously, organizations should provision high quality 'fit for purpose' test data, regardless of which methodologies or tools they use. In practice, there are constraints and challenges to manage test data manually. For those reasons, it is expected that number of organizations investing more on test data management activities will increase drastically.

HOW DO YOU MANAGE DATA USED BY **AUTOMATED FUNCTIONAL** TESTS?

41% — *Manual maintenance (Integrated data)*

24% — *Database manipulation (Before and After Scripts)*

18% — *Data-driven approach (External data source)*

17% — *Dynamic data generation (Just-in-time data manipulation)*

ANALYSIS OF THE CURRENT SITUATION

Survey indicates that 41% of the organizations which are automating their test cases, have manually inserted test data into their test scripts. 24% of the organizations seem to refresh their test database before and after the automated test execution. Database refresh is realized by the help of DB scripts to create a clean data set and to rollback the 'dirty/used' data after the test runs. 18% of the organizations are using external data sources like csv's or properties files and the rest of the participants are creating dynamic data sets for their automation scripts to have brand new data set for their test runs.
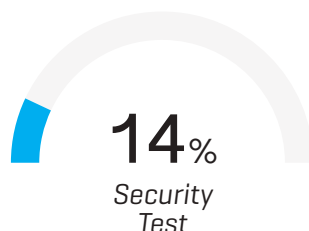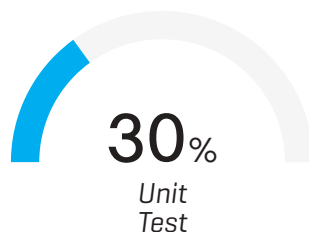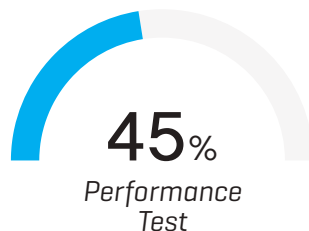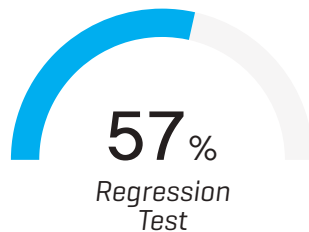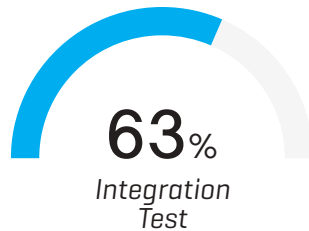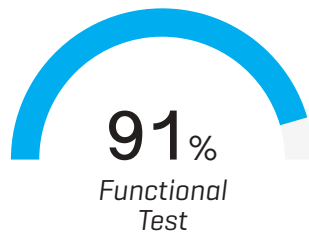
FUTURE PREDICTIONS

In today's environment of shorter SDLCs, test automation becomes an increasingly critical and strategical necessity. A good test automation framework can be achieved by the separation of test data and test scripts. As some organizations see the potential benefits of test data abstraction, they will try to adopt Data-Driven Test Automation framework into their test automation logic. As a consequence, they will achieve more reusable test scripts. Besides, test maintenance costs will be less and test coverage will be better. With the rise of the automation, it is also expected to see decreases on database manipulation that demands qualified workforce.

WHAT ARE YOUR
TOP THREE TEST LEVELS/TYPES
THAT YOU NEED
TO GENERATE YOUR TEST DATA?

multiple selection was allowed

91% Functional Test

63% Integration Test

57% Regression Test

45% Performance Test

30% Unit Test

14% Security Test
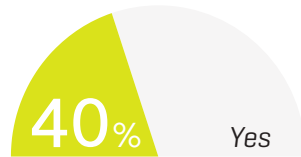
ANALYSIS OF THE CURRENT SITUATION

As it can be clearly observed from the results, almost all the critical testing activities require test data. Functional testing has the most hits, since it is one of the most commonly used test types among different industries and domains. Furthermore, integration testing, security testing, performance testing and unit testing seem to be other high priority technical testing activities in which reliable and consistent test data is essentially required.

FUTURE PREDICTIONS

With the increasing importance and necessity of test data in any software testing activity; test experts will focus more on identifying test data requirements and test data management solutions will provide more flexible solutions for test management tool integrations. By the help of these potential evolvements, technical test activities and especially regression testing will catch up with functional testing and top three test types/levels will be head-to-head.

10 DO YOU USE **UNMASKED LIVE DATA** IN YOUR TESTS?

40% Yes

60% No

ANALYSIS OF THE CURRENT SITUATION

One of the most important questions that the respondent are concerned with is the use of live data during software testing activities. Referring to the answers given on this issue, we observe that nearly half of the institutions are testing with live data. Even during these tests, another important indicator that needs to be considered is that the live data is used without masking/de-sensitizing of sensitive data.

FUTURE PREDICTIONS

Although the Banking, Finance, Insurance, Telecommunications and Defense Industry sectors are regulated by supervisory institutions for the use of sensitive data, though, still data to be used in many sectors in the testing process is transported from the live environments with unplanned and unstructured processes. Ultimately, in the near future it is anticipated that; laws and regulations affecting the IT organizations will be more widespread among different industries. Moreover, these regulations will gain more maturity and they will grow in a way that affect all organizations that handle sensitive data.

CONTRIBUTORS

Ahmet İzgi
Logo

Barış Sarıalioğlu
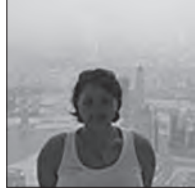*Keytorc*

Berk Dülger
Keytorc

Besim Kosova
Keytorc

Canberk Akduygu
Keytorc

Duygu Onaral

Emine Çelik
Turkcell Global Bilgi

Emrah Yayıcı
BA-Works

Fatma Molu
Kuveyt Türk

Halit Özdemir
OdeaBank

Kadir Herkiloğlu
Havelsan

Koray Yitmen
TTB (Chair)

Merve İçöz
TTB

Okan Coşkun
Scala Mice

Pınar Cinali
BA-Works

Sera Seren
TTB

Vilma Orancı
TEB

Tolga Mırmırık
WeBeYou

ABOUT

Turkish Testing Board (TTB) is the regional body representing and supporting software testing professionals in Turkey. The TTB was constituted in Istanbul in September 2006 as a non-profit organization and a member of the International Software Testing Qualifications Board (ISTQB).

TTB is responsible for certification of testing professionals to the standards and syllabi laid down by the ISTQB. TTB also acts to generate public awareness of the economic and risk mitigation benefits that professional software testing practice offers.

www.**turkishtestingboard**.org

TestIstanbul is the largest conference in South East Europe and Middle East on software testing. TestIstanbul introduces the region not only to the advancements in software testing but also to the advancements in other streams of SDLC like business analysis, design, development and usability. With its almost 400 participants from all over the world every year, TestIstanbul creates a healthy discussion and networking platform for IT professionals and organizations.

www.**testistanbul**.org

ISTQB is a global, non-profit organization responsible for enabling test professionals, through globally accepted software testing certification standards to support their career development. As of June 2015, ISTQB has issued over 410.000 certifications in more than 100 countries.

www.**istqb**.org

TURKEY SOFTWARE QUALITY REPORT

Test Data Management
2016-17

TURKEY
SOFTWARE QUALITY
REPORT

Test Data Management 2016-17

ISTQB™
**International Software
Testing Qualifications Board**

www.**turkishtestingboard**.org | info**@**turkishtestingboard.org
Phone: +90 212 290 76 62 | **Fax:** +90 212 290 76 63

TTB
Turkish Testing Board
www.turkishtestingboard.org