

# Successfully control data quality

**New in RapidRep 5.0:  
Data comparison wizard**

**The RapidRep Test Suite is an innovative software for automated back-end testing.**

RapidRep can access data sets of different systems and connect them as required. Among the data sources supported are all common databases (Oracle, DB2, Microsoft Access, Sybase, SQL Server and many more) as well as structured text files.

RapidRep's flexibility can be seen in its numerous possible applications. For example, RapidRep can:

- verify the results of data processing.
- demonstrate the correct implementation of a migration.
- measure and improve the quality of data (raw data and results).
- repeat tests automatically and transfer their results to a test management system.

RapidRep is developed, distributed and continuously improved by FINARIS.

Based on your purchasing strategy, you can either buy or rent the RapidRep Test Suite.

You can find our current training program on our website. On demand, we offer workshops specially tailored to your needs at your or our offices.

## Test management

Companies use test management systems (TMS) to plan, manage and track test cases. Only in a very limited way is testing the back-end supported by these systems. RapidRep on the other hand evaluates test cases by machine and documents the result in one of the systems supported like HP Quality Center, IBM Rational Quality Manager or Microsoft Team Foundation Server. RapidRep creates a detailed Excel workbook for each executed test case. In the TMS, these workbooks provide the revision-safe test proof.

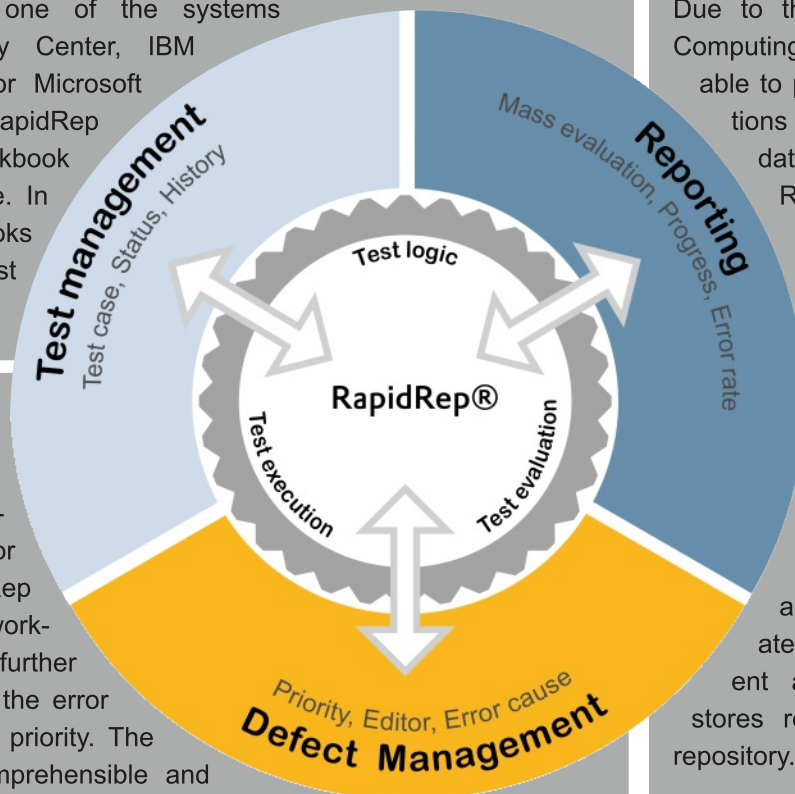
## Defect management

If RapidRep finds at least one deviation from the target result during the test evaluation, it creates a defect for the failed test case. RapidRep attaches the created Excel workbook and if required, it adds further information like for example the error cause, responsible editor or priority. The detailed workbooks are comprehensible and facilitate troubleshooting. RapidRep supports numerous defect management systems.

## Reporting

RapidRep can store results and interim results of the test execution in a database and then evaluate them. Due to the optional use of Cloud Computing technologies, RapidRep is able to perform very many calculations in parallel on different data sets.

RapidRep presents the evaluation results in user-defined Excel workbooks, providing you with an overview of large amounts of data. Examples for such mass evaluations are reports on the status of your data migration or data quality evaluations, for example. The workbooks created are absolutely transparent and replicable. RapidRep stores reports revision-proof to a repository.



Corporate processes require business data of high quality. They shall be up to date, coherent and consistent. In reality, however, it is often quite different. As a consequence, successful companies are more and more viewing data quality as a process and strategically essential task, which it is worth investing in.

With RapidRep, companies of any area of business can improve their data quality in a cost-effective and sustainable manner.

Data have to meet technical and contentual requirements. Therefore it is an advantage when technical and subject-specified employees work closely together in improving the data quality.

An effective process for data quality improvement should thus be able to include employees from the business and IT departments. Such a process basically consists of four phases (see illustration below).



### Definition of the Data requirements

Members of the data quality team determine which data sources shall be reviewed and which properties the data must have. It can also be helpful to describe exactly those constellations which characterize flawed or incomplete data.

Excel presents itself as a commonly accepted and suitable tool for the determination of rules and the communication between employees of the IT and business departments.

In Excel, they define...

...which data (tables, queries or files) shall be tested.

...which tests shall be applied to the data and what the corresponding error notification messages shall be if data records do not pass the tests.

... threshold values for the number of defective data records, whose exceedance shall generate a notification.

Rule #	Test	Precondition	Additional condition	Conclusion
1815	PLZ1	COUNTRY = 'DE'	LENGTH(PLZ) <> 5	FAILED
1816	PLZ1	COUNTRY = 'DE'	PLZ IN (#DOMÄNE('DE-PLZ'))	PASSED
1817	PLZ1	COUNTRY = 'DE'	-	FAILED
1818	PLZ1	Remainder	PLZ IS NULL	FAILED
1911	T1_ISIN	LENGTH(ISIN) <> 12	-	FAILED
1912	T1_ISIN	Remainder	#ISINChecksum(ISIN) <> OK	FAILED

... rules for the individual tests which RapidRep uses to clearly identify erroneous data.

Using the example of a zip code

(German: PLZ) test, the procedure shall be explained by means of a small set of rules:

1. RapidRep determines all rules that exist for the test, in this case 'PLZ1'. In our example they are the rules 1815 to 1818. RapidRep always evaluates rules in ascending order.

2. For each rule, RapidRep checks, whether the data record in question meets the conditions in the column "Precondition" as well as in the column "Additional condition".

If so, the evaluation for this record is complete at this point. The result is derived from the column "Conclusion" (PASSED or FAILED).

The distinction between "Precondition" and "Additional condition" increases readability. By this way, case differentiations and remaining amounts can be easily identified.

3. If neither precondition nor additional condition are met, RapidRep evaluates the next rule analogous to point 2. If no applicable rule

exists, the data record is considered valid.

The table below presents the zip code (PLZ) test at the example of five data constellations. RapidRep determines a rule violation for three data records.

Country	PLZ (zip)	Rule	Result
'DE'	'99999'	1817	FAILED
'DE'	'60594'	1816	PASSED
'DE'	'1129'	1815	FAILED
'CH'	empty	1818	FAILED
'GB'	'SE1 7TJ'	-	PASSED

The method presented is characterized by its clarity and transparency. The rules are clear, unambiguous and flexibly extensible. When building rules macros can be used, as in this example #DOMAIN('DE-PLZ') for querying valid values for German zip codes.

Using a similar approach, statements can also be made about amounts of data records. Such rules can for example determine, whether exactly one record exists per year and customer, whether a specific database field contains positive as well as negative values or whether ranges of values in fields vary as expected.

## Perform measurements

The defined rules build the logic for measuring the data quality. Evaluations on the distribution of data characteristics over time can be used for plausibility checks. RapidRep performs measurements by machine and completely automatic. For continuous quality measurement, RapidRep can be executed periodically (e.g. daily), time-controlled or individually parameterized. RapidRep stores all defective data records in a database.

When executing RapidRep, parameters control the extent of the evaluations and mark the results of a measurement, in order to be able to exactly monitor changes from previous evaluations.

## Analyze results

The task then is to derive insights from the many measurement results which can serve as a basis for improvement measures later.

To this end, RapidRep collects and presents the measurement results in a detailed Excel workbook.

The summary sheet lists the number of defective data records per test and contrasts them with the successfully tested data records.

The Delta worksheet displays changes from previous analysis runs. If these exceed an absolute or relative threshold value, they are highlighted in color (see fig. below).

For each data source tested, a separate worksheet exists which lists

Data source	Test	# Errors (current)	# Errors (previous)	Error text
ADRESSES	PLZ1	148	150	PLZ is missing or invalid
CLEARING	T1_ISIN	9	5	ISIN incorrect

the defective data records for this data source.

Another worksheet lists the currently valid data quality rules which RapidRep used for the evaluation. Due to the compact and comprehensible presentation, the error evaluation can also be done by employees without special technical expertise.

In practice, it is often the case that irregular data constellations shall be excluded from the next series of measurements. For this case, there is an exclusion table whose entries prevent for example that RapidRep repeatedly identifies well-known and accepted data inconsistencies.

Correction Rule #	Violation of Rule #	Precondition	Additional condition	Correction
7099	1815	LENGTH(PLZ) = 4	'0' & PLZ IN (#DOMAIN('DE-PLZ'))	UPDATE ADRESSES SET PLZ = '0' & PLZ

## Eliminate defects

To detect defects, only intensive testing helps. By the approach introduced here, the results which a program generates and stores can be made plausible via rules.

In order to improve erroneous data, it is possible to define correction rules. Thereby the fact is used that RapidRep knows the unique rule ID against which a record has failed.

For each rule ID the data quality team can determine which countermeasures can be taken in the form of updates, default values etc.

Example: A data set with the value '1129' (Dresden) in the field PLZ (zip) is delivered via a faulty delivery system without a leading zero (0) and therefore violates rule 1815.

With rule 7099, the correction mechanism checks, whether the zip code has 4 digits and whether '01129' would be a valid zip code. If so, then the record is supplemented by the missing zero.

The solution presented here by a simple example is suitable for data checks as complex as required.



In several projects, RapidRep could significantly decrease the cost for testing the back-end. The increase in quality accompanying these cost savings is a direct result of the following features of RapidRep.

## Automation

RapidRep has an efficient batch interface which allows for a highly automated degree of the evaluation of test cases and the generation of reports. Thereby, RapidRep performs the underlying calculations without any manual intervention. The evaluation of the complete test case portfolio can be repeated as often as required.

## High level of integration

RapidRep communicates with all renowned test and defect management systems. As a consequence, RapidRep can be integrated smoothly and without manual intervention into an overall test process. The RapidRep user does not need any knowledge about the underlying test or defect management system.

## Efficient integration of resources

In RapidRep, Excel workbooks play an important role. They allow to early on include members from the business and IT departments in designing the Excel templates together.

This ensures that the workbooks, which get automatically filled during the test evaluation, contain all details necessary to replicate the executed tests. The main effort lies then only in the error evaluation and not in the otherwise elaborate implementation and documentation of the tests.

## Global Delivery Model

Due to its modular architecture of different specialized components, RapidRep enables the organizational separation of test case design and test execution.

Besides, RapidRep saves all artifact versions to a repository. This allows different persons to apply changes to the test implementation from everywhere.

## Reusability

Many elements of a test implementation can be reused in other projects. Consequently, the supplied function libraries grow steadily to a collection for returning tasks. For the following standard problems, complete solutions are already available:

- Wizard for comparing data from different source systems (Data comparison wizard)
- Wizard for rule-based data quality evaluations (Data quality wizard)

After a one-time configuration, both solutions are ready for immediate use.

## USP

In the market for test automation there is currently no solution comparable to RapidRep. The provable successes in projects show that a highly automated level of back-end testing can be realized with RapidRep. The license costs for purchasing or renting RapidRep amortize within a short period time.

HP Quality Center, IBM Rational Quality Manager, Microsoft Excel, Microsoft Team Foundation Server, Oracle, DB2, Sybase, SQL Server and RapidRep are (registered) trademarks of their respective owners.

## System requirements



- 1 GB RAM, 400 MB hard drive space
- Windows XP SP3, Vista, 7 and 8
- RedHat- and SUSE-Linux GTK > 2.2.1
- IBM AIX
- Solaris / Open Solaris

## Evaluation / Proof of Concept



Would you like to evaluate RapidRep® at a specific application case in your company?

Please contact us.

We gladly implement a RapidRep® test installation and convince you personally of our software's benefits.