



VOIE TECHNOLOGIQUE

Série STMG : Sciences et technologies du management et de la gestion

2^{DE}

1^{RE}

T^{LE}

*Management, sciences de gestion
et numérique*

ENSEIGNEMENT

SPECIALITE

SELOGER.COM ARRIVE À FAIRE RIMER RGPD AVEC AGILITÉ¹

Article du site LeMagIT.fr, dédié à l'information pour les décideurs IT2
Par Alain Clapaud, publié le : 18 mai 2020

Pour beaucoup d'acteurs du Web dont l'activité est de plus en plus liée à l'exploitation des données clients, RGPD rime avec contraintes techniques et lourdeur de gestion. Le Groupe SeLogger mise sur une gouvernance particulière pour transformer cette contrainte en atout.

L'entrée en application du RGPD en mai 2018 a coïncidé pour le groupe SeLogger avec la mise en place d'un département Data qui allait mettre en musique tous les usages de la donnée.

Le groupe SeLogger est une entreprise de 800 personnes (dont 250 développeurs), qui gère une quinzaine de marques, avec des sites web et des applications mobiles pour chacune d'elles, et des données acquises via des opérations de croissance externe.

« Nous agissons un peu comme un regroupement d'entreprises dont les activités sont différentes. Ce qui, en termes de système d'information et de mise en conformité avec le RGPD, a ajouté de la complexité du fait des spécificités de chaque Business Unit », explique Charles Sauthier, IT Manager Data du groupe.

De fait, il y a 20 ans, le groupe ne proposait que les sites Seloger.com et SelogerNeuf.com. Puis sont venus s'ajouter Belles Demeures, SeLogger Vacances, Amivac, Logic-Immo et quelques autres. Le chantier d'uniformisation du SI est désormais permanent. Mais cette complexité a aussi impacté très directement la mise en conformité RGPD – même si beaucoup des règles étaient déjà traitées dans le cadre de la loi « Informatique et Libertés ».

1. Article consulté le 09/12/2020 : <https://www.lemagit.fr/etude/Groupe-SeLogger-quand-RGPD-rime-avec-agilite>

2. IT ou technologie de l'information.

Le nouveau règlement européen a été pris très au sérieux et préparé très en amont par le Groupe. Les premières réflexions étant engagées deux années avant sa date d'entrée en vigueur.

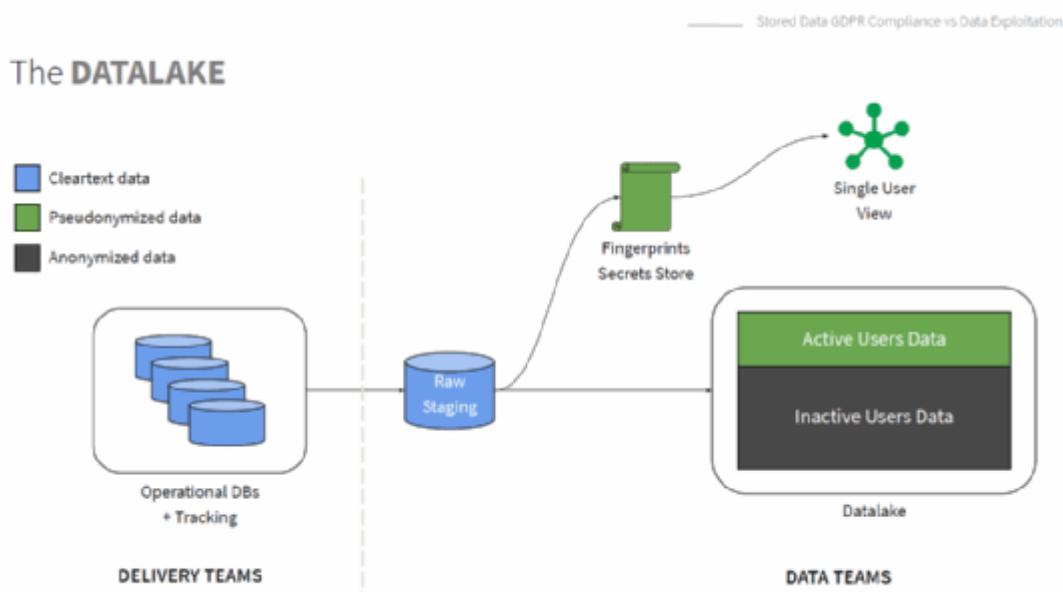
« Nous étions très heureux de la mise en place du RGPD, car cela nous a permis de nous appuyer sur des guidelines claires, savoir quelles données devaient être sécurisées et comment le faire. Jusqu'alors, chaque entreprise pouvait faire ses propres choix sur ce qui doit être sécurisé, ce qui peut être conservé ou pas. Avant le RGPD, un service pouvait choisir de conserver le plus de données possible alors que l'IT et la sécurité cherchaient à limiter au maximum le stockage. Avec le RGPD, tout le monde est au diapason, il n'y a plus de débats, tout le monde doit aller dans la même direction. »

Un Data Lake pour centraliser la donnée

Le groupe engage alors beaucoup de moyens. Pas uniquement pour assurer sa mise en conformité, mais aussi pour créer un département Data, copiloté par Charles Sauthier pour le volet technique et par Florent Guiocheau, CDO du groupe, en charge de l'aspect Business.

Alors que chaque site Web du groupe gérait ses propres données, priorité est alors donnée à une centralisation. Un Data Lake est créé en implémentant, dès sa conception, les règles du RGPD. Outre un chiffrement des données réalisé au niveau de l'infrastructure, le Data Lake est entièrement pseudonymisé par défaut, qu'il s'agisse de données internautes ou de données clients.

Charles Sauthier explique le principe de fonctionnement mis en place : « lorsque quelqu'un accède à la liste des inscrits pour une alerte immobilière, il y aura bien une ligne par inscrit, mais les noms, prénoms et emails seront chiffrés. La clé de chiffrement est en accès restreint et dès qu'un internaute est inactif depuis 3 ans, sa clé est détruite, donc automatiquement les milliers de lignes qui le concernent dans le Data Lake sont anonymisées, car nous ne sommes plus techniquement capables de retrouver le nom ».



Retrouvez éducol sur



Schéma de fonctionnement de base du Data Lake mis en place par le Groupe SeLoger, qui a été un élément clé de la mise en conformité RGPD en implémentant, dès sa conception, les processus de base du règlement européen.

Le processus est automatique : dès qu'une nouvelle table est créée dans le *Data Lake*, toutes les colonnes qui contiennent des données personnelles sont marquées comme telles. Dès lors, toutes les informations stockées dans ces colonnes sont automatiquement chiffrées au moment de l'écriture.

Les clefs sont, pour leur part, stockées dans un emplacement sanctuarisé du *Data Lake* dont l'accès est géré directement par le responsable de la sécurité.

Ce *Data Lake* a été conçu par les équipes du groupe SeLoger et s'appuie sur le format de stockage Apache Parquet, le tout étant stocké sur Amazon S3.

En parallèle, un grand nettoyage des bases de données de production a été lancé. Le RGPD indique que celles-ci ne doivent contenir que les données strictement nécessaires au fonctionnement des applications. Ainsi, les données des internautes récoltées via des formulaires ne sont conservées que pour les stricts besoins de fonctionnement des sites et des applications mobiles. L'historisation de ces données n'est faite que de manière pseudonymisée ou anonymisée au niveau du *Data Lake* lui-même.

Autre idée forte de la stratégie sécurité du groupe : les développeurs ne doivent par défaut pas avoir accès aux données opérationnelles (données directement produites par les sites et applications en production). S'ils ont besoin d'accéder à des données opérationnelles, ce ne peut être qu'à titre temporaire et sur un projet précis.

« Une telle approche n'est pas antinomique avec DevOps dès lors que l'on peut faire de l'automatisation », estime néanmoins Charles Sauthier. « Par principe, la gestion de la sécurité des données ne peut être disséminée entre les équipes de développement. Les développeurs travaillent sur les environnements de dev et de recette sur lesquels il n'y a pas de données personnelles ».

Passer de l'environnement de recette à la production est réalisé par un processus CI/CD. Si, exceptionnellement, un développeur doit accéder à la production, il peut obtenir des droits temporaires via un *repository Git* dédié à cette gestion des droits. Le développeur fait une *pull request* sur Git pour accéder à la base. Seul le service sécurité peut alors accepter sa demande si celle-ci est fondée.

Du RGPD à la gouvernance de données, un grand pas à franchir

Si la mise en conformité avec le RGPD a permis de remettre à plat les bases de données et les schémas de circulation de données, c'est à la mise en place d'une véritable gouvernance au niveau groupe que la direction Data s'est attelée.

« Il a 250 personnes à l'IT réparties dans une quarantaine de "Bulles" (équipes agiles) accompagnées d'une centaine de gens du Produit (Product Owners, UX/UI,...) qui travaillent sur l'ensemble de nos sites Web et applications mobiles. Il a donc fallu mettre en place une gouvernance de la donnée pour mettre tout le monde au diapason. C'est une vaste démarche qui est encore en cours ».

« Notre approche [de la gouvernance] n'est pas antinomique avec DevOps dès lors que l'on peut faire de l'automatisation. »
Charles Sauthier, Groupe SeLoger

Retrouvez éducol sur



Cataloguer les données personnelles pour le RGPD n'a pas été une tâche excessivement complexe, mais mettre en place cette gouvernance l'est visiblement beaucoup plus. Des *Data Owners* vont devoir définir les termes métiers et tenir à jour les catalogues de données. Des référents doivent être nommés dans les *Feature Teams* pour travailler avec le Data Stewart et maintenir le catalogue de données à jour. Un *Tech Data Owner*, un profil plus technique, aura également pour rôle de suivre les évolutions des structures des bases de données et avertir le *Data Stewart* central lorsqu'un changement est apporté.

En termes d'outillage, la mise en place du RGPD s'est essentiellement appuyée sur des outils de communication : un canal *Slack* dédié et une feuille *Google Sheet* partagée avec la liste de toutes les tables contenant des données personnelles.

Ces outils simples ont permis aux équipes de communiquer et échanger sur la gestion de ces données personnelles de manière efficace. « C'était pragmatique et ça a bien fonctionné pour le RGPD, mais il fallait rendre nos process plus robustes pour passer à l'échelle du groupe », ajoute l'expert.

C'est de la maison mère du groupe SeLogger – AVIV Group (elle-même filiale du groupe Axel Springer) – que va venir la solution.

Cette structure commune a fait le choix du logiciel de *Data Management* belge Collibra. Ce logiciel était déjà mis en œuvre au sein d'AVIV Group pour faire de la documentation de données sur les volets *Business Terms* et *Data Dictionary*, et plus globalement faire de la gouvernance de la donnée.

Le groupe SeLogger a commencé à mettre en œuvre la solution en tant que référentiel (*repository*) pour ses dictionnaires de données et à déployer, petit à petit, ses fonctionnalités progressivement notamment sur le volet processus. Chaque filiale du groupe dispose de son compte administrateur et peut configurer l'outil à sa guise.

Charles Sauthier qui avait eu une expérience malheureuse avec une plateforme de *Data Management* trop rigide et à l'interface désuète a été séduit par la plateforme choisie par AVIV Group : « Collibra se montre modulaire. [...]. D'autre part, le logiciel est entièrement pilotable par API. Nous pouvons donc automatiser ce que l'on souhaite ».

Le confinement n'a pas entravé ce déploiement, car l'activité du groupe SeLogger n'a été ralentie que partiellement par la pandémie, du fait de la fermeture temporaire des agences immobilières et de la forte réduction du trafic. « D'une certaine manière c'est presque une opportunité, car il n'y a plus toutes les interruptions du quotidien liées aux urgences, aux demandes clients et nous pouvons nous concentrer sur les sujets de fond », constate même Charles Sauthier.

Le département Data vient en tout cas de nommer ses premiers Data Stewards et va pouvoir engager la montée en puissance de sa structure de gouvernance dans les prochains mois.