

# Test Data Management

**Representative data, compliant for test environments**



```
01011011010011
01001011010010
01011011010011
  0010
    11
      010
        01011011010011
          01001011010010
            01011011010011
```

## Test Data Management

To maintain a competitive edge in today's information economy, companies and other organizations need to gather vast swathes of customer and corporate data from their existing databases and new social media, public and cloud-based data sources; rapidly analyze that data; and use that knowledge to improve their products and services.

To maximize their use of information, companies also need to integrate their multiple data sources together, across a range of interconnected systems, both inside and outside the organization.

That means constantly updating their core business applications, which in turn puts much greater emphasis – and pressure – on their software development and testing.

To successfully test, you must have the right data. That is the key challenge addressed by Capgemini Group's Test Data Management (TDM) service.

### Business benefits

The Test Data Management service enables organizations to increase their testing speed by up to 25%, cut costs by 5-10% and maintain data security. It achieves this by solving two key challenges:



### Challenge Number 1: How to improve software testing speed and quality while minimizing cost.

More sophisticated business intelligence (BI) systems, combined with lower storage costs, have led to businesses collecting ever-increasing volumes of data relating to their day-to-day operations and customer preferences. Managers more than ever are under pressure to analyze this data, interpret how it can be used, and share it with different people and groups.

Take financial instruments, for example. Most are now traded/priced in real-time (high-frequency trading) rather than as previously at pre-defined time periods (minute-by-minute) – significantly adding to the amount of historic pricing data.

The result is that, unlike the good old days when companies had one mainframe with a single core application holding a few megabytes of data that was updated perhaps annually – today's businesses have terabytes of data spread across 20, 30 or even 40 or more applications, with increasingly 'agile' development resulting in frequent new software releases, sometimes weekly.

All this makes it difficult to test live data without the danger of disrupting production systems. Using 'live' production data for testing also requires terabytes of costly storage, with organizations typically needing a number of test environments for every production environment.

The alternative is to use synthetic data – data constructed for testing – but that brings its own challenges. How do you ensure the test data properly replicates the live production systems so the test is valid? How do you ensure the test data is broad and representative enough to test all possible scenarios and outcomes? How do you ensure you can re-use the same test data time after time, while keeping it consistent across all those 20, 30 or 40 operational systems it is representing?

Simply put, companies can't create test data fast enough for this ever-growing development schedule. Many clients have summed up the challenge: *"We know about our core systems. We know about the data in our systems. But we don't know how to test these systems."*

## Challenge Number 2: How to keep test data secure.

The threat of data breaches is real and growing. Latest UK government figures estimate that 93% of large organizations and 87% of small business suffered a data breach in the year to April 2013. Global figures from the Ponemon Institute show that the average non-malicious data breach costs \$470,000 – mainly from lost reputation, brand value and image – and the average malicious data breach costs \$840,000.

Software testing is a data security Achilles heel. Organizations frequently use production data in larger application test environments because of the difficulty of creating complex data sets for testing. This data represents real information about customers and the business, and may be restricted for confidentiality or security reasons. But while many companies have extensive procedures to protect their production data, once it becomes test data, no-one is really concerned with what happens to it.

Often, test data ends up on a USB stick which may be transported to external outsourcing partners and countries without people knowing about it, or accessed by employees who do not have clearance to look at production data. At the same time, organizations can be targeted by external parties seeking to gain unlawful access to confidential and personal data for fraudulent criminal purposes.

Companies are also faced with more and more laws and compliancy regulations regarding the use of data; and they are being inspected more frequently to see if the rules are being obeyed.

Organizations therefore run a real reputational and legal risk if they suffer a test data breach or get their privacy controls wrong.

## The solution

The Capgemini and Sogeti Test Data Management service tackles these key testing challenges:

- It minimizes the cost of testing – by utilizing smaller subsets of production data that can be re-used test after test.
- It ensures the quality of testing – the subsets of data are still representative of the 'live' production data, making the test fully valid.
- It improves testing coverage – test data is assembled that incorporates all relevant risks and test cases.
- It complies with privacy laws and reduces the risk of data breaches – by 'masking' and anonymizing the test data being used, and sub-setting (not using complete production data).

## How does Test Data Management deliver?

Test Data Management is a professional services-led approach that:

- Analyzes organizations' current software testing and test data management.
- Proposes what actions and toolsets are needed to improve testing.
- Helps customers choose the right testing tools.
- Offers a pilot or proof of concept to show that the selected tool(s) can deliver the test data required within the client organization, and the proposed process can deliver the expected benefits.
- Provides a full TDM rollout.
- Supports and trains customers all the way through the process and even after the rollout.

The TDM approach means clients can test better, faster and more securely. Crucially, it helps 'right-size' the number and size of test environments needed, by introducing a smart solution for getting the right data available. This saves a lot of money.

The more applications organizations have, the more difficult it is to get good data for testing end-to-end environments. The same customer data sets have to appear in every test in an end-to-end scenario, otherwise testing suffers because you can only test parts of your systems.

Using TDM, Capgemini and Sogeti can help customers improve their testing by lining up the data across all their systems. This means you can test faster and identify bugs or risks quicker, and correct them before going live.

Capgemini and Sogeti will identify the complete data landscape across all the client's applications and by doing so, are able to introduce a mask across all systems. So Customer 1 in the masking of application 1 is the same Customer number 1 in the testing of application 2 – and so the test data is aligned across all systems, if that is needed to achieve the goals of the test strategy.

TDM works by taking 'real' data from the production environment, applying field filters and masking it (where appropriate), so it can be used in the test environment.

It also utilizes techniques to create accurate test data where production data is insufficient or unavailable. Test data is assembled by:

- Extracting – TDM takes a sub-set of data from live production data (e.g. all customers with surnames starting with the letter A).

- Masking – it removes data contained in certain records or data fields (e.g. telephone number).
- Enhancing – it generates ‘substitute data’ to re-populate certain cells or adds new fields to create test data sets.

TDM not only works with a copy of production data, it can even help to create accurate test data if the production data is insufficient or unavailable. In that case, test data is also provided by:

- Cloning – TDM copies data sets to create similar or identical test cases to enhance the test coverage.
- Scripting – it uses the actual applications to load data and so ensure data integrity.

TDM is especially also relevant for enterprises that handle confidential personal data such as government, public sector and healthcare organizations – though data security and protecting commercial information is a priority for all organizations.

### Why is TDM from Capgemini and Sogeti different?

**Better service.** TDM not only helps to compile test data, it improves the test process. Capgemini and Sogeti can create the best set-up for the customer; for example, they can execute a scan, select tools and educate and coach employees.

The set-up of TDM is a one-time process, but the service also supports ongoing maintenance. This not only helps the customer create their test process, it also offers the opportunity to improve it. This combination makes the TDM approach unique.

**Full service.** TDM uses the leading TMap® and TPI® test methodologies originally developed by Sogeti to manage the entire test environment – whether that is infrastructure, data or just ensuring that the live production systems are replicated as closely as possible.

Capgemini Group is the only service provider to cover the full testing lifecycle. Other suppliers may offer implementation services for tooling or test data management; and there are tool vendors and specialists in data masking. But no single toolset can handle the whole task, and data masking is only part of the overall solution.

Capgemini and Sogeti take responsibility for the full service delivery, incorporating testing tools, the management of test data and understanding all the issues. They do this by partnering with major tools vendors and helping to integrate the customer’s testing process.

*“We know testing. So we know test data management.”*

### Test Data Management benefits

- Faster test data preparation.
- Quicker data refresh.
- Better test data quality.
- Broader test coverage.
- Flexible test environments.
- Repeatable data quality.
- Lower test data preparation costs.
- Secure and compliant test data.

### Test data challenges

- Speed, agility.
- Using production data in test environments.
- Data in more than one type of database: how to keep it in sync.
- Protecting production and test data.
- Lengthy test data preparation time.
- Size of environments/test sets.
- Complexity of end-to-end testing.
- How to keep end-to-end test data consistent.
- How to provide consistently ‘good’ test data.
- Improving test data coverage.
- Developing the business case.

### Test Data Management’s end-to-end service

Capgemini and Sogeti deliver TDM using a four-stage approach:

1. Initiation.
2. Pilot.
3. Implementation.
4. Operations.

Each stage includes a number of integrated and separately available services:

#### Stage 1: Initiation.

The goal of this phase is to identify the opportunities to improve the customer’s test data creation and maintenance, and speed up the process. Based on this, Capgemini and Sogeti identify quick wins and develop an action plan for the future. They also formulate the conditions for a business case and set up an approach for the tool selection.

One key service in this stage is **Quick Scan**, which reports back to the customer what problems need to be addressed and the steps to take. Quick Scan:

- Analyzes the client's current software testing and test data management situation, including:
  - The existing application/data landscape, environments and infrastructure.
  - Which testing processes and roles are already in place.
  - How often test data is re-used in new applications, releases or maintenance bug fixing.
  - What legal rules and regulations have to be met by the use of test data.
  - The requirements of the test team.
- Proposes what actions are needed to embed TDM into the client's organization and what TDM tools are required to achieve this.
- Identifies the benefits, if any, of introducing the TDM process and using TDM tools – including describing the current problems, proposed improvements and expected results; and describing the expected business benefits, from quick wins up to total implementation of TDM.

Another key service in the Initiation stage is the **Implementation Plan**, where Capgemini and Sogeti draw up a strategy for the TDM implementation or improvement process, including:

- Project set-up.
- Detailed business case for the improvement/introduction of the TDM process and corresponding TDM tooling.
- Tools selection process.
- Proof of concept.
- Pilot.
- Rollout/implementing the TDM process.
- Operational after-care.

### Business case

Presenting a business case for TDM is known to be difficult. The burden to set up test data lies mainly within the testing department, but responsibility for compliance or the cost of environments is often located in different parts of the organization. So the benefits and costs are difficult to combine as the investment and payback are felt on different sides of the company...

Capgemini and Sogeti can help clear up this confusion and present decision makers with a clear business case of the costs and benefits of using TDM in their organization.

### Stage 2: Pilot.

During this phase, Capgemini and Sogeti carry out the tool selection, draw up the financial benefits in a business case, and define the whole implementation process for the customer.

The **Tools Selection** service helps the client choose the right TDM toolset by:

- Determining their goals and tool selection criteria.
- Proposing a shortlist of TDM tools that meet the client's needs.
- Scoring the shortlisted tools against the selection criteria.
- Analyzing the scores and providing selection advice to the client's stakeholders.

Capgemini and Sogeti also offer **Pilot** and **Proof of Concept** services:

#### Proof of Concept

Sometimes clients want proof that, by using specific technology, the pre-selected TDM toolset can deliver the test data they need. Capgemini and Sogeti provide this proof of concept, including:

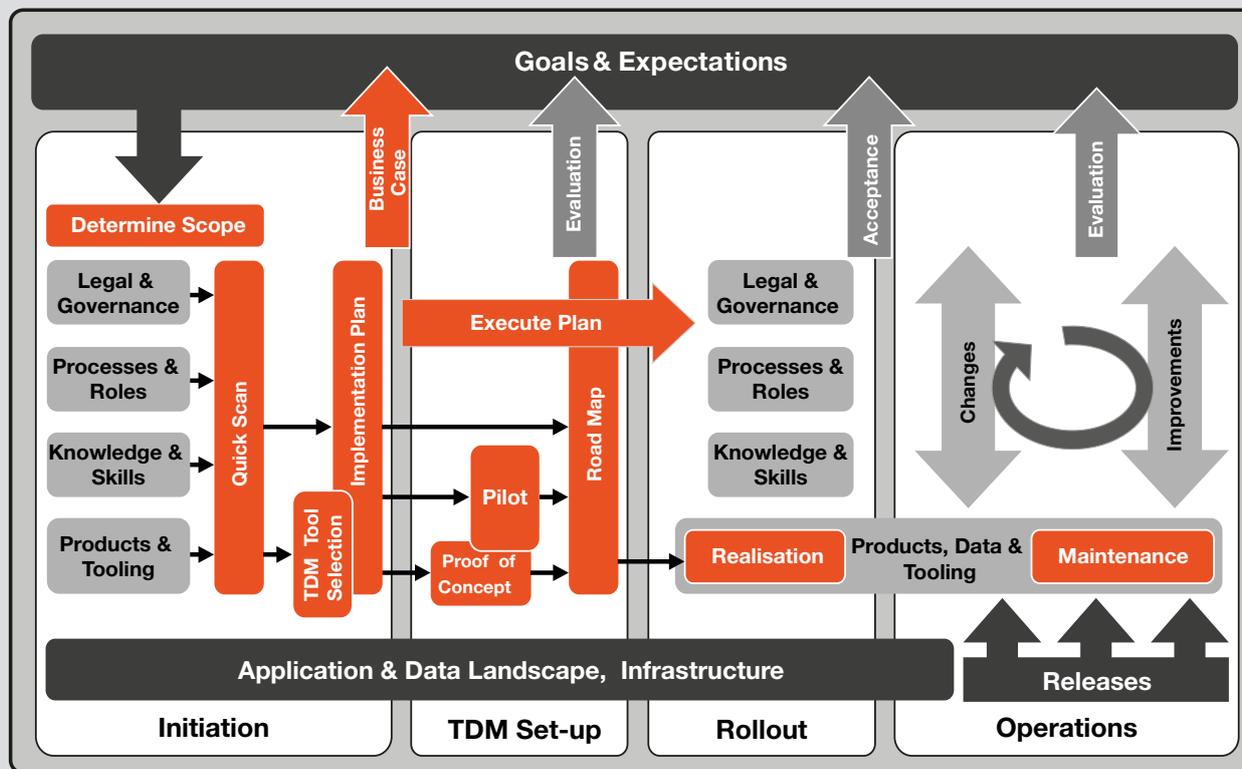
- Determining the solution models for selecting, masking, generating and distributing test data, and drawing up the data design.
- Creating data scenarios for selecting, masking and generating test data, and creating and testing the scripts.
- Preparing the distribution of the test data.
- Executing the prepared scripts and distributing the test data to the designated test environment.
- Presenting the results and evaluating them with the client.

#### Pilot

Sometimes clients want a demonstration that the proposed TDM process is workable and can deliver the expected benefits, Capgemini and Sogeti offer the required TDM Pilot, including:

- Defining the TDM Pilot – determining its goals, tool environment, scope, data coverage and control measures used.
- Preparing the TDM Pilot – installing and configuring the tooling, defining the TDM process outlines for the pilot organization, creating the TDM workspace on the client's infrastructure, and determining and analyzing the general test data requirements.
- Performing the TDM Pilot – executing all stages of the TDM model within a pilot environment, including performing the most risk-full activities in all stages of the model.

Figure 2: Test Data Management Rollout



### Stage 3: Implementation.

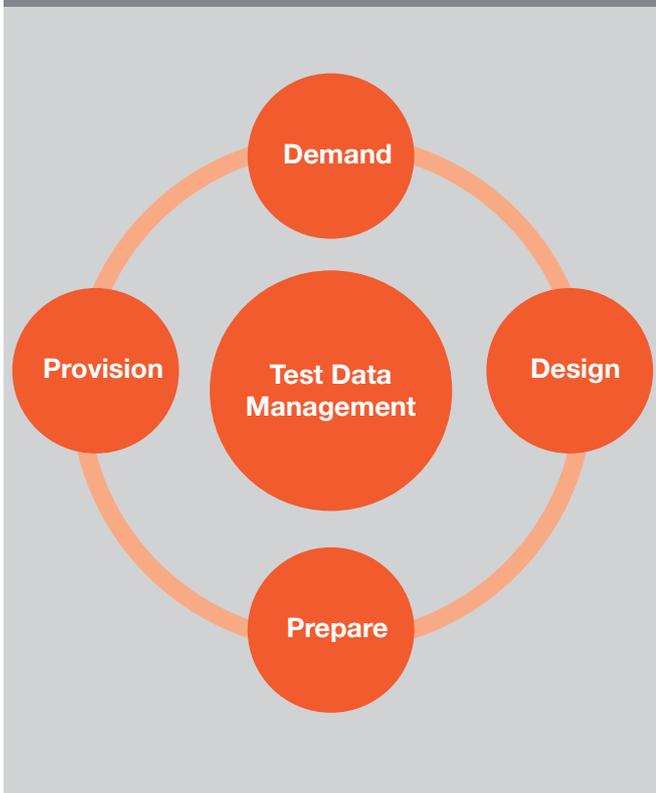
During this phase, Capgemini and Sogeti introduce the selected toolset and the TDM service acts as an instrument for test data. Services provided include **Tools Implementation** and full **TDM Rollout**.

#### Test Data Management Rollout

Capgemini Group introduces the full TDM cycle into the customer organization. Once the TDM process has been set up, it can be applied to new software development or maintaining existing applications. Capgemini sets up the data for an application/new release via a four-step process:

1. **Demand.** Capgemini or Sogeti gains an understanding of the demand for test data in the particular client situation, then determines and details the test-specific requirements – the number of test cases, scope of the data set, etc.
2. **Design.** To turn demand into a design of the test data, Capgemini or Sogeti determines the solutions for selecting, masking, generating and distributing the test data based on what is in scope.
3. **Prepare.** This phase consists of creating the data scenarios for selecting, masking and generating the test data in scope, and creating and testing the scripts. Tools are set up to generate the exact dataset required – either by selecting the right data or creating the data, or a combination of both. The distribution of the test data is also prepared.
4. **Provision.** After the TDM tooling has been set up and tested, data is generated and provided into the right environment at the right time. Other available services during this stage include:
  - Managing the provision of test environments with test data.
  - Keeping test data sets up-to-date.
  - Reseeding test environments with data from the central test data repository.
  - Rolling back the environment to the virgin data state.

Figure 3: TDM's four-stage approach



**Stage 4: Operations.**

Cappgemini or Sogeti defines the required changes to keep compliance with data privacy laws and regulations and continues to improve the test process. Services offered include **Maintenance of Set-up** and **Training & Support**.

**Benefits of Test Data Management**

Two business case studies prove the benefits of TDM:

**Client 1: Large international bank with 3,000-plus applications.**

**Challenges**

- Create representative test data sets for end-to-end testing. Test data creation takes weeks to months.
- Privacy regulations require production data to be masked before use.
- Test data preparation reduced by 50%.
- Test efficiency improved by 25%.

**Project characteristics**

3,000-plus applications, with migration to new platforms continuing every day. Platforms range from mainframe to web applications, cloud solutions and emerging technologies.

**TDM approach**

Identified business risks:

- Production data in test environments.
- Effective end-to-end testing impossible due to complexity of test data.

**Goals**

- Set up a process and corresponding TDM service where test teams can acquire masking expertise.
- Set up a central masking process and tools and services for masking and sub-setting data.
- Reduce the need for test environments by sub-setting production data to functionally correct test data sets.

**Results and benefits**

- Compliance with EU and country-specific privacy regulations.
- Single approach for masking/sub-setting across all company platforms.
- End-to-end test data quality improved.
- Masking across domains reduces test data defects.
- Smaller test environments due to sub-setting.
- Reduced set-up time by using specialized masking teams.
- Cost reduction due to less need for storage in testing.
- Test data preparation reduced by 50%.
- Test efficiency improved by 25%.

**Client 2: Semi-government institution**

**Challenge**

- Comply with privacy regulations banning the use of production data in test environments.

**Project characteristics**

- 30 applications, knowledge of data structure and business knowledge of applications is limited within the institution.
- Complex end-to-end test data prevents organization from performing real end-to-end test runs.

**TDM approach**

Identified business risks:

- Production data in test environments.
- Effective end-to-end testing impossible due to complexity of test data.

### Goals

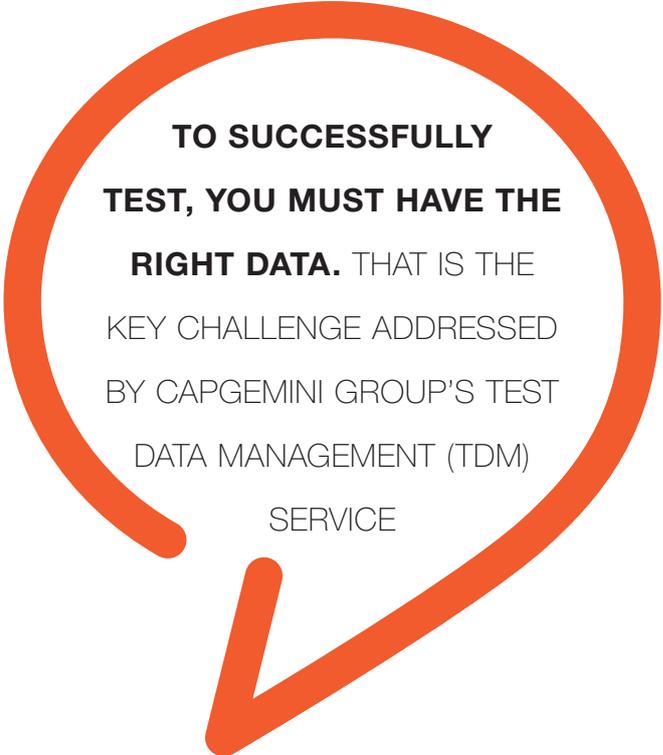
- Set up a process and corresponding TDM service where test teams can acquire masked test data sets for use in testing.
- Due to the complex structure of data, sub-setting from production needs to maintain data integrity to produce functionally correct test data sets.
- Reduce need for test environments by sub-setting production data to functionally correct test data sets.

### Solution

- Sogeti demonstrated in a proof of concept (POC) together with a tool supplier how in a short amount of time masking and sub-setting could produce functionally correct data sets for use in test environments.
- After this POC, the TDM approach is being rolled out to the entire organization including tooling, training and support.

### Results and benefits

- Compliance with EU and country-specific privacy regulations.
- End-to-end test data quality improved.
- 80% smaller test environments needed due to sub-setting.
- 66% reduction in set-up time for preparing test data.



**TO SUCCESSFULLY  
TEST, YOU MUST HAVE THE  
RIGHT DATA.** THAT IS THE  
KEY CHALLENGE ADDRESSED  
BY CAPGEMINI GROUP'S TEST  
DATA MANAGEMENT (TDM)  
SERVICE

## About Capgemini and Sogeti

With more than 125,000 people in 44 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2012 global revenues of EUR 10.3 billion. Together with its clients, Capgemini creates and delivers business and technology solutions that fit needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

Sogeti is a wholly-owned subsidiary of Cap Gemini S.A., providing local professional services, specializing in Application Management, Infrastructure Management and High-Tech Engineering. Sogeti offers cutting-edge solutions around Testing, Business Intelligence, Mobility, Cloud and Security. Sogeti brings together more than 20,000 professionals in 15 countries and is present in over 100 locations.

The Capgemini Group has created one of the largest dedicated testing practices in the world, with over 11,000 test professionals and a further 14,500 application specialists, notably through a common center of excellence with testing specialists developed in India.

Together Capgemini and Sogeti have developed innovative, business-driven quality assurance (QA) and testing services, combining best-in-breed testing methodologies (TMap® and TPI®)

Learn more about us at  
[www.capgemini.com/testing](http://www.capgemini.com/testing) or  
[www.sogeti.com/testing](http://www.sogeti.com/testing)

### Contact

For more information about how Capgemini and Sogeti's Testing Services can help organizations achieve their testing and QA goals, please contact your local Capgemini or Sogeti account manager or our Global Testing Services Team:

**Mark Buenen**  
VP, Business Development  
Testing Global Service Line  
[mark.buenen@sogeti.com](mailto:mark.buenen@sogeti.com)