WHITEPAPER

sqs.com

Systematic Test Data Management

Using anonymised data for testing

Author: Sven Euteneuer
(Global Head of Technical Quality)
sven.euteneuer@sqs.com
SQS Software Quality Systems, Germany

Published: November 2016

Management Summary

With a new pan-European data protection legislation looming on the horizon, the topic of test data acquisition has crept up to the top of IT agendas once again. With the new EU regulation coming into force in May 2018, companies have little time to lose in order to ensure compliance.

Many aspects of the upcoming regulation are evolutions of the existing data protection directives. However, a large share of organizations are not yet compliant to either, for instance by reusing unaltered copies of production data. The new regulation will provide for a strengthened catalog of fines and penalties of up to 4 % of worldwide turnover."[1]

Achieving a state of compliant and cost-effective use of test data is possible – it requires implementing a systematic approach to test data management. A wealth of methods and techniques such as synthetization or anonymization of data underpin such an approach and make it sustainable. While synthetic test data has many advantages (e.g. being purely fictional and thus perfectly privacy-compliant, full coverage of test case data requirements), it calls for mature testing processes and precise test data requirements in test specifications. Test Data Anonymisation (TDA) hence is the method of choice to quickly achieve compliance without possibly having to reorganise testing radically.

In this white paper we would like to discuss a concrete case of establishing such a TDA solution at a specific customer. In order to realise data privacy compliance, SQS was contracted to pilot test data anonymisation for a client in the financial industry. During a brief time frame of three months, various technical questions on the test data anonymisation topic were supposed to be answered. Important issues to investigate were the suitability of the client's established platform for test data anonymisation as well as the general applicability of anonymised data for system testing. Furthermore, the test data anonymisation pilot was supposed to already yield a concrete, working anonymisation solution for the selected pilot application. The concepts employed in the pilot, however, were supposed to support scalability for a company-wide roll-out as well.

1) http://mlawgroup.de/news/publications/detail.php?we_objectID=227

The achievements of the pilot are: Two focused 'subject areas' of the selected pilot application were anonymised completely and consistently according to the application owner's test data requirements. Limited to these subject areas, the pilot offers a productively working test data anonymisation solution. In order to achieve full data privacy compliance, however, the solution would need to be extended to cover the remaining subject areas as well. Anonymised data provided by the pilot solution was utilised successfully in existing test cases of the application. No major concerns were raised with respect to a broad usage of anonymised data in system testing. Implementing and testing the anonymisation solution confirmed the tool validation results regarding the suitability of the selected tool. It proved to be reliable, sufficiently powerful with respect to functional and non-functional anonymisation requirements, and efficient to use during development and testing. An anonymisation framework was designed around the selected tool and developed during the project. The framework is well suited to support a company-wide roll-out and efficient central management of the numerous anonymisation solutions required throughout the customer's IT.

Introduction

With new EU regulation coming into force in 2018, organisations across Europe are struggling to ensure that they are able to demonstrate compliance. For the first time, data protection legislation has been enacted in a shape that will result in a homogeneous body of law across all markets without the need or even the possibility of national parliaments introducing variations. This regulation affects all organisations dealing with personal data in the EU, regardless whether they are public or private or whether they are headquartered in the EU or not. At the same time, the new General Data Protection Regulation specifies sanctions on non-compliant organisations ranging from warnings and regular audits to monetary fines ranging up to 4% of the total worldwide annual turnover of the preceding financial year. Of course, fines and criminal proceedings are not the only possible consequences – any publicity of data breaches will negatively affect brand value and business.

The "Regulation (EU) 2016/679", informally known as "General Data Protection Regulation" or GPDR thus repeals the previous "Directive 95/46/EC" and affects how organizations can use data of their users, clients, associates or employees. It states that, i.a.:

Personal data shall be:

- **processed lawfully, fairly and in a transparent manner** in relation to the data subject ('lawfulness, fairness and transparency');
- **collected for specified, explicit and legitimate purposes and not further processed** in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
- **adequate, relevant and limited to what is necessary** in relation to the purposes for which they are processed ('data minimisation');
- **processed in a manner that ensures appropriate security of the personal data,** including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

The controller shall be responsible for, and be able to demonstrate compliance with, paragraph 1 ('accountability').

Thus, one major principle of data protection is the earmarking, which in particular prohibits storing and processing personal data for software development and testing purposes unless explicitly requested from and allowed by the affected persons. The practice of reusing a copy of productive application databases for testing still may be common in many companies, but it clearly violates the letter of the law.

At the same time, IT projects (and sometimes operations request an ever larger amount of data, usually in order to develop, maintain or test IT applications, the end result of which is an increasing demand for "test data". In order to satisfy this demand, a systematic test data management (TDM) approach is required. Such a TDM approach typically has to cover a number of areas, the most common ones are shown in Figure 1.

As a consequence, a test data acquisition mechanism is desperately needed. At the same time and to achieve compliance with data privacy legislation, test data acquisition has to be organised using one of the following methods:

- Synthesis: Test data is being generated avoiding references to actual persons from the start.
- Anonymisation/Obfuscation: Productive, individual-related data is transformed so that identification of data subjects is effectively made impossible.

In order to illustrate these methods, consider a database containing 'Person' objects, each consisting of a ['First Name', 'Last Name', 'Zip Code'] tuple. In this case, applying the above methods would mean:

- Synthesis: Clear all existing values and generate 100 new 'Persons' defined as ['Alice', 'Bobman', '12345'].
- Anonymisation: Replace all 'First Name' and 'Last Name' values by random strings. 'Zip Code' is no identifying field, so it is retained.

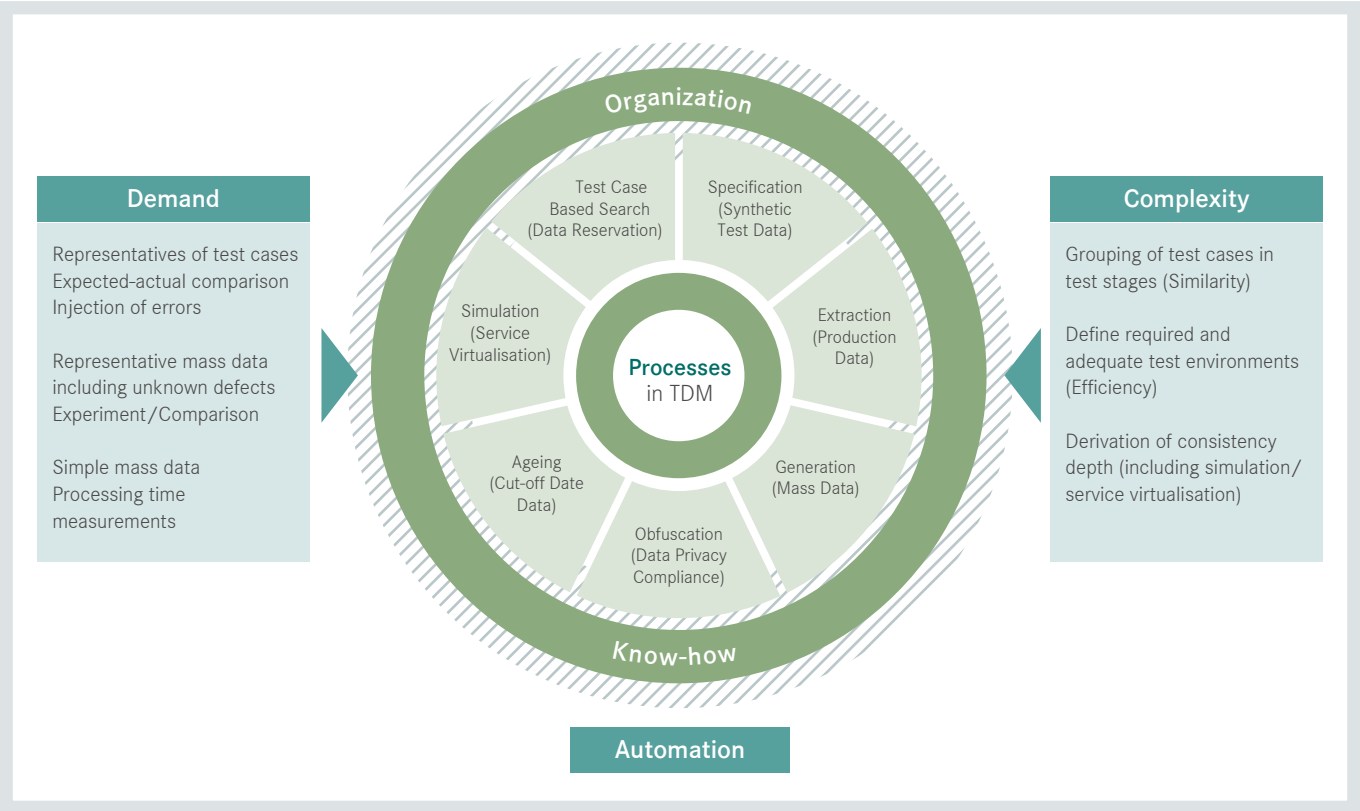

Figure 1: Test Data Management disciplines satisfy demand for data while having to account for the complexity of existing IT ecosystems.

Synthesis obviously is the cleanest approach to avoid personal data:

- Synthetic data are fictional, hence privacy compliance is given even in the absence of any additional security measures.
- Test data is generated to match test case requirements; there is no need for actually existing constellations to match test data demands.
- Consequently (and typically contrary to production data), synthetic test data can achieve a 100 % coverage for the defined test cases, optimally supporting functional test coverage and quality.
- Test data volume and related costs are significantly lower compared to using production copies.
- Test data provisioning is significantly more time-efficient, supporting test execution with respect to test environment set-up times.

Unfortunately, in many cases synthesis cannot be applied. Typical reasons are test case specifications missing comprehensive and precise test data requirements, or tests relying on production-like variety and amounts of data. Furthermore, complexity (and sometimes even incomplete knowledge) of data relationships can complicate and eventu-ally thwart efforts to generate comprehensive and consistent data sets. In such cases synthesis of test data may become prohibitively expensive or even impossible.

Until data quality, documentation and testing processes have become more mature, test data anonymisation often is the most feasible and pragmatic way to achieve compliance with data protection regulation.

Market – Current Status and Outlook

According to Nelson Hall, testing services represent roughly 6% of overall IT services spending and are growing by 2% annually. Given this, total global testing spending is poised to reach a volume of $40 billion by 2017.

The 2015/16 World Quality Report also researches the adoption of TDM techniques to provide test data. It finds that more than 30% of respondents use copies of production data without applying any kind of masking or anonymization, thus potentially violating the EU GPDR if they are doing business on the territory of the EU. In the same report, respondents also clearly indicate that providing test data in a compliant, yet cost-effective manner is a critical concern, that for many organizations remains unsolved as of today.

While several tool providers are available (e.g. Informatica, IBM, Compuware), providing a concrete test data anonymisation solution turns out to be quite similar to software development in miniature. Available technologies only provide frameworks demanding customization in order to match specific test data requirements for each organization. Designing and establishing test data anonymisation easily can require 100 to 200 man-days of expert's work for a single application.

With implications of data privacy violations – e.g. identity theft, whistleblowing or blackmail – hitting the headlines on a regular basis, established practice can no longer be regarded as secure. The later privacy compliance is achieved, the higher the probability for severe material or reputational damages. The EU GPDR will come into force in May of 2018, thus giving a deadline at which compliance must be achieved to remain within the constraints of EU law.

Hence, in order to control the increasing risks, companies will be required to account for data privacy until then – urgency depending on data sensitivity (amount of damage) and data protection measures already in place (probability for event of damage).

The Challenge

The client discussed here is a large, multinational finance corporation. Its IT portfolio contains more than 100 applications being maintained by internal IT management units. Management have taken a strategic decision to establish an IT outsourcing option and therefore prepare applications to be tested by third-party service providers.

An internal survey of the information security department lists about 40 applications to process sensitive personal data. Measures to provide privacy-compliant test data were applied only for a few applications, and were using specialised and isolated solutions. No consistent and standardised approach was available, neither in organisational nor technical terms, neither for test data management in general nor for anonymisation in particular. No external testing could be applied for applications processing sensitive data, therefore it was deemed crucial to change the way in which test data is derived from live data.

While the corporate-wide ETL platform Informatica PowerCenter was considered a tool candidate for test data anonymisation as well, its suitability for the task of anonymisation in the client's actual context was unproven.

One application was selected as a pilot for evaluating test data anonymisation. Existing test procedures included weekly 1:1 copies of production databases for system and acceptance testing.

The Assignment

Based on the starting situation outlined before, SQS was engaged to evaluate the centrally set ETL platform (here: Informatica PowerCenter including its Data Masking Option) in terms of whether it could serve as the company-wide test data anonymisation platform. The evaluation was split up into an abstract tool validation and a concrete technical pilot.

Goals of the tool validation were:

- Generally assessing the suitability of the existing ETL tool for test data anonymisation
- Conducting a market comparison of the existing ETL tool to other anonymisation tools

Goals of the technical pilot were:

- Specifically verifying the suitability of the existing ETL tool for test data anonymisation
- Surveying the feasibility of anonymisation for the purpose of system testing
- Delivering a working anonymisation solution that achieves compliance for parts of the pilot application already
- Observing scalability for corporate-level implementation

Technical Pilot

This section describes the technical pilot for test data anonymisation, referred to as 'TDA pilot'. The pilot application selected by the client is a data warehouse system (furthermore referred to as 'DWS') loading, consolidating and exporting data from or to various neighbouring systems. Its architecture of data layers corresponds to subsequent processing stages. Physically, each layer is represented by a separate database schema. Orthogonally to the layers, data can be divided into functional 'subject areas'. Subject areas can be viewed as a vertical partitioning permeating all the layers of the application. The data volume to be handled was approximately 500 GB, and was expected to be growing continuously. In addition to the production environment, two more environments exist for testing purposes:

- 'DWS provisioning environment' is used for system testing and UAT. Apart from application databases, it contains excerpts of relevant neighbouring systems for end-to-end testing. Databases were filled using 1:1 copies of production on a weekly basis.
- 'DWS test environment' is structurally identical to 'DWS provisioning', which also serves as a 1:1 copy template on demand. It is used for component and integration testing.

From the structures described above, the immediate consequence for data from the neighbouring system is that excerpts must be anonymised as well, thereby adding to the overall project effort.

Tool validation

The tool validation was conducted in several steps (see Figure 2).

From a long list of available tools capable of data masking (i.e. means to anonymise data), three tools were selected for a more detailed comparison based on market analysis and the project experience of SQS. Informatica PowerCenter was set as the corporate ETL platform by the client as part of a unified enterprise integration architecture. Based on that decision, Informatica PowerCenter was chosen as one of the candidates.

Features and capabilities of these tools then were rated in a two-stage process. The rating was based on a catalogue with approximately 100 criteria, seven of which were classified as 'knockout' criteria. As shown in Figure 1, the KO criteria were used in an absolute evaluation, whereas the remaining criteria were applied in a relative evaluation. In the end, the results were distilled into a tool recommendation.

The following selection of knockout criteria reflects the most important requirements of the customer, as well as the evaluation for the tool candidates. Each of the three candidates fulfils these KO criteria (even if only by means of workarounds in the case of one candidate):

- Client-specific knockout criteria
- Custom rules can be implemented
- Support for DB2, Oracle and flat files
- Support for restart/recovery/error handling
- Data sources and targets can be changed via configuration
- No limit for data volume
- Anonymisation is repeatable via automation
- Product is subject to active further development
- Affected attribute
- Satisfying anonymisation requirements
- Basic applicability, coverage of IT landscape
- Administrability
- Protection of investments
- Delivery in time and budget
- Maintainability
- Basic applicability

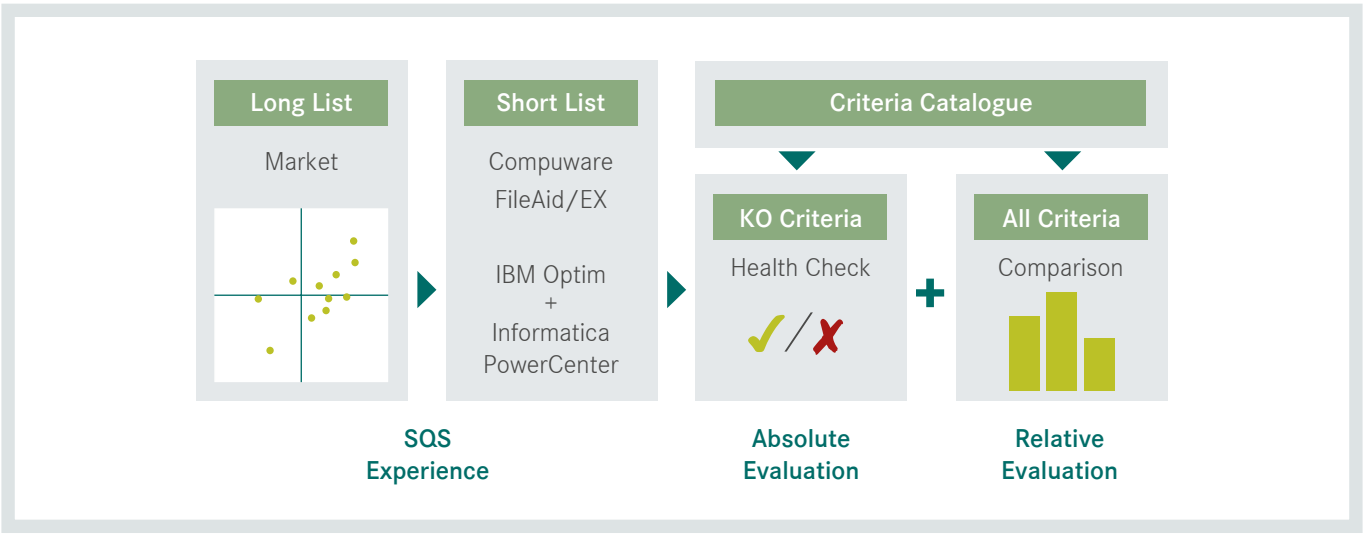

Figure 2: Tool validation process

The full criteria catalogue for relative evaluation was organised in seven requirement categories. The categories were weighted to reflect the client's specific requirements. According to this prioritisation, data provisioning and rule set were judged to be the most important, followed by (in descending order) administration, references, training, infrastructure and database analysis.

The existing ETL tool received the best rating in five of seven requirement categories (especially the most important ones).

Hence, before having conducted the technical pilot, both absolute and relative evaluation already supported employing the existing ETL tool for company-level test data anonymisation, especially in the context of the customer.

Pilot Scope

In order to satisfy the conflicting goals of yielding sufficient insights with regard to a corporatewide roll-out and of concluding the pilot as soon as possible, only a carefully chosen subset of the DWS, as well as its sources and targets, was included in the TDA pilot scope.

Vertically, only the two most important subject areas 'Person' and 'Deposit' were included. Based on a prior data structure analysis, twelve logical data fields containing personal data were expected to translate to 90 to 100 anonymisation-relevant physical data fields.

Horizontally, only those layers or database schemata were anonymised which directly participated in processing data from the selected subject areas.

Figure 3 shows a high-level view of the data storages in a DWS test environment. Vertical bars highlight the system parts participating in the TDA pilot. The reasons for tailoring the scope as described were:

- Reduced effort to implement anonymisation rules
- Touching various types of anonymisation requirements
- Full end-to-end test with regard to the chosen subject areas in order to validate anonymisation results
- Significant performance measurements possible (transfer of several complete database schemata)

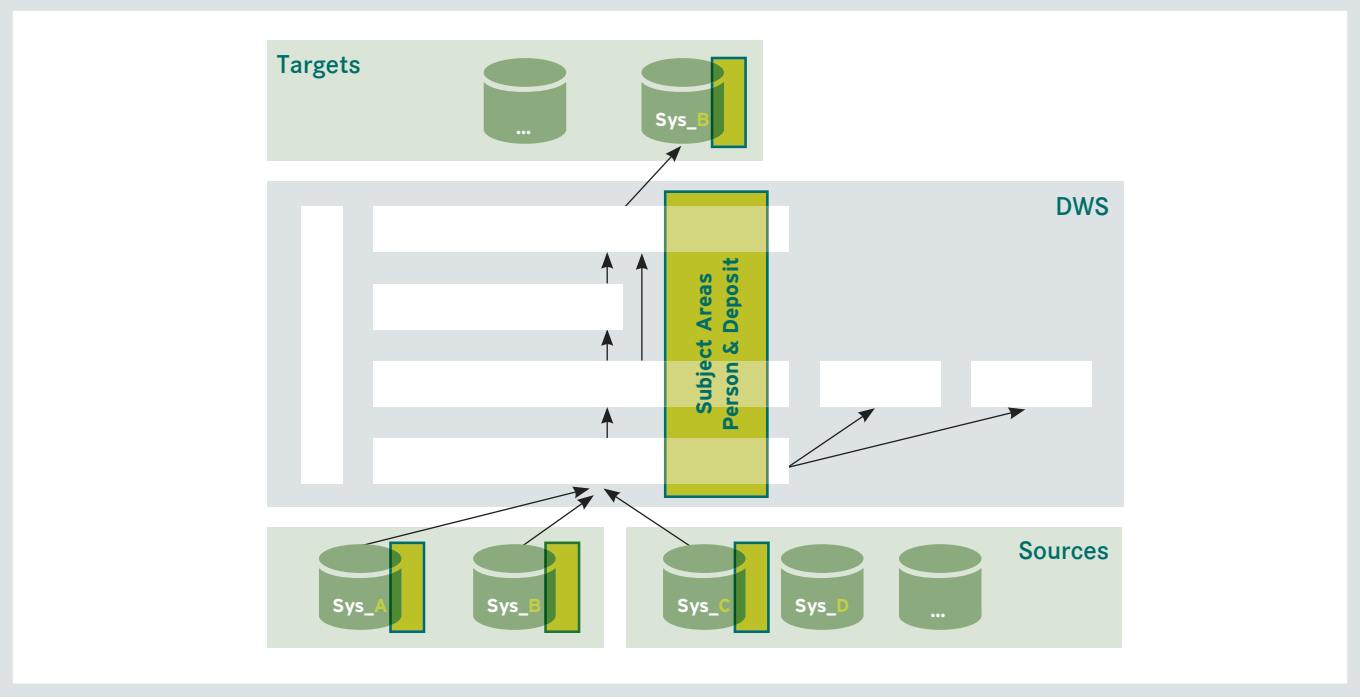

Figure 3: Scope of the test data anonymisation pilot

Concept

The purpose of the TDA pilot was to provide an anonymised test or development database environment for the piloted application 'DWS'. Anonymisation hence can be seen as a transport layer between two DWS database environments.

As shown in Figure 4, the TDA pilot development took place completely detached from the production environment. The DWS 'provisioning environment' used for acceptance testing also served as the anonymisation transfer source. This way, any negative side effects for production performance and data integrity could be ruled out effectively from the start.

Basic data flow

For the export and import of data, two basic data transport technologies were available: the ETL tool, and native database dump tools. Depending on the project requirements and technology characteristics, different concepts of technical architecture were balanced against each other. While changes in project parameters may require another design, the implemented solution promises to be reusable for other contexts as well.

The data transfer was designed to follow the KISS principle: 'Keep It Simple and Straightforward.' All database content is transferred directly from source to target using only Informatica PowerCenter. Anonymisation is applied during the transfer. This way, sensitive data never exists in the target environment.

Further concerns addressed by the chosen design were portability and reusability. Using the client's platform-independent standard ETL tool for transferring data promises manageable efforts when extending the solution to further applications. Only for some pre- and post-processing steps, specific Oracle scripts are currently used. For future extensions, these processing steps are to be ported to native Informatica, reducing platform dependency even further.

Anonymisation rules

The anonymisation rules were defined on the basis of the pilot goal to deliver a working anonymisation solution accomplishing compliance for the chosen subject areas.

The following logical data fields were included in the TDA pilot:

- Subject area 'Person': First name, last name, organisational unit, address (street, house number, city, zip code, PO box)
- Subject area 'Deposit': Deposit account number, sub-account number, account key number
- Both subject areas: Comment

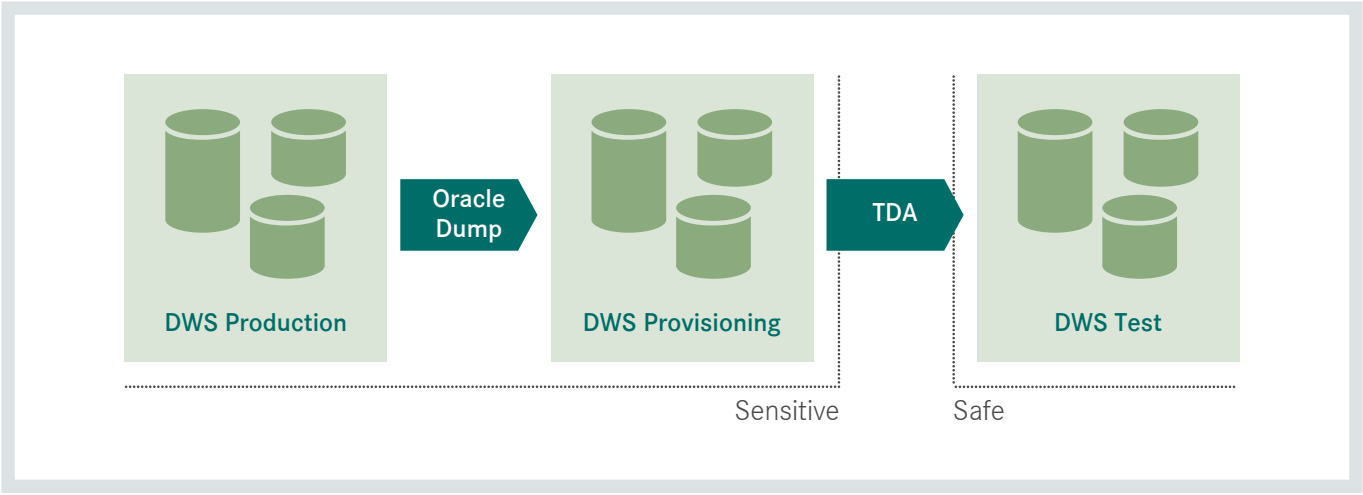

Figure 4: TDA pilot system environment and data flow

Every physical data field relating to one of these logical fields was recorded in a classification table. Serving as a pivotal point of specification for the test data anonymisation, the classification table essentially maps physical to logical data fields, and logical data fields to anonymisation rules. The table was extended during development and test to include newly found, additional physical instances of anonymised logical fields. Acceptance of the anonymisation solution was accomplished with 142 included physical fields. It should be noted that the sensitive physical field count is approximately 50 % higher compared to the definition of the project scope. Depending on existing knowledge about application data models, similar increases should be expected in other contexts as well.

On a side note: technical ID fields were not considered for anonymisation because they contain no personal data. Furthermore, risks for overall data consistency would arise if technical IDs (i.e. database primary and foreign keys) were included in anonymisation.

For every field, pseudo-random masking with numeric or alphanumeric characters was selected as anonymisation method (e.g. a name value 'John Doe' could be replaced by 'yZdPvnsH'). The method was judged 'good enough' for the planned purpose of testing processing flows and calculations of the Data Warehouse System. Although not part of the TDA pilot, additional requirements to data formats could easily be implemented to combine the masking with further common ETL transformation types.

In Informatica PowerCenter terms, this translates to 'Data Masking' transformations of the type 'Key Masking'. Values generated this way cannot be tracked back to the original values, effectively achieving anonymity and privacy compliance for that data record.

In case of logically identical fields (i.e. having the same logical field classification), equal source values are transformed into equal target values. Taking the above example, all occurrences of 'John Doe' values in 'Name' fields would be replaced by 'yZd-PvnsH', regardless of their location in the physical data model (database schema and table). This way, functional key relationships are effectively preserved. In Informatica PowerCenter, this is achieved using the same 'Random Seed' for logically identical data fields.

Design

In Informatica PowerCenter, the actual anonymisation procedure is stored as a modular composition of Informatica objects. Each database table is transferred using a 'Mapping' containing 'Source' and 'Target' definitions, connected by 'Transformations' (with anonymisation, if applicable). Each mapping is embedded in a 'Session', and all sessions belonging to one schema are compiled in one executable 'Workflow'. Everything taken together constitutes the Informatica configuration used for creating an anonymised copy of a DWS database environment.

Considering the complexity of the application, this results in a significant number of configuration objects. In case of application data model changes, a lot of these objects potentially require maintenance work. Rather than manually creating and maintaining the configuration, an XML generator framework was developed during the TDA pilot using Groovy scripts and XSL Transformations. It generates a complete Informatica PowerCenter XML configuration based on the following inputs:

- Environment configuration (small XML configuration file)
- Database schema (extracted automatically)
- Mapping of anonymisation rules to physical data fields (specification in Microsoft Excel)

The generated XML configuration then is imported into the tool in order to execute it.

Manual maintenance is required only for environment configuration (e.g. connection parameters) and anonymisation rule mappings. In case of the latter, the Excel Spreadsheet file format was chosen in order to allow all stakeholders to contribute easily to the mission-critical rule-to-fieldmapping.

For efficient reusability, anonymisation rules are designed separately and application-unspecific. They are stored as 'Reusable Transformations' and 'Mapplets' in a central Informatica PowerCenter repository, and referenced in the individual mappings as described above.

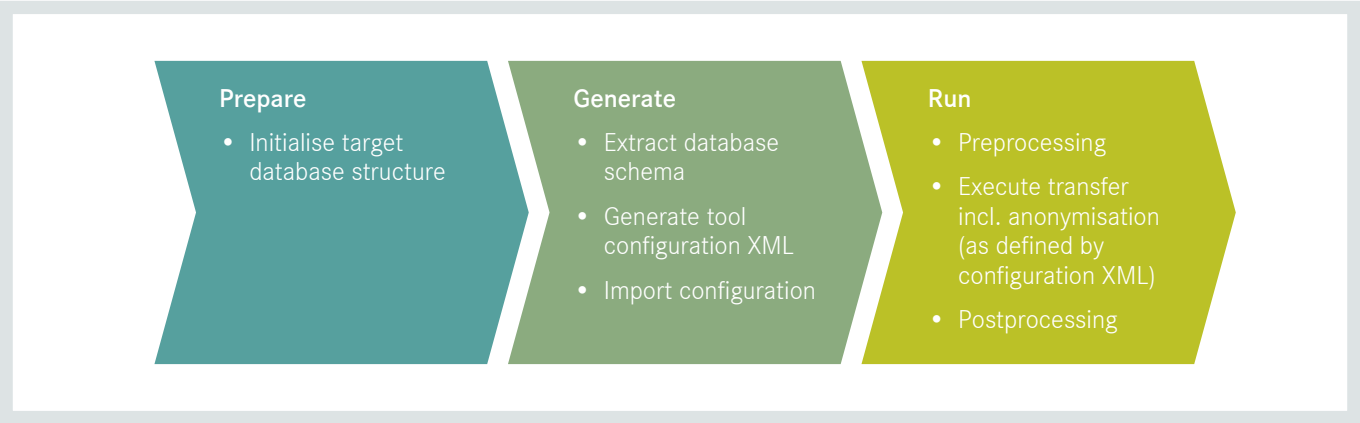

Figure 5: Anonymisation execution sequence

Figure 5 shows the complete anonymisation sequence. Step 1, 'Prepare', ensures existing and up-todate target database structures. Step 2, 'Generate', is optional: if data model, environment and anonymisation specifications have not changed since the last run, the existing Informatica XML configuration can be reused. Step 3, 'Run', then realises the actual data transport shown as the 'TDA' arrow in Figure 4: TDA pilot system environment and data flow.

Test concept

From the nature of test data anonymisation, two main test missions arose:

1. Testing the anonymisation process: White-box test of data transport mechanisms
2. Functional regression testing of the application using anonymised data: Black-box test using existing test cases of the application

The missions stated above yield the following test objects:

- Data storage: Completeness (quantity) of data transfer
- Data record: Correctness (quality) of anonymised data
- Anonymisation rules: Correctness and compliance of anonymisation rule design
- Application: Proper operation of the DWS application using anonymised data

Given the project goals and test missions, testing focused on the quality attribute of functionality. Efficiency was measured, but not tested explicitly. Maintainability was not explicitly tested either, but evaluated during the incremental development of the pilot. The quality attributes of reliability, usability and portability were not tested during the TDA pilot.

After assessing possible risks relating to the test objects described above, the test methods in Figure 6 were chosen for component and system/acceptance tests (risk-based testing).

The test specification and procedure details were documented in a separate test concept according to the client's documentation standards.

Test Stage / Test Object	Component Test	System and Acceptance Test
Data Storage	Explorative test	Decision table test
Data Record	Explorative test, checklists	Checklists, use case test
Anonymisation Rules	Error guessing	Equivalence class test, compliance review
Application	n/a	Use case test

Figure 6: Applied test methods

Results

Implementing and testing the anonymisation solution confirmed the tool validation results (see above) with regard to Informatica PowerCenter. It proved to be reliable, sufficiently powerful with respect to functional and non-functional anonymisation requirements, and efficient to use. Weaknesses of its Data Masking Option – e.g. with respect to anonymisation functionality – could be compensated by other capabilities of Informatica PowerCenter (e.g. 'Lookup' or 'Expression' transformations).

The status of the anonymisation test objects for release:

- Data storage: All table contents of the source environment were transferred successfully.
- Data record: The data fields participating in the TDA pilot were transformed correctly according to agreed anonymisation rules.
- Anonymisation rules: The correctness of anonymisation rule design was implicitly tested in the context of the test object 'Data record'. All rules transformed source values as specified.
- Application: Both subject areas were tested according to DWS specification with positive results.

During the test of the participating subject areas 'Person' and 'Deposit', anonymisation-relevant data fields were identified which either had previously been incorrectly classified as 'non-sensitive' or had not existed at the time of specification. For example, several data fields were added in a new release of a neighbouring system of the DWS application. Because excerpts of the neighbouring system are mirrored in the DWS test environment and hence are subject to anonymisation, the TDA pilot had to include 19 new fields of the neighbouring system as well. Furthermore, several previously unspecified consistency requirements were revealed, again increasing the set of anonymised data fields.

As a benefit from the efficient XML configuration generation process, these fields could be included in the anonymisation with marginal effort. However, additional test cycles were required, utilising the project's risk buffers.

Testing the anonymisation solution yielded 38 issues, 33 of which could be resolved before project closure. The remaining five issues were either rated as 'minor' or did not affect acceptance. For acceptance, eight test cases of the application were executed successfully after anonymisation, all in all performing a complete source-to-sink warehouse load process. Anonymised data belonging to both subject areas and all included databases was utilised without having any impact on application functionality. The test conception proved successful and was considered to significantly contribute to overall satisfaction of the (internal) customer team DWS. Limited to the subject areas included in the pilot scope, application testing can now be carried out in accordance with privacy compliance test data requirements. The remaining subject areas can easily be included in the course of the corporate roll-out of test data anonymisation.

Conclusion and Outlook

Relating to the goals of test data anonymisation as stated in the section “The Assignment”, the TDA pilot achieved the following statements:

- **Verification of suitability**
  Informatica PowerCenter is a suitable tool for test data anonymisation. Every functional and non-functional requirement from the application owner’s side could be implemented successfully. Remaining issues are of minor relevance, and can be resolved in follow-up projects.
- **Feasibility of anonymisation for system testing**
  Anonymised data was utilised successfully in test cases of the application (see above). The pilot yielded no fundamental concerns against a company-wide usage of anonymised data for system testing.
- **Working anonymisation solution for parts of the pilot application**
  The subject areas ‘Person’ and ‘Deposit’ were anonymised completely and consistently. Hence, limited to these subject areas, the anonymisation solution is working as required for compliance. In order to achieve full data privacy compliance, however, the anonymisation solution shall be extended to cover the remaining subject areas as well.
- **Scalability for corporate-level implementation**
  The framework developed during the TDA pilot is well suited for introducing test data anonymisation to other applications as well. Special focus was placed on portability, reusability and extendability, as well as a high degree of automation during execution. Due to the applied technologies and designs, the TDA framework can be efficiently upgraded for the company-wide roll-out.

The TDA pilot goals thus were successfully achieved.

In subsequent projects, the pilot may easily be enhanced to satisfy operational capability requirements in an enterprise level environment of operation (supported platforms, central monitoring and control, etc.). Optionally, integrating data reduction can be evaluated using the ‘Informatica Data Subset’.

Eventually, test data anonymisation shall be rolled out on a corporate level, reaping the benefits of the reusable and portable design applied to the pilot. Privacy compliance can be obtained now using a centralised, systematic and scalable approach. Several valuable lessons learned from the pilot hold true for arbitrary corporate roll-out scenarios as well:

- Completeness and correctness of anonymisation requirements are crucial for the applicability of anonymised data. Gaps with respect to physical data fields as well as implicit and explicit consistency requirements can cause serious error conditions in the application. Apart from rendering the anonymised data unusable, general acceptance of anonymised data for system testing can be questioned. Support of concerned IT management units hence is mandatory and must be accounted for in structures and budgets.
- Centralised management of anonymisation rules does not contradict nation-specific anonymisation rules. Different legal spaces (e.g. because of different national legislation) can be accounted for using separate but centrally managed anonymisation rule template sets. This way, consistency and reusability can be retained, maximising efficiency and quality of test data anonymisation.

In general, this case study demonstrates that achieving compliant test data acquisition is not beyond the reach of organizations, even if they are starting from a basic level of implementation. However, it is crucial not only to select the right way to implement concrete services such as Test Data Acquisition but also to ensure that the right services are implemented and that they are integrated into a Test Data Management concept that serves the requirements of the organization. Now may be the perfect time to tackle this issue as the impending EU regulation offers good visibility of regulatory requirements while any solution can be tailored to cover the remaining requirements towards test data management as well.

References

1. European Parliament, Council. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. EUR-Lex – Access to European Union Law. [Online] 24/10/1995. [Cited: 02/11/2011.] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:NOT.
2. Consolidated Version of the Treaty on the Functioning of the European Union. EUR-Lex – Access to European Union Law. [Online] 09/05/2008. Cited: 22/11/2011.] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:115:0047:0199:EN:PDF.
3. European Commission. Commission's First Report on the Transposition of the Data Protection Directive. European Commission Justice. [Online] 03/05/2003. [Cited: 22/11/2011.] http://ec.europa.eu/justice/policies/privacy/lawreport/report_en.htm.
4. Karsten Leclerque, Pierre Audoin Consultants (PAC) GmbH. Outsourcing, Beratung, Wartung – Was die Zukunft den Providern bringt. Computerwoche.[Online] 26/10/2011. [Cited: 11/11/2011.] http://www.computerwoche.de/management/it-services/2369789/.Pierre Audoin Consultants (PAC) GmbH. Growth Market Software-Testing –Market Trends, Service Providers, and Success Factors. Cologne, Germany: SQS Software Quality Systems AG, 2011.

SQS Software Quality Systems AG
Phone: +49 2203 9154-0 | Fax: +49 2203 9154-55
info@sqs.com | www.sqs.com