LIVRE BLANC

Libelle Data*Masking*
L'anonymisation automatisée et optimisée des données

Le Règlement général sur la protection des données (RGPD) définit un standard minimum de protection des données pour l'Union Européenne (UE). Ce Règlement pourra également être renforcé au niveau national. Pour les sociétés, ces évolutions réglementaires influent nécessairement sur la gestion des données sensibles. Cela est notamment vrai pour des environnements jusqu'alors peu protégés : les environnements hors-production tels que les environnements de tests et de développement.

Afin de prévenir les risques d'accès frauduleux aux données ou encore de fuite des données, le Règlement propose entre autres la technique de la pseudonymisation des données (art. 4.5). Cela étant, la pseudonymisation est un mécanisme *de facto* réversible. Dès lors, il permet de tirer des conclusions quant aux données d'origine. L'anonymisation, quant à elle, offre un niveau de sécurité supplémentaire en ajoutant le critère de l'irréversibilité.

Quel que soit le mécanisme retenu par une organisation, les relations logiques entre les données doivent être prises en compte. En effet, si elles sont ignorées, les résultats des tests portant sur le processus de qualité, la cohérence des données, ou encore l'exactitude logique d'un processus métier, ne seront que partiellement fiables.

Libelle **Data***Masking* (LDM) permet l'anonymisation et la pseudonymisation des données critiques dans les systèmes hors-production en prenant en compte le contexte logique de la donnée. LDM peut être utilisé indépendamment de l'application utilisée, tant pour des systèmes isolés que pour des paysages variés.

Libelle **Data***Masking* permet d'obtenir des données anonymisées et cohérentes tant pour le développement que pour le test des applications et des processus métiers. Les méthodes d'anonymisation utilisées fournissent des valeurs réalistes qui respectent la logique des données remplacées.

Enfin, Libelle **Data***Masking* prend en charge les données réparties sur des systèmes hétérogènes.

De nouvelles règles, un nouveau cadre : Focus sur la protection des données

Aujourd'hui, il ne s'agit pas de remettre en cause les nombreux investissements réalisés dans la sécurité des systèmes de production informatique. Par contre, la sécurité des systèmes hors-production tels que les systèmes de développement, de test, ou encore de projets a, elle, souvent été mise de côté.

D'une part, l'expérience montre que généralement peu d'investissements ont été réalisés dans les outils de protection physique et logique de ce type de système. D'autre part, les conditions d'accès aux systèmes hors-production sont nettement moins contraignantes que les conditions d'accès à la production. Or, lorsqu'une société décide de copier des données de production sur des systèmes hors-production, les données sensibles deviennent alors plus vulnérables. Cela ouvre notamment la porte au vol de données. C'est ce type de situations que le Règlement européen sur la protection des données (RGPD) cherche à éviter.

Le RGPD fournit le socle minimum d'exigence pour la protection des données au sein de l'Union Européenne. Ce Règlement est entré en vigueur le 25 mai 2018. Dans de nombreux cas, ce socle d'exigence sera complété par des exigences nationales spécifiques.

Ce règlement a un impact sur les entreprises qui sont établies dans les pays de l'Union Européenne. Les entreprises devront être en accord avec le RGPD et les lois s'appliquant dans chaque pays où les données sont traitées.

Dans un premier temps, le RGPD propose la pseudonymisation comme outil de protection des données. Ce mécanisme de transformation de la donnée permet un retour à la donnée d'origine. Ainsi, il s'agit d'un outil de protection des données critiquable, dans le sens où l'accès à la donnée originale est encore possible. Par ailleurs, certaines règles nationales (en Allemagne notamment) et internationales (le PCI DSS par exemple) ne toléreraient pas la pseudonymisation des données comme mécanisme de sécurité.

Dès lors, il convient de trouver d'autres méthodes qui permettraient de prévenir le vol de données ou de supprimer l'aspect critique des données.

Une des options envisageables serait de supprimer purement et simplement les données critiques présentes dans les environnements hors–production. Les jeux de données perdraient alors leur caractère interdépendant. Et les scénarios de tests perdraient de leur valeur.

Dès lors, comment tester les fonctionnalités et les processus métier avec des données cohérentes sans pour autant donner accès à des données originales sensibles ? Afin que les environnements hors-production puissent continuer à remplir leurs fonctions

(Développement, Test…), il est nécessaire que les logiques métiers soient respectées et, pour cela, que les données nécessaires à la logique soient disponibles.
Dans ce cadre, l'anonymisation apporte une réponse.

Contrairement à la pseudonymisation, l'anonymisation ne permet plus de retour en arrière. Dans certains cas, il est vrai que le nombre de processus métier à évaluer, le type et la taille de l'application ou encore le degré d'interconnexion entre les données peuvent rendre l'activité d'anonymisation complexe.

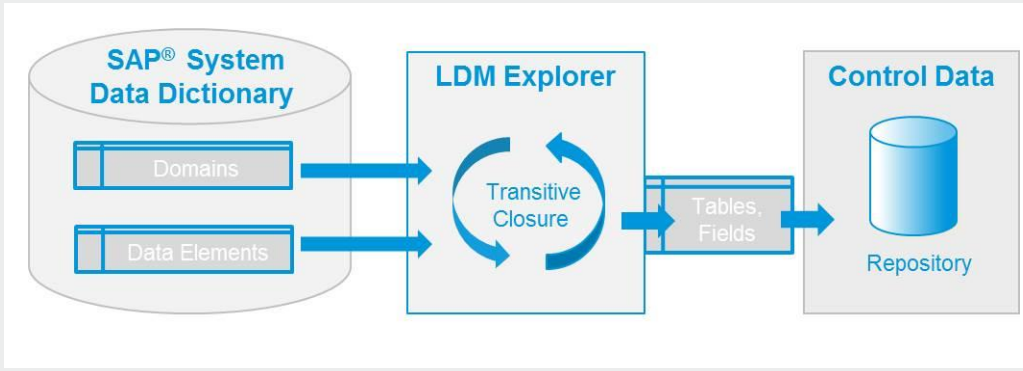
Afin de se départir de cette complexité, il peut être intéressant d'opter pour l'utilisation d'une solution dédiée.

L'anonymisation avec Libelle Data*Masking*

Libelle **Data***Masking* (LDM) est un logiciel capable de traiter des masses de données de complexité diverse. La solution utilise des méthodes d'anonymisation variées et offre des mécanismes d'anonymisation standards. Ces mécanismes peuvent facilement être appliqués aux données estimées sensibles. Par ailleurs, les responsables métiers comme les responsables techniques peuvent créer des règles spécifiques et de tout type de complexité.

La détection des données sensibles : LDM Explorer et LDM Associator

Un des composants centraux de Libelle **Data***Masking* est LDM Explorer. LDM Explorer analyse les entrepôts de données qui pourront être anonymisées. L'analyse est réalisée une seule fois lors de la configuration initiale de la solution puis à chaque changement de modèle de données. Pour les systèmes non-SAP, l'analyse est menée depuis le modèle physique de données. Pour les systèmes de type SAP, l'analyse se fonde sur les dictionnaires de données. L'exploration donnera une vue d'ensemble des types de champs qui contiennent les données potentiellement sensibles.



LDM Explorer

LDM Explorer s'exécute de manière itérative afin d'atteindre un résultat complet.

Les résultats obtenus peuvent être ajoutés, modifiés ou exclus du résultat final.

L'outil LDM Associator, lui, associera directement les résultats fournis par LDM Explorer aux procédures d'anonymisation standard proposées dans la solution.

Les champs détectés par LDM Explorer et les associations réalisées par LDM Associator peuvent être modifiées par des utilisateurs autorisés et des champs qui n'auraient pas été répertoriés pourront être ajoutés.

LDM et la cohérence des données

Libelle Data*Masking* peut prendre en charge plus que des entrepôts de données isolés. La solution anonymise également des ensembles de systèmes complets, homogènes ou hétérogènes. Lorsque **Libelle Data***Masking* est utilisée dans des environnements interdépendants, LDM Explorer s'assure que l'occurrence d'un objet présent dans plusieurs environnements est reconnue. En respectant ce principe de cohérence des données, **Libelle Data***Masking* permet une anonymisation logique des données pour des environnements complets.

En ce sens, **Libelle Data***Masking* permet aux utilisateurs d'exécuter et de tester des processus métier de bout en bout. Cela permet d'atteindre une qualité globale des processus importante. Il est à noter que tous les index existants seront reconstruits après anonymisation.

Par exemple, les processus métier qui incluent à la fois le portail Web, le système CRM, le système de Gestion des entrepôts, la Production et la Comptabilité seront anonymisés tout en respectant chacun des processus.

Les profils d'anonymisation Libelle Data*Masking*

Habituellement, les données identiques conservent leur caractère sensible quel que soit l'entrepôt de données où elles se trouvent et quelle que soit l'application. Afin de faciliter la gestion de ces caractéristiques, **Libelle Data***Masking* offre des profils d'anonymisation standard. Dans ces profils, les types de champ et les champs reliés sont décrits de façon complète. Par exemple, le profil « Connexion bancaire » comprend les champs suivants :

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

Profil d'anonymisation

Exemple du profil «Connexion Bancaire». Il contient 6 attributs, dont le BIC et l'IBAN.

LDM offre 76 attributs d'anonymisation standards.

Libelle maintient en permanence les attributs contenus dans ces profils, soit 76 attributs standards à date.

Les procédures d'anonymisation de Libelle Data*Masking*

Les méthodes d'anonymisation de LDM offrent des valeurs réalistes. Les **procédures standards** sont, entre autres :

- L'anonymisation des noms par des prénoms et des noms de famille réalistes d'individus fondés sur des valeurs de référence interne ou externe. Les valeurs externes sont disponibles pour de nombreux pays ;
- L'anonymisation de l'adresse par des adresses qui répondent aux standards postaux, fondées sur des valeurs de référence interne ou externe. Les valeurs externes sont disponibles pour de nombreux pays.
- Des changements de date en paramétrant certains critères.
- L'anonymisation des chiffres particuliers tels que le numéro d'employé unique dans un cadre de valeurs prédéfini ;
- L'anonymisation des données bancaires sur la base de spécifications générales. Par exemple, les numéros de cartes bancaires comprenant les clés de vérification.
- La suppression des données
- Le remplacement de données par une valeur constante.

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	

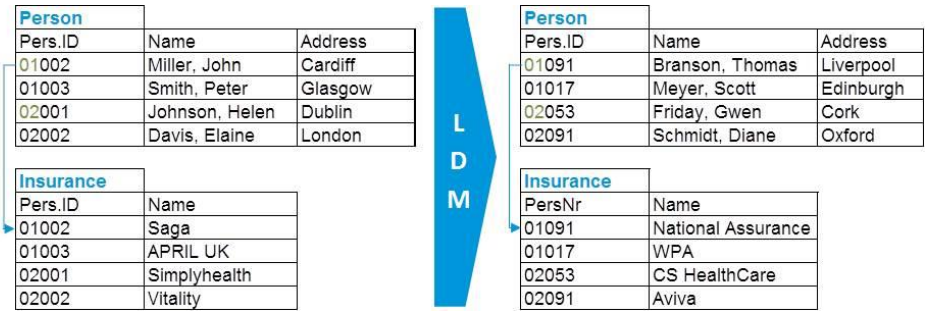
Méthode d'anonymisation

Exemple du changement de date

Dans sa version actuelle, Libelle **Data***Masking* offre 25 procédures d'anonymisation standard.

De plus, des procédures personnalisées peuvent être créées. On peut entre autres créer de nouvelles règles permettant :

- La création de numéros de contrat structurés dans le secteur de l'assurance et du trading par exemple ;
- Le calcul d'index dans un cadre de valeur de classes prédéfinies telles que les dommages dans le secteur de l'assurance et du trading ;
- Le calcul d'une valeur de compteur électrique dans le secteur de l'énergie
- La création d'autres procédures fournies par le client et par la communauté Libelle **Data***Masking*.



Cohérence des données

Exemple de conservation de la cohérence des données avec l'anonymisation simultanée de plusieurs occurrences d'une même donnée se trouvant dans plusieurs entrepôts de données séparés.

L'exécution de l'anonymisation : LDM « A-RUN »

L'objectif premier de Libelle **Data***Masking* est d'anonymiser des données productives se trouvant dans les environnements hors-production. Plusieurs mises en œuvre sont possibles. Les plus fréquentes sont les suivantes :

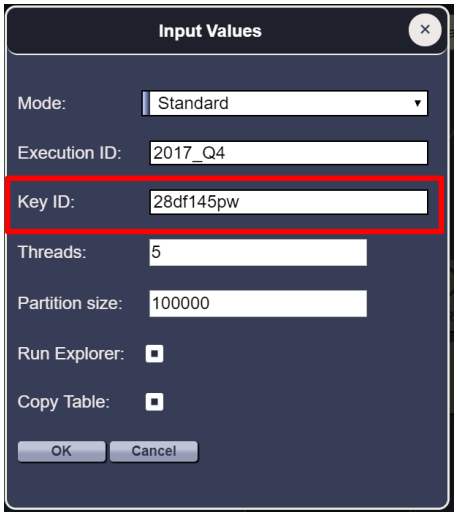
- L'anonymisation locale : l'anonymisation est réalisée localement depuis les entrepôts de données. Aucun système en amont ou en aval n'est impliqué.

- L'anonymisation « à la volée » : LDM Explorer et LDM Associator lisent les données depuis le système source. Les données originales sont anonymisées pendant le processus d'anonymisation (A-RUN) et les données anonymisées sont écrites sur le système cible.

- Le rafraîchissement de système puis l'anonymisation. Il s'agit d'un cas classique, notamment lors de copies homogènes de système SAP. L'anonymisation des données vient compléter le rafraîchissement de système.

Chaque exécution d'anonymisation est documentée. Cela permet de fournir un rapport sur l'état actuel des entrepôts de données vis-à-vis de leur anonymisation à n'importe quel moment.

Chaque exécution d'anonymisation est fonction d'une **clé d'anonymisation**. La clé est un champ alphanumérique que l'utilisateur est libre de déterminer (chiffres, lettres, caractères spéciaux). Si l'utilisateur utilise la même clé pour deux entrepôts de données identiques et pour une configuration identique, alors l'exécution de l'anonymisation mènera à un résultat identique. Dans ce sens, il est possible de mettre en œuvre des scénarios de test pour lesquels une entité quelconque vérifie la bonne application des procédures d'anonymisation individuelles.



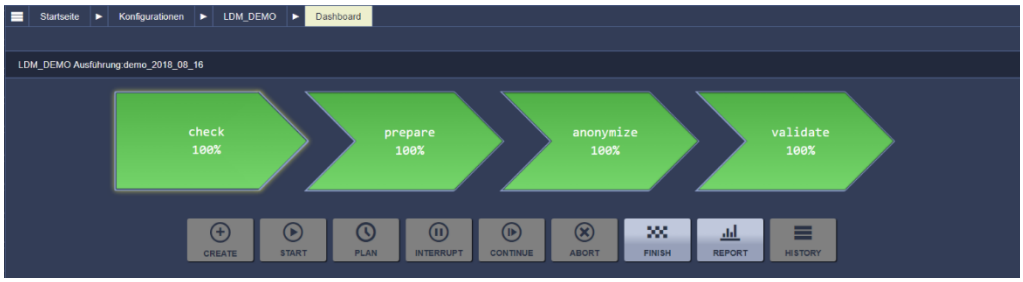
La clé d'anonymisation

Chaque anonymisation s'accompagne d'une clé d'anonymisation qu'il convient de renseigner.

Ainsi, afin d'obtenir une anonymisation sécurisée, il est recommandé d'entrer une clé aléatoire ou, a minima, de changer la clé à chaque exécution d'une anonymisation.

La configuration de Libelle Data*Masking* et les modèles de configuration : l'anonymisation sur simple pression d'un bouton

Une ou plusieurs configurations sont créées pour l'ensemble des bases de données, fichiers plats et profils d'anonymisation. Cela inclut l'ensemble des données techniques et des spécifications indiquées pour l'anonymisation. Une fois qu'une configuration a été définie et testée, elle peut être réutilisée à l'infini. La configuration est sauvegardée dans une « base de données de contrôle ». Ainsi, l'exécution de l'anonymisation peut être réalisée automatiquement et sans assistance – la nuit par exemple.



Appuyez sur le bouton

Après une configuration initiale, chaque anonymization est démarrée sur un simple clic.

Depuis l'interface de LDM, vous pouvez configurer, lancer et surveiller l'anonymisation.

Pour les environnements SAP tels que HCM, FI CO, etc., Libelle a développé des modèles de configuration standard. Ces derniers réduisent le temps de configuration du logiciel. Pour d'autres applications standard (ERP, CRM, Solution RH du marché), les utilisateurs peuvent créer des modèles de configuration individuels et les partager avec d'autres utilisateurs via Libelle.

Contactez-nous !

Libelle
37 rue de Neuilly
92110 Clichy
France

T: +33 1 41 05 39 07
france@fr.libelle.com

www.Libelle.com