WHITE PAPER

Libelle Data*Masking*
Automated and Optimized Data Anonymization

The issue of data protection and data anonymization is becoming more and more important, and this focus did not start with the adoption or entry into force of the EU General Data Protection Regulation (EU-GDPR). The General Regulation defines a Europe-wide minimum standard for data protection, which can be further tightened regionally. This influences the handling of sensitive data in companies and affects IT environments with relatively few access restrictions to potentially sensitive data: the non-productive test and development systems. In order to be able to adequately counteract the real danger of unwanted data access or even data outflow, the ordinance stipulates the data stock to be pseudonymized (Article 4.5). However, pseudonymization is per se reversible and allows conclusions to be drawn about original data. Anonymization of data stock goes one step further towards irreversibility.

Whichever variant a company chooses to meet the objectives, logical relationships should also be taken into account in the pseudonymized or anonymized data stock also in the test systems. Otherwise, test-based statements on process quality, consistency, and logical correctness of business processes within IT environments remain only partially reliable.

Libelle **Data***Masking* (LDM) enables this anonymization and pseudonymization of critical data in non-productive systems while taking logical contexts into account. LDM can be used universally and independently of the application software, for both individual systems and complete system landscapes.

The result is anonymized, consistent data for the development and testing of software and business processes across all platforms. The anonymization methods used provide realistic, logically correct values to describe and test almost all business cases.

Libelle **Data***Masking* supports comprehensive and cross-system data sets in SAP® **and** non-SAP systems.

New standards, new rules, growing challenges: Focus on data protection

Corporate investment in a wide variety of measures to protect productive environments from unwanted data access and data outflow should not be seriously questioned these days. What is often put off until later, however, is an examination of non-productive environments such as development, testing, or project systems.

For one thing, experience has shown that much less effort is put into physical and logical protection mechanisms, and for another, the requirements for access privileges, for internal and external developers and testers, for instance, are considerably less stringent than in production environments. However, if companies decide to use system and landscape copies to add real data to non-productive environments, sensible data becomes essentially openly available. This opens the door to unintentional data outflow. And this is exactly where the EU General Data Protection Regulation comes in.

The issue of data protection is becoming more and more important, and this focus did not start with the adoption or entry into force of the EU General Data Protection Regulation (2016/679). This basic regulation provides the first minimum data protection requirement that is binding throughout Europe, and it actually must be implemented in all member states by 25 May 2018. In many cases, it even includes additional country-specific requirements.

This has an impact on all companies that do business anywhere in the European market: They must meet the requirements of all countries in which they conduct legal transactions.

The primary focus of the EU GDPR is on pseudonymization, which still allows a general back-calculation of the data. Moreover, in Germany, as in other countries, there are other official regulations and recommendations. More specific international anonymization rules, such as the Payment Card Industry Data Security Standard (PCI-DSS), the Health Insurance Portability and Accountability Act (HIPAA) and the Family Educational Rights and Privacy Act (FERP) must also be observed.

External auditors, data privacy activists, auditing departments and, not least important, the management levels of companies are also very interested in more far-reaching methods that prevent unwanted data outflow or at least remove criticality when it occurs.

One way to do this is to simply delete or overwrite critical live data in non-productive environments. However, the individual data sets then completely lose their logical connections and dependencies to each other, and thus their usefulness. This makes most of the test scenarios worthless that focus on business functions or even end-to-end business processes.

The question is, how can testing of business functions and business processes continue with logically meaningful data, but without “unauthorized” access to original information?

In order for non-productive environments to continue to fulfil their purpose, it is therefore necessary to ensure that reasonable data is made available from a logical point of view. At this point, it is therefore not enough to simply delete or overwrite data as mentioned above; instead, it is imperative that attention be given to logical meaningfulness and connections.

This is where the pseudonymization requirements take effect (or one step further along, anonymization). The latter no longer allows back-calculation of original data, which is what many examiners and responsible individuals want to avoid.

However, depending on the type and size of the application environment and the business processes being evaluated and the number and intensity of relationships between various data sets, the matter can become complex.

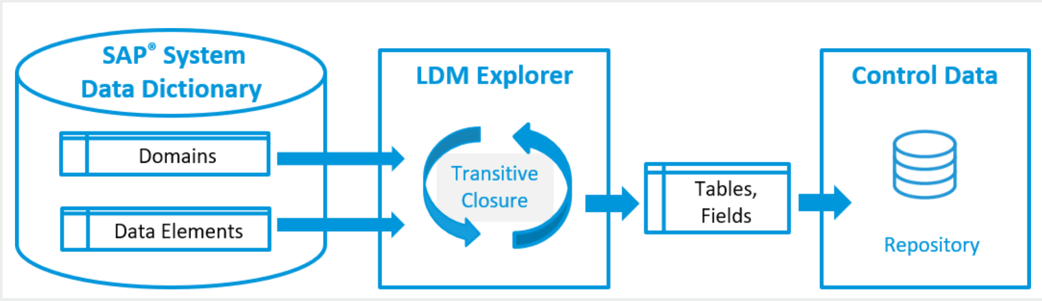
A comparatively simple way to cope with this complexity is to use tools that are specifically designed for the purpose.

Anonymization with Libelle Data*Masking*

The Libelle **Data***Masking* solution makes it possible to easily process data stock of any complexity with anonymizing methods of any complexity. Standard out-of-the-box mechanisms are already available for this purpose. They can be quickly and easily applied to data fields identified as critical and sensitive. In addition, functional and/or technical managers can set up rules of any complexity with which, for example, even highly structured and identifying data sets can be calculated.

Detection of sensitive data fields: LDM Explorer and LDM Associator

One of the central components of LDM is LDM Explorer. It analyses the data stores that are to be anonymized once during the initial set-up and regularly after relevant changes to the data model of the corresponding application. This analysis is usually based on Data Dictionary in SAP environments, for example, and the physical data model in applications without Data Dictionary. The result of the Explorer run is an overview of field types that have been determined to be potentially sensitive.



LDM Explorer

LDM Explorer runs iteratively to achieve a complete final result.

The results can be added to and edited to include or exclude customer-specific data.

The results of LDM Explorer are automatically linked to typical anonymization procedures via the LDM Associator.

The results of LDM Explorer and LDM Associator (that is, the list of detected fields, the associated field types and the allocated anonymization procedures) can be modified by authorized individuals, and unrecognized fields can be added.

The IT world is complex. LDM ensures consistency

However, LDM does not only support the anonymization of individual, independent data stock. One of its central features is the ability to anonymize completely across system boundaries. When LDM is used in complex, interdependent landscapes, LDM Explorer ensures that the occurrence of a given data object in different environments is recognized. This forms the basis for subsequent, logically thorough and consistent anonymization of related data records across system boundaries.

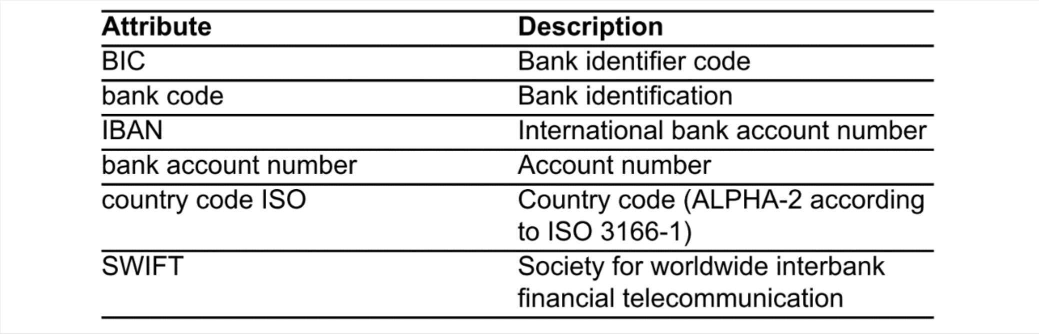
In this way, LDM enables the user companies to run and test end-to-end business processes, which can lead to a significant increase in overall process quality: The simultaneous anonymization of all occurrences of an attribute across all landscapes means the data “fit together” just as well after anonymization as real data would. For instance, all affected indexes are rebuilt after anonymization, of course.

An example of this are those business process transactions that are completely anonymized across several systems, such as an order process from a web portal to the CRM system, warehouse management, or production, through accounting, to the data stores of the logistics and after-sales departments. This allows test environments to evaluate not only individual systems, but also complete business processes.

Typically critical. The LDM anonymization profiles

Usually, similar data structures keep being determined to be sensitive and critical in different applications and data stores. To make it easier to find and manage these structures, LDM offers typical anonymization profiles out-of-the-box.

In these profiles, logically related fields and field types are described in a comprehensive way. For example, a frequently used “Bank Connection” anonymization profile consists of the following fields:



Attribute	Description
BIC	Bank identifier code
bank code	Bank identification
IBAN	International bank account number
bank account number	Account number
country code ISO	Country code (ALPHA-2 according to ISO 3166-1)
SWIFT	Society for worldwide interbank financial telecommunication

Anonymization profile

Example of the “Bank Connection” anonymization profile.

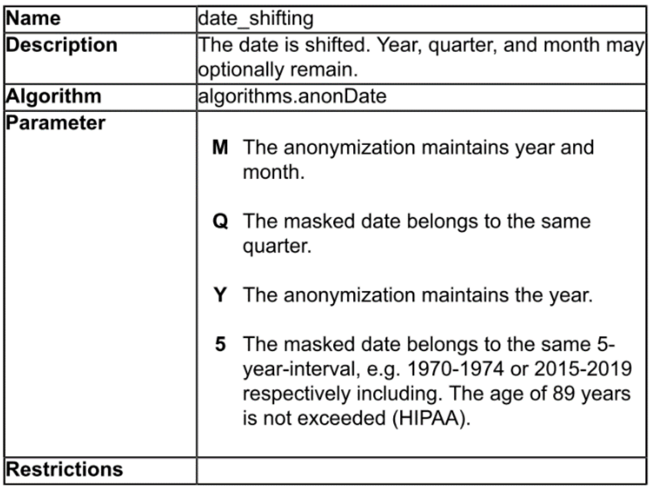
LDM currently offers around 75 standardized anonymization profiles out-of-the-box.

Libelle constantly maintains this profile, and about 75 others. In this way, changes in business application data structures are continuously taken into account and made available to LDM users.

No longer real, but still realistic: The LDM anonymization procedures

The anonymization methods of LDM offer realistic-looking values. Among the most important of the **standard procedures** available out-of-the-box for Libelle **Data***Masking* are:

- Delete data
- Overwrite data with constants
- Name anonymization with realistic first and last names for individuals based on internal or external reference values. External values available for many regions.
- Address anonymization with addresses that comply with postal standards, based on internal or external reference values. External values available for many regions.
- Date shifts within parametrizable limits
- General and particular number anonymization, such as unique person/ personnel numbers within predefined framework values
- Anonymization of bank data on the basis of generally regulated specifications, such as realistic account details with correct check digits, realistic credit card numbers with correct check digits
- Uniformity for ICD codes in the medical field



Name	date_shifting
Description	The date is shifted. Year, quarter, and month may optionally remain.
Algorithm	algorithms.anonDate
Parameter	**M** The anonymization maintains year and month. **Q** The masked date belongs to the same quarter. **Y** The anonymization maintains the year. **5** The masked date belongs to the same 5-year-interval, e.g. 1970-1974 or 2015-2019 respectively including. The age of 89 years is not exceeded (HIPAA).
Restrictions	

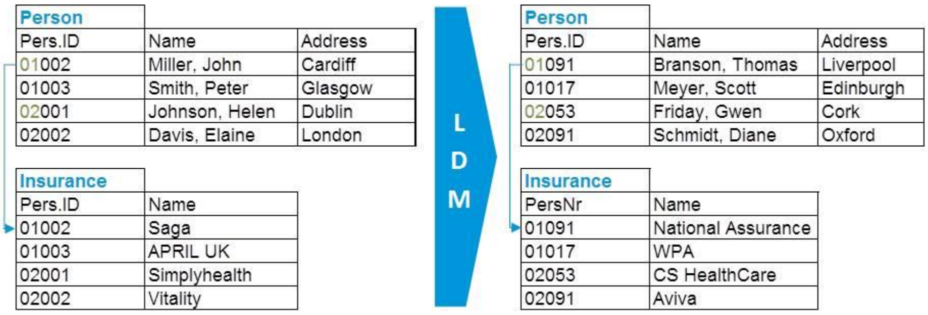
Anonymization methods

Example: Date shifting

LDM currently offers around 25 standardized anonymization profiles out-of-the-box.

In addition, further procedures of any complexity can be defined, such as the following:

- Calculation of structured, descriptive customer/contract numbers in certain sub-number ranges in the insurance and trading sectors
- Calculation of index values within the framework of predefined value classes, such as damage or other defined value classes, in the insurance and trading sectors
- Calculation of structured electricity meter numbers in the energy sector
- Other procedures provided by LDM customers and the community in general



Person		
Pers.ID	Name	Address
01002	Miller, John	Cardiff
01003	Smith, Peter	Glasgow
02001	Johnson, Helen	Dublin
02002	Davis, Elaine	London

Insurance	
Pers.ID	Name
01002	Saga
01003	APRIL UK
02001	Simplyhealth
02002	Vitality

Person		
Pers.ID	Name	Address
01091	Branson, Thomas	Liverpool
01017	Meyer, Scott	Edinburgh
02053	Friday, Gwen	Cork
02091	Schmidt, Diane	Oxford

Insurance	
PersNr	Name
01091	National Assurance
01017	WPA
02053	CS HealthCare
02091	Aviva

Consistency

Consistency through simultaneous anonymization of all occurrences of a data object in all relevant data stores.

The actual anonymization run – LDM A-RUN

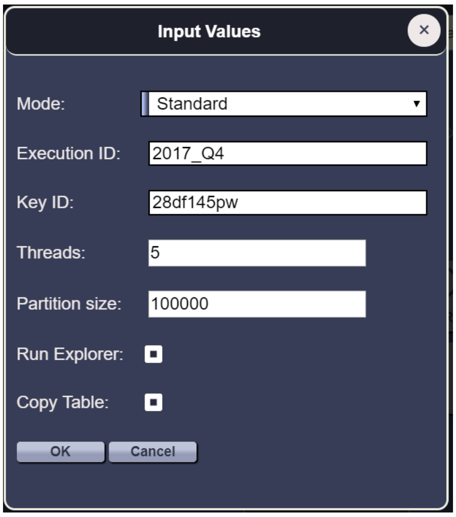
The basic objective of LDM is to anonymize productive data in non-productive environments. Several application variants are conceivable, of which the two most frequent are described here:

- Variant 1 – local anonymization: Anonymization is performed locally in the data stores, and no other upstream or downstream systems are involved.

- Variant 2 – System refresh and subsequent anonymization: This is a classic case in a homogeneous SAP system copy. Integration of LDM into a system-refresh procedure – seamlessly as an add-on in the Libelle **System***Copy* solution, for instance, allowing complete end-to-end automation.

Each individual anonymization run is created explicitly and documented. This ensures the ability to provide information and report on the current status of a data stock with regard to its anonymization at all times.

When an anonymization run is created, a so called Anonymization Key is also specified, among other things. This key consists of freely selectable phrases (numbers, letters, special characters) of any length. Given unchanged data stock and settings, anonymization runs that have identical keys lead to identical anonymization results. In this way, it is possible to carry out test runs in which, for example, the correct application of individual, new anonymization procedures can be verified by appropriate bodies.


Input Values

Mode: Standard

Execution ID: 2017_Q4

Key ID: 28df145pw

Threads: 5

Partition size: 100000

Run Explorer: □

Copy Table: □

A-RUN with anonymization key

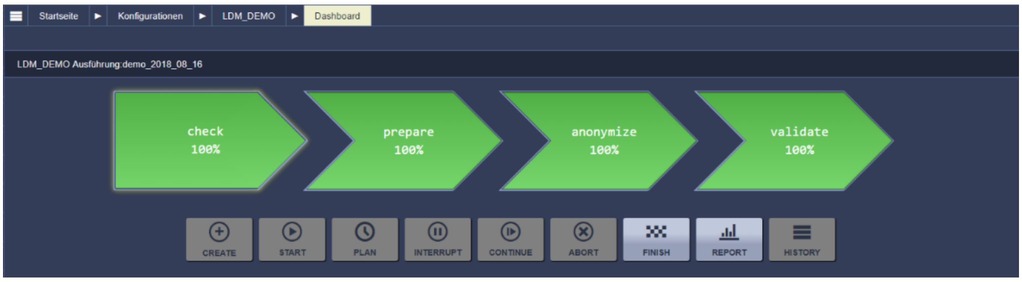
Each anonymization run is created explicitly and provided with an anony-mization key.

Identical data sets and identical keys lead to identical results.

Random or at least regularly changed keys should be used for the ultimately "real" anonymization of "real" data stock, of course.

Everything at the push of a button. LDM configurations and configuration patterns

One or more configurations are created for all related databases, files, and anonymization profiles. These include all technical data and specifications that are relevant for anonymization. Once a configuration has been defined and tested, it can be used as often as required for anonymization runs. The configuration is stored in the control database, so the anonymization run can be performed automatically and unattended – at night, for example.



Push the button

After one-time configuration, it is only the push of a button to start an anonymization run.

Executing the preparation, the anonymization itself until validation, manage and monitor all the process via LDM GUI

Standard configuration patterns prepared for SAP® modules (HCM, FI, CO, etc.) shorten the time for individual customizing of the data to be anonymized. For other standard applications, users can create individual configuration patterns and exchange them with other users via Libelle.

Contact us!

Libelle AG
Gewerbestr. 42
70565 Stuttgart
Germany

T: +49 711/78335-0
sales@libelle.com

www.Libelle.com