White Paper

SAP test data: best practices

White Paper by EPI-USE Labs and Sogeti

sogeti Part of Capgemini

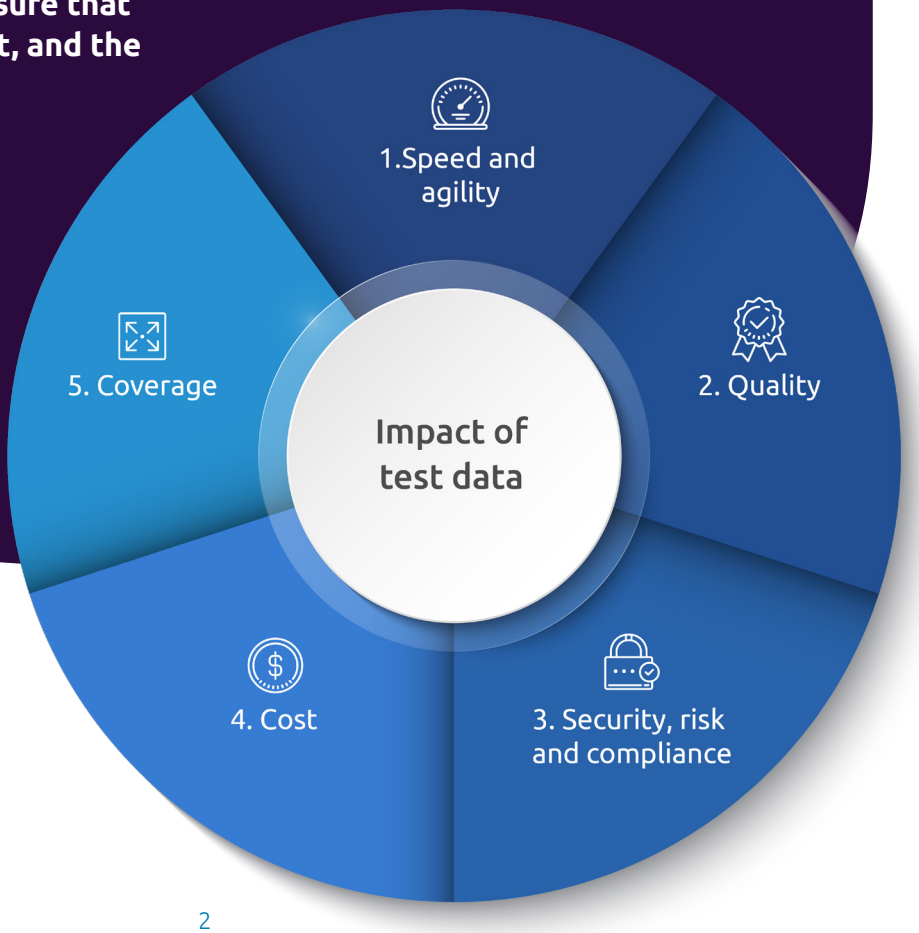
EPI-USE LABS value through innovation

Innovation gives your business a competitive edge. Companies are always looking at ways to minimise the risk and cost associated with innovation, while keeping up with the speed of change. To align IT systems with the needs of the business, initial implementations of SAP® must run smoothly, as well as keeping pace with continual innovation.

A clear test data management strategy allows you to lower the risk of changes to configuration damaging functionality in Production, which can be costly, and increases implementation time. Effective quality assurance and testing strategies make sure that new developments, customisations and optimisations are ready for the Production environment. With today's fast-paced business environments, building confidence with your business users helps to ensure that testing is effective and efficient, and the business can meet its goals.

In this whitepaper, we explore the impact of test data in five key areas:

- Speed and agility
- Quality
- Security, risk and compliance
- Cost
- Coverage.



Contents

Speed and agility 4

Quality 6

Security, risk and compliance 8

Cost 10

Coverage 12

In closing 12

About EPI-USE Labs 13

About Sogeti 13

Speed and agility

The challenge

The speed of business changes is increasing rapidly. This requires that bringing live new features, bug releases and application changes into production should take place faster and faster. These changes should be tested. Modern agile development practices promote and enable rapid changes to cope with this demand.

Whether your business is using the Waterfall, Agile methodology or Continuous Integration/Continuous Development (CI/CD), testing earlier and faster is important:

- You shouldn't leave testing to be performed at the end of the process.
- Testing has to become integrated throughout the process and should be continuous.
- The process is iterative, and the results of new developments should be tested before the majority of investment already has taken place.
- Larger development trajectories are broken down into smaller pieces that can be tested with unit tests as you go.

SAP system copies

Traditionally, SAP system copies are used to copy the whole Production environment to the non-production system. The process is lengthy, resource intensive, and causes system downtime. Once the data has been copied, BDLS and other post-processing activities take a long time, making sure the new system behaves as it should. Many organisations don't have enough time to refresh the data in the non-production environment through a SAP System Copy, or do it yearly at most. The result is that data gets out of date, it slows down the process of testing, and impacts the quality of the test data.

Testing earlier with the shift-left approach

The shift-left approach speeds up the testing process. With the shift-left approach, testing is performed earlier in the development process and systems. The only prerequisite is to have test data available earlier in the development cycle. Faster, earlier testing and testing of a larger scope often means the demands on test data are changing. To have test data available that is appropriate and relevant, both from a quantitative and a qualitative perspective, in every test cycle, means manual preparation belongs to the past. Getting data prepared in time simply becomes unfeasible.

The use of an automated test data solution can speed up provisioning of test data for the quality assurance and testing activities. There are principally two approaches to providing test data for test scripts:

- **Test data from the test automation application**
  The first approach uses the application under test itself to create and/or modify test data for use as appropriate test cases for the tests at hand. Not only does this mean that the application itself needs to be of sufficient quality to be able support calls by test automation scripts to generate the test data; it also means an intermediate step to prepare the test data after the application creation action has been performed. This adds time to the testing process and reduces the savings required to speed up the complete quality cycle.

- **Test data as part of the application rollout**
  A second solution is to deliver test data as part of the application rollout on an environment of one or more SAP applications. As part of the deployment process, fit-for-purpose test data sets are prepared and rolled out to the test environment(s):

  - the right size in terms of test cases;
  - the right complexity to be able to test business processes accurately;
  - the right consistency to identify possible application defects without triggering false positives

  Having the test data deployed as part of the overall deployment process means the data is ready for use and testing can start as soon as the test environment is available. This significantly reduces the period where test environments are unavailable due to test data preparation. It also speeds up testing of SAP applications.

Impact on speed of no test data

The result of not having the appropriate level of test data available can severely alter project timelines and cause frustration in the business. One of the best practice approaches of testing is having test systems available on demand and leveraging the advantages of the cloud.

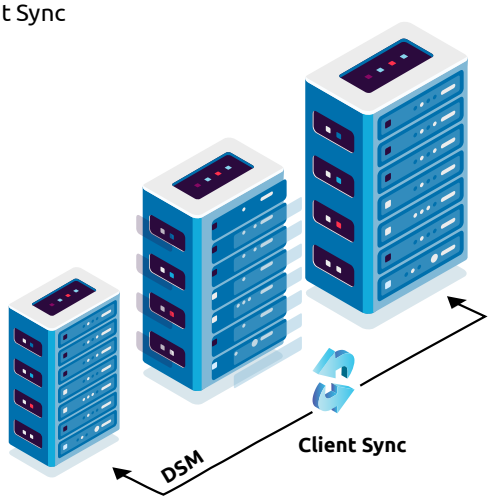
Data refresh with Data Sync Manager: Client Sync

Data Sync Manager™ (DSM) Client Sync™ is an industry-leading solution by EPI-USE Labs that creates a client copy. You can take a subset of Production data and copy it to your non-production data. The process can be much faster than the traditional system copy. It is also less demanding on resources and automates all post-processing activities. For example, Agfa found that a client refresh was 93% faster by eliminating post-processing tasks.

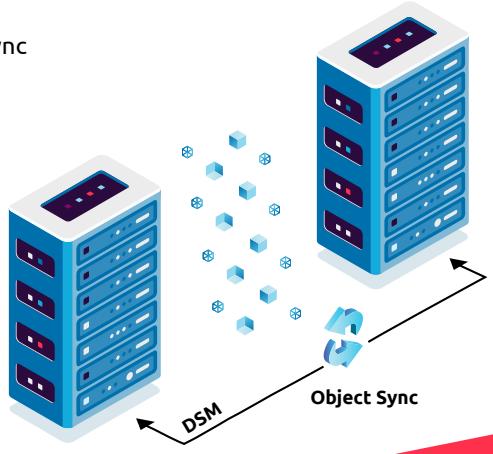
Data on demand with Data Sync Manager: Object Sync

Another alternative is to use DSM™ Object Sync™ to copy only a selection of Data Objects that are required for a test. Business users can perform the copy creation themselves to make sure they copy the data elements that are required. Object Sync copies the data in an intelligent way to connect that data so that the linkages between the data are still in place. It is fast, and doesn't break the data integrity.

Client Sync



Object Sync



Summary

When looking at the speed aspect, the best practices recommended are:

- Make testing part of your process by shifting left.
- Make sure you have test data available when a project starts to avoid delays.
- Copy only a subset of data for smaller agile systems.
- Leverage a data copy solution like DSM Client Sync that speeds up the process of creating test data and avoiding large system copies.
- 'Top up' data with data on demand copied with DSM Object Sync.

Quality

When testing applications, there are two questions to be answered:

- Are the tests I performed correct, and
- is the data I used to perform the tests correct?

Only correct tests and correct data lead to the correct outcome or findings. If the test data used is not of the required quality, unnecessary time will have to be spent on the analysis of the test results. Moreover, this will negatively affect confidence in the tested application.

Having a good test data set ensures that there are significant time savings and efficiency gains in quality assurance and testing. Not having a good data set is one of the main bottlenecks in achieving continuous testing (also within DevOps environments).

Good quality test data ensures confidence in test results

By having correct test data available, it is possible to place more trust in the results of testing. By having correct test data available for each test round, tests are performed correctly each time and the test results and the application can be viewed with more confidence. that data so that the linkage between the data is still in place. It is fast, and doesn't break the data integrity.

Good quality test data can mean many things

Test data needs to be recent data. You want to perform testing for your business today, and this means using recent HR data and actual payroll information, in the case of HCM systems. Or for example with logistics, you want to be testing seasonal products at the correct time of year, not summer articles in winter!

Quality test data also means enough data for testing. In your testing process, you will consume the data during your testing activities. So, the set of data needs to hold enough master data and transactional situations. At the same time, we don't require all data to be copied. A full copy from Production data takes a long time to create, and due to the cost and duration to refresh the test data, and therefore is often avoided, meaning. the data gets out of date/old over time.

The last point we want to emphasise is that the data needs to be consistent, within the system you are testing, and in the process and multiple systems you are testing. And when your sensitive data is scrambled in your test data, this needs to be consistent as well.

How to get good quality test data

Good quality means that your test will give you the same result as when you go live. So having Production-quality data in your system gives you that confidence in the result of the test.

Creating test data manually is risky and time consuming. It's not easy to replicate and won't give you the Production quality data you need. The risk of false-negative and/or false-positive findings will generate unnecessary reworking in testing and in development work. It means development quality will suffer, timelines will be delayed and even issue solving will be difficult, time consuming and frustrating. With the risk, you end up validating results in your Production landscape.

Why is SAP different?

Creating synthetic data is an option some organisations use to 'bypass' the security risk (see below). In the world of SAP, the same information is often stored in different places, and this makes it increasingly difficult to keep the data consistent. If we take the data model of Employee as an example, the first name is stored in over 25 fields alone, not counting concatenated fields, other modules or connected systems like an SAP ERP or SAP® SuccessFactors® system.

Therefore, companies need solutions to get data from Production down to its quality assurance or development systems to get consistent test data of Production quality. But by doing so, data privacy comes into play, which we will cover in the next section in detail. Worth mentioning here is the fact that according to SAP's data processing agreement, you are not allowed to store Production data in its original form in a non-production system. Meaning, you have to scramble your data.

Scrambling data can have an impact on your test scenarios, so it is crucial that the Data Protection Officer gets into discussion with the quality assurance team to find the middle ground. The data is of no use if it is scrambled to a degree that the test scenarios give different results from Production. If we want to scramble the address, we need to first consider if it will impact any calculations on the systems (for example, the delivery cost or the insurance fee) and the same may apply to different kinds of information on your system.

Summary

When looking at the quality aspect, the best practices recommended are:

- Ensure the test data is of Production quality.
- Create systems to bring up-to-date data into your testing environment with a solution such as Data Sync Manager.
- Take the quality into consideration when discussing scrambling data.
- Make sure the linkages in your data stay intact when data is copied so it still behaves in the same way to give you accurate results in your testing process.

Security, risk and compliance

Everyday you can read in newspapers about the increasing frequency and severity of cybercrime. Losing data or access to data is a serious threat to the continuity and success of companies. This makes the priority of cybersecurity increasingly important.

SAP systems are a core component of an organisation's IT infrastructure. Thus, the security of the systems is key. Care and consideration are taken to assure systems are kept secure from the outside and the correct authorisations are in place to limit visibility of sensitive information. The key challenge lies in getting quality test data from Production but not exposing the data to people who should not have access.

System copies refresh the data from Production but do not address the security risk by having Production copies throughout your landscape. An additional layer of anonymisation is required to mask the sensitive data present in these Production copies. When data is anonymised, there can still be a concern that a specific pattern of data can point to the sensitive data in the system.

Synthetic data

An option for securing the data is by creating synthetic data. This process looks at a business process and maps all the paths that are possible during a transaction. Once you have the mapping, you create the data that is required. The concept is that you will have a much fuller set of data, even more than from Production should people only use certain transactions in Production. It solves the problem of security in that you don't have 'real data', but you run the risk of over-complicating the process of creating data and you will create data that won't be used in the Production system, making it superfluous.

Pollute data on purpose

Another option to keep people away from reverse engineering the masking in place is to inject the copied data with made-up data, thus polluting the data set in a way that people won't be able to know the difference between the copied data and the fictitious data. Data pollution is, however, a risk for the application quality as care must be taken to ensure the extra data is structurally and intrinsically correct to avoid the risk of tests being negatively impacted if the data is of insufficient quality.

Data privacy compliance (GDPR)

In 2018, the General Data Protection Regulation (GDPR) came into effect, protecting European citizens' personal data. Similar legislation is increasingly being introduced globally. As part of being compliant with data privacy regulations, you need to identify where you have sensitive data in your non-production systems and who has access to this sensitive data. This can be accomplished by doing a risk assessment to understand your risk.

Key considerations for GDPR:

- Make sure you have valid grounds to store the data, referred to as 'lawful basis'. It is important to communicate from the start what data you will gather and how it will be used. It is important to communicate that the data will be used in a testing environment.
- The data that you need is supposed to be limited to what is adequate, relevant and only what is necessary. This data minimisation enables you to minimise your risk in the testing environment.
- Investigate the retention period you need for data. This could impact testing as you require a certain amount of historical data, based on your business type.
- It is important to have the appropriate roles and user access in the test environments. These roles can degenerate much faster than in Production environments where tighter controls are in place.
- Providing access to testers outside of the EU can be seen as a data transfer and more permissions are required if this is the case.

Authorisations

This leads us to the topic of authorisations and how to manage it in non-production environments. This can be a costly and time-consuming way to keep control over the risk of data leakage. By anonymising the data, you remove the risk, reduce the cost and less focus can be placed on keeping the authorisations strict. SAP states in their own contract that all data should be anonymised:

"Customer shall not grant SAP access to Licensee systems or personal information (of Customer or any third party) unless such access is essential for the performance of SAP Services. Customer shall not store any Personal Data in non-production environments."

Pseudo-anonymised data versus fully-anonymised data

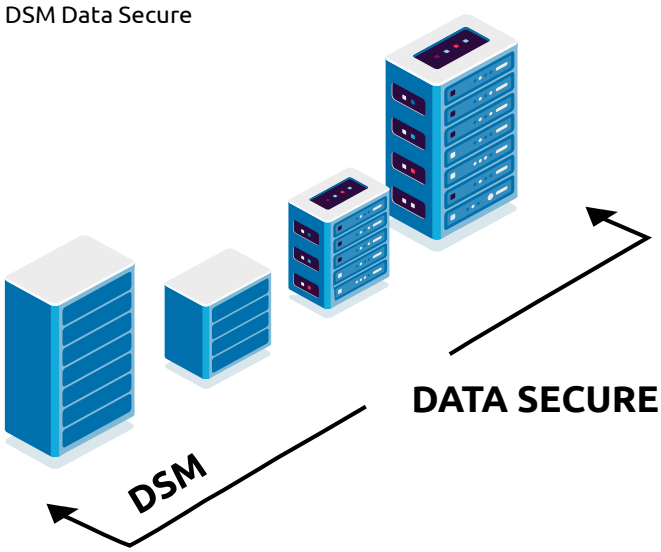
This terminology is also used in GDPR where they differentiate between the two. With pseudo-anonymised data, it means there is still a way to get the real data back when you have the seed table or the key of how the data was masked. For example, if every 'John' in the system becomes 'Peter' and you can work out the key, one can get back to the original data. This is not deemed as secure. The alternative is fully-anonymised data where you can't get back to the original because in that situation pattern-breaking technology is included. For example, the first 'John' becomes 'Peter' and the second one become 'Harry' and so forth. If the data is fully anonymised, it means that the data has been changed to such a degree that it no longer presents a risk to the business.

Secure your non-production environment with DSM Data Secure

DSM Data Secure™ is an industry-leading solution that automates the masking of SAP data, either 'in place' or more importantly 'at the source' so that the data is masked before it ever leaves the source system. The original sensitive data is never duplicated so there is less chance of it falling into the wrong hands. Additionally, most organisations have integrated SAP landscapes where data is distributed across ERP, CRM, SRM and other environments. So these systems should also be included in the anonymisation process.

Data Secure's unique 'pattern-breaking' technology ensures that the same starting value does not result in the same new value unless the records are for the same person, even across systems. This makes reverse engineering of a seed table impossible; for each person a randomised value has been used. The cross-system integration ensures all references remain intact after scrambling.

DSM Data Secure



Summary

When looking at the security aspect, the best practices recommended are:

- Copy data from the Production environment but anonymise data before it leaves the Production system.
- Minimise the data footprint by creating subsets of data that are needed for testing.
- Choose a robust solution like DSM Data Secure that offers an automated method to anonymise the data in your non-productive environment.

Cost

22% of a large enterprises' IT budget will go directly to testing"

SAP is synonymous with the ability to adapt and develop custom processes and applications to fit your business requirements. To adopt new developments fast and with limited risk, testing is crucial. According to the Capgemini Quality Report for 2020-2021, large enterprises allocate 22% of IT budget to QA activities. The cost can be much higher, but as a business, you need to balance cost.

Reducing the footprint of non-production

As mentioned before, for optimum testing, you need accurate and recent test data. To achieve this, full system copies can be made from Production. In terms of cost, these are not optimum. Full copies are just that – a duplication of the whole system, meaning that your non-productive environment is the same size as Production.

With Data Sync Manager Client Sync, you can do a selection based on the last number of months' transactional data and all the master data to create a system with a reduced footprint of your Production environment. You can save up to 80% of storage space when using this method.

Additionally, this lean new system will reduce the cost of hosting and services costs if you are hosting it in the cloud. When it comes to S/4HANA® systems, the cost associated with the hardware is much higher than before, thus the savings that can be realised are more substantial.

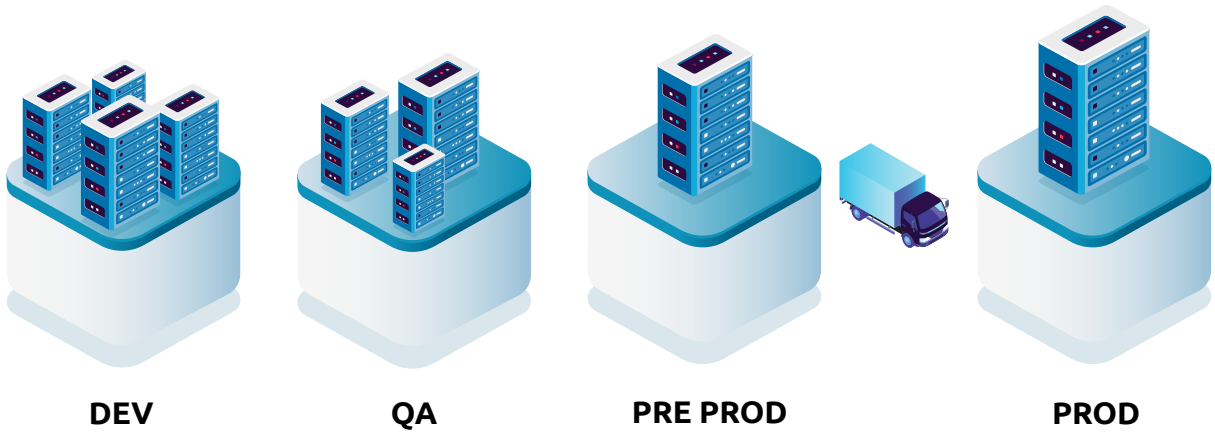
More efficient refreshes lead to cost savings

The process of full system copies is cumbersome and tends to take a long time and consumes a lot of man hours to perform all the post-processing tasks.

A relevant, intelligent and smaller selection of test data will mean faster refreshes. This means less time is used and lowers costs in terms of manpower while refreshing your test data. System copies also mean there is system downtime involved when the data is exported. This is one of the reasons organisations choose not to refresh the data as the business can't afford to have downtime. This downtime can be eliminated by creating a client refresh with built-in automation for the post-processing steps.

When the process is more efficient, it also means that with recent data you can reduce the amount of time needed to resolve issues. If an issue is not resolved in a timely manner, it can potentially lead to business loss. Thus, the sooner an issue can be resolved the better.

Reducing the footprint of non-production systems



Improved coverage

Working in agile teams to deliver new releases more quickly, and faster client refreshes, allows data to be refreshed more often as it is needed. This means every new sprint can start with a fresh set of test data, and test data based on the latest released functionality. That extra object added to a screen generates more and new data in Production, with a recent refresh this new object will have actual test data from Production. Your new sprint is now tested on real data, so on actual reality. This gives confidence to the testing staff and the development team to move releases to Production faster.

When you can move your developments into Production faster, it shows on the bottom line. It enables IT to support their business in a cost-effective manner.

Continuous improvement

The improvement in test coverage means you will have fewer findings and less rework. This is continuous improvement, also your applications management team will be able to test issue solving on more recent test data. They can use test data on demand by copying and cloning data from Production to the test systems and and thus resolving issues is much faster.

Cost of compliance

There are cost savings for testing with scrambled test data. Although this is predominately a security and compliance concern, the potential cost for the company when in breach of GDPR is 4% of annual global turnover or up to €20 million (whichever is higher). By scrambling the data, you reduce your risk of a potentially large fine. Obviously, there is more to data privacy compliance than just anonymising the non-production data. But it is one aspect that is easy to achieve when it comes to testing data generation.

Summary

The best practices when it comes to reducing cost for testing data are:

- Reduce your footprint: less data means less storage means lower costs.
- Improve your processes to include automation and reduce the amount of manual work and therefore reduce the cost.
- Embrace the latest agile methodologies in CI/CD to get new functionality to the market faster.
- The cost of non-compliance is too high so make sure you build your test environment with security and compliance at its core.

Coverage

The dilemma has existed ever since testing was established: how to test efficiently and achieve the highest test coverage possible. Given enough resources, budget and time, testing all aspects of software is possible. Within the current demands on speed, available budget and the strained resources available, however, choices will need to be made.

Test coverage from a business perspective can be defined as the level at which business functionality is covered by test cases. To stay within the constraints of time and budget, tests can be made for the most critical business functionality first, then moving on to less critical processes and on until a sufficient set of test cases covers the business-critical functionality.

If the business-critical functionality is defined, then these test cases can be created. Many a tester has tried to create a sensible set of tests from a business perspective, only to be disappointed by the lack of available test data.

The process of creating and maintaining test data needs to accommodate the ability to select appropriate test data for the test cases that the business wants to perform. There is no need to have gigabytes of test data in your test system which are simply duplicates of one another from a functional point of view. There is more value in supplying the test team with accurate test data for all the test cases that need to be performed.

With that in mind, selecting appropriate test cases from Production sources, moving them to the appropriate test system and masking privacy-sensitive information is the way to create a fit-for-purpose test data set ready to be used for the tests that were designed. By automating those steps in Data Sync Manager, test data can be supplied into any test environment by running pre-configured test data runs. This reduces preparation and execution time and reduces errors by eliminating the amount of manual intervention in these steps.

In closing

With many organisations moving to SAP S/4HANA or SAP SuccessFactors, testing and test data is important to make the project a success. Throughout the SAP lifecycle, having an accurate subset of data for testing provides valuable savings, speeds up the total testing process and reduces the risk of implementing new developments.

Contact EPI-USE Labs and/or Sogeti to learn how you can improve your SAP test data.

Summary

Test data is turned into an enabler instead of an encumbrance by:

- Defining the exact data needed for testing.
- Defining critical business functionality.
- Automating the way in which this data is provided into the test environment before testing.
- Navigating complex datasets and avoiding time-consuming test data finding sessions.

About EPI-USE Labs

At EPI-USE Labs, we help you transform your SAP landscape and solve your business challenges. Our software, value-added solutions and managed services create more powerful and secure SAP and SAP SuccessFactors systems, saving you time and resources. We simplify and speed up data management, giving you the tools you need to navigate small and large challenges successfully. And we've done this for 40 years, all over the world.

EPI-USE Labs is a global company with hubs throughout Europe, the United Kingdom, the Americas, Australia, the Philippines, South Africa, the Middle East and Turkey. With over 1,350 clients across 49 countries, our 97% client renewal rate speaks for itself.

EPI-USE Labs is a member of the groupelephant.com family. Find out how we can help you solve your business challenges: get in touch today.

About Sogeti

Sogeti is a thought leader in quality engineering and testing. Working closely with clients and partners to take full advantage of the opportunities of technology, Sogeti combines agility and speed of implementation to tailor innovative, future-focused solutions in Quality Engineering and Testing, Cloud and Cybersecurity, all fuelled by AI and automation.

With its hands-on 'making technology work' approach and passion for technology, Sogeti helps organisations implement their digital journeys at speed.

sogeti Part of Capgemini | EPI-USE LABS value through innovation