RGPD

Réglementation Générale sur la Protection des Données

GDPR - General Data Protection Regulation

Sources : Bozhidar Bozhanov, développeur senior et conseiller du vice ex-premier ministre de Bulgarie

Introduction

- Loi entrée en vigueur le 25 mai 2018
- Doit être suivie dans tous les pays européens et qui s'applique aux entreprises hors de l'Europe qui ont des utilisateurs en Europe).
- Développeurs ont un rôle clé dans les recommandations qu'ils doivent faire lors de l'établissement d'un projet ou d'une fonctionnalité.
- L'utilisateur / le client est mentionné en tant que « sujet de données » dans la réglementation.

Droits des utilisateurs

- Droit à l’effacement.
  - le droit d'être oublié/supprimé du système.
- Droit à la restriction du traitement.
  - vous gardez toujours les données, mais les marquez comme « limitées » et ne les touchez pas sans un consentement supplémentaire de l'utilisateur.
- Droit à la portabilité des données.
  - la possibilité de les exporter dans un format lisible informatiquement.
- Droit à la rectification.
  - la possibilité d'avoir ses données personnelles corrigées.
- Droit d’être informé.
  - avoir des informations humainement lisibles, plutôt que de longues conditions.
- Droit d’accès.
  - l'utilisateur devrait pouvoir voir toutes les données que vous avez sur lui.

Principes de base

- minimisation des données
    - ne pas collecter plus de données que nécessaire, savoir justifier ces choix
- intégrité et confidentialité
    - toutes les mesures de sécurité imaginables pour protéger les données et les mesures pour garantir que ces données ne seront pas modifiées de manière inappropriée

La réglementation requiert la mise en place de certains processus à l'intérieur des structures traitant des données personnelles, et ceci inclut de garder un enregistrement de **tous** les traitements réalisés, incluant les transferts à d'autres intervenants tiers, ce qui inclut les fournisseurs de services cloud.

Aucune exception n'est prévue en fonction de la taille de la structure qui traite les données.

« Je suis petit, la RGPD ne me concerne pas » est un mythe.

Définition

Donnée personnelle :

Une donnée pouvant être utilisée pour identifier une personne de façon unique ou d'une donnée étant déjà une personne identifiée.

Ce sont les données que l'utilisateur a explicitement fournies, mais aussi les données qui sont collectées sur lui par le biais d'un tiers, ou basées sur son activité sur un site (ce qu'il a consulté, ce qu’il a acheté, etc.).

Fonctionnalités à implémenter

- Oubliez-moi
- Notification des suppressions aux tiers
- Restriction de traitement
- Exportation des données
- Edition du profil
- Consentement par cases
- Redemande de consentement
- Voir toutes mes données
- Conservation temporaire
- Cookies

Oubliez-moi !
(Art. 17 GDPR)

Avoir une procédure qui prend en paramètre un identifiant utilisateur et efface toutes ses données personnelles.

- Fonctionnalités intéressantes pour les tests d'intégration (nettoyage après test), mais difficile à implémenter selon le modèle de données. Dans un modèle de données ordinaire, supprimer un enregistrement peut être facile, mais certaines clés étrangères peuvent être enfreintes :
    - Autoriser les clés étrangères pouvant être à NULL.
    - S’assurer de la destruction de toutes les données relatives (en cascade)

Vous devez globalement constamment penser à comment supprimer les données personnelles !

Et à propos des sauvegardes ?

Idéalement, vous devez garder une table séparée pour les identifiants des utilisateurs oubliés, donc à chaque restauration d'une sauvegarde, vous réoubliez les utilisateurs oubliés.

Cela signifie que la table devrait être dans une base de données séparée ou avoir un processus de sauvegarde/restauration séparé.

Notification des suppressions aux tiers (Art. 19 GDPR)

Lors de la suppression des données d’une personne, il faut informer tous les tiers à qui les données ont été poussées que celles-ci sont supprimées.

Donc, si vous avez envoyé des données personnelles à des services externes tels que Salesforce, Hubspot, Twitter, ou tout autre fournisseur de service cloud, vous devrez appeler une API de leur cru permettant la suppression de données personnelles.

Appeler l'API tierce pour supprimer des données n'est pas tout.

Vous devez également vous assurer que l'information n'apparaisse pas dans des résultats de recherche.

Maintenant, c'est délicat, car Google n'a pas d'API de suppression, seulement un **processus manuel**.

Heureusement, c'est seulement pour les pages de profils publics navigables et accessibles par Google (et d'autres moteurs de recherches, bien entendu), mais vous avez toujours à prendre des mesures.

Restreindre le traitement (Art. 18 GDPR)

Dans votre panneau d'administration, où il y a une liste d'utilisateurs, il devrait y avoir un bouton « restreindre le traitement ».

La page réglages utilisateur devrait également l'avoir.

Lors de son clic (après lecture de l'information appropriée), il devrait marquer le profil comme restreint.

Cela signifie qu'il ne devrait plus être visible par l'équipe du backoffice, ou publiquement.

Export des données (Art. 20 GDPR)

Dans le profil d'un utilisateur, un bouton d'export des données, doit permettre le téléchargement de toutes les données possédées sur lui.

Généralement, cela concerne toutes les données qui font partie de la fonctionnalité « oubliez-moi ».

Quelquefois l'export peut prendre du temps, le bouton peut donc déclencher un processus en tâche de fond pour être envoyé par email quand les données sont prêtes.

Permettre aux utilisateurs d'éditer leur profil (Art. 16 GDPR)

Cela semble une règle évidente, mais elle n'est pas toujours suivie.

Les utilisateurs doivent pouvoir modifier toutes les données les concernant, incluant les données que vous avez collectées depuis d'autres sources (par exemple en utilisant un login avec Facebook vous devriez avoir récupéré leur nom et adresse).

Consentement par cases à cocher (Art. 7 GDPR)

« J'accepte les termes et conditions » n’est plus suffisant pour prétendre que l'utilisateur a donné son consentement au traitement de ses données.

Donc, pour chaque activité particulière de traitement, il devrait y avoir une case à cocher séparée sur l'écran d'enregistrement (ou le profil utilisateur).

Vous devrez garder ces cases à cocher dans des colonnes séparées de la base de données, et laisser l'utilisateur retirer son consentement (en décochant ces cases depuis son profil).

Idéalement, ces cases à cocher devraient provenir directement de l'activité de traitement de l'enregistrement (si vous en gardez un).

Notez que ces cases à cocher ne devront pas être présélectionnées, ceci ne comptant pas comme un « consentement ».

Concernant le *machine learning :*

Si vous allez utiliser les données des utilisateurs pour entraîner vos modèles ML, vous devrez également obtenir le consentement pour cela (sauf à des fins scientifiques, lesquelles ont un traitement spécifique).

Notez ici le traitement appelé « intérêt légitime ». C'est à l'équipe juridique de décider ce qui est d'un intérêt légitime, mais le marketing direct est inclus dans cette catégorie, comme tout traitement de bon sens relatif à l'activité commerciale. Si par exemple vous collectez des adresses d'expédition, c'est évidemment légitime. Toutes les activités de traitement ne nécessitent pas de cases à cocher de consentement.

Redemande de consentement (Art. 7 GDPR)

Si le consentement qu'ont donné les utilisateurs n'était pas clair (exemple : simple accord aux termes et conditions), vous devrez le réobtenir.

Préférez donc une fonctionnalité de *mass-mailing* pour demander aux utilisateurs d'aller dans leur page de profil et vérifier toutes les cases à cocher pour les traitements sur leurs données personnelles.

Voir toutes mes données (Art. 15 GDPR)

- Ceci est similaire au bouton « export », mis à part que les données devraient être visibles dans une interface graphique régulière.
- Je ne dirai pas que c'est obligatoire, et vous pouvez le laisser comme une fonctionnalité « souhaitable ». Par exemple, Google Map vous montre votre historique de position, tous les endroits où vous avez été. C'est une bonne implémentation du droit d'accès (bien que Google soit **loin d'être parfait** quand la vie privée est concernée).
- Ce n'est pas tout à propos du droit d'accès. Vous devez laisser les utilisateurs non enregistrés savoir si vous avez des données les concernant, mais ce serait plus un processus manuel. Le minimum idéal serait d'avoir une fonctionnalité « vérification par mail », ou vous contrôleriez si vous avez des données concernant une adresse mail particulière. Vous devez dire à l'utilisateur de quelle façon vous traitez ses données. Vous pouvez simplement imprimer tous les enregistrements dans votre registre de traitement des données pour lequel l'utilisateur a consenti.

Contrôle de l'âge (Art. 8 GDPR)

- Demander l'âge des utilisateurs, et si l'utilisateur est un enfant (moins de seize ans), demander la permission des parents.

- Évidemment, les enfants pourront tricher sur leur date de naissance, ou fournir une fausse adresse mail parentale, mais vous aurez fait votre travail conformément à la législation (c'est un des vœux pieux de la réglementation).

Ne pas garder les données plus que nécessaire (Art. 5 GDPR)

Si vous avez collecté les données pour un usage spécifique (ex. livrer un produit), vous devez les supprimer/les anonymiser aussitôt que possible.

Beaucoup de sites d'e-commerce offrent une option « acheter sans inscription », dans quel cas, le consentement n'est valable que pour la commande particulière.

Vous devrez donc avoir une tâche programmée pour anonymiser périodiquement les données (supprimer les noms et adresses), mais seulement après la réunion de certaines conditions, exemple : confirmation de livraison du produit.

Les cookies (directive 2009/136/CE)

Le principe :

- d'un consentement préalable de l'utilisateur avant le stockage d'informations sur l'équipement d'un utilisateur ou l'accès à des informations déjà stockées.
- sauf, si ces actions sont strictement nécessaires pour la délivrance d'un service de la société de l'information expressément demandé par l'abonné ou l'utilisateur.
- L'article 32-II de la loi du 6 janvier 1978, modifié par l'ordonnance n°2011-1012 du 24 août 2011 qui a transposé la directive 2009/136/CE reprend ce principe.
- En application de la loi informatique et libertés, les traceurs (cookies ou autres) nécessitant un recueil du consentement ne peuvent donc être déposés ou lus sur son terminal, tant que la personne n’a pas donné son consentement.

Les choses à faire

Voici maintenant une liste de choses à faire.

La plupart concernent les mesures techniques nécessaires pour protéger les données personnelles (décrites dans l'**article 32**).

Il s'agit plus d'infrastructure que du développement, mais souvent l'application devra être étendue pour les supporter.

Ce n'est pas exigé par la réglementation, mais c'est en tout cas des bonnes pratiques et aide à protéger les données personnelles.

Chiffrage des données en transit

- Cela signifie que la communication entre votre couche applicative et votre base de données (ou votre queue de messages, ou tout composant que vous avez) devrait se faire à travers **TLS / SSL**.
- Les certificats peuvent être autosignés (et éventuellement épinglés), ou vous pourriez avoir une autorité de certification interne.
- Différentes bases de données ont des configurations différentes. Certaines bases de données ont besoin de dialoguer entre plusieurs nœuds, qui devraient aussi être configurés pour utiliser du chiffrement.

Chiffrer les données au repos

- Cela dépend encore de la base de données (certaines offrent un chiffrement au niveau des tables), mais peut aussi être au niveau machine.
- Ex. en utilisant **LUKS**. La clé privée peut être stockée dans votre infrastructure, ou dans un service cloud comme AWS KMS.

Chiffrer vos sauvegardes

- …

Implémentation de la pseudonymisation

- Le cas d'utilisation le plus évident est quand vous voulez utiliser des données en production pour des serveurs de tests/mises en scène. Vous devrez changer les données personnelles en « pseudonymes », de façon à ce que les gens ne puissent pas être identifiés. Quand vous poussez des données à des fins d'apprentissage automatique (à des tiers ou non), vous pouvez aussi le faire.
- Techniquement, cela pourrait signifier que votre objet utilisateur peut avoir une méthode « pseudonymise » qui applique un hash+sel/**bcrypt**/**PBKDF2** pour les données pouvant être utilisées pour identifier une personne. La pseudonymisation peut être réversible ou non, dépendant de l'usage (la définition dans la régulation implique la réversibilité basée sur une information secrète, mais pas en cas de données de test/mises en scène).
- Certaines bases de données ont de telles fonctionnalités intégrées, **ex. : Oracle**.

Protéger l’intégrité des données

- C'est un point très large, et peut simplement signifier « avoir un système d'authentification pour modifier les données ».
- Mais vous pouvez faire plus, même aussi simple qu'un checksum, ou une solution plus compliquée.
- Cela dépend des enjeux, de la façon dont les données sont accédées, d'un système particulier, etc.
- Le checksum peut être de la forme d'un hash de toutes les données d'un enregistrement de la base, qui devra être mis à jour à chaque fois que l'enregistrement est modifié via l'application. Ce n'est pas une garantie forte, mais c'est au moins quelque chose.

Avoir votre registre RGPD des activités de traitement dans autre chose qu'Excel

- L'**article 30** stipule que vous devriez garder un enregistrement de tous les types d'activités où vous utilisez des données personnelles. Cela sonne comme de la bureaucratie, mais peut être utile. Vous serez capable de lier certains aspects de votre application avec ce registre (ex. les cases à cocher de consentement, ou votre trace d'enregistrement d'audit).
- Mettre en place un simple registre ne devrait pas prendre beaucoup de temps, mais les besoins de l'entreprise pour cela viennent de quiconque est responsable de la conformité RGPD. Mais vous pouvez aussi les informer que l'avoir dans un fichier Excel ne le rendra pas facile pour vous en tant que développeur (imaginez avoir à fouiller dans le fichier Excel, pour pouvoir le parser et implémenter une fonctionnalité). Un tel registre pourrait être un microservice/une petite application déployée séparément dans votre infrastructure.

Loguer les accès aux données personnelles

- Chaque opération de lecture d'un enregistrement de données personnelles devrait être loguée, pour que vous sachiez qui y a accédé et pour quel usage. Cela ne découle pas directement de la réglementation, mais c'est implicite concernant les responsabilités.
- Qu'en est-il des résultats de recherche (ou listes) contenant des données personnelles sur plusieurs sujets ? Simplement loguer « utilisateur X a fait une recherche sur le critère Y » devrait suffire. Mais n'affichez pas trop de données personnelles dans les listes, par exemple voir comment Facebook vous fait passer par des cercles pour obtenir l'anniversaire d'une personne.
- Il y a d'autres articles sur le règlement qui laissent entendre que garder un log d'audit est une meilleure pratique (pour protéger l'intégrité des données aussi bien que s'assurer qu'elles n'ont pas été accédées sans une raison valable).

Répertorier toutes les API clients

- Vous ne devriez pas autoriser les API anonymes à accéder aux données personnelles. Il faut connaitre le nom de l'entreprise et le nom du contact pour chaque utilisateur de l'API lors de l'inscription, et les ajouter au registre de traitement des données.

Les choses à ne pas faire

N'utilisez pas les données pour d'autres buts que ceux acceptés par l'utilisateur

- Ceci est supposé être l'esprit de la réglementation.
- Si vous voulez exposer une nouvelle API à un nouveau type de clients, ou voulez utiliser les données pour de l'apprentissage automatique, ou décidez d'ajouter des publicités à votre site basé sur le comportement des utilisateurs, ou encore vendre votre base de données à un tiers, ATTENTION ! Lors de l'ajout d'une nouvelle activité de traitement (et l'ajout à votre registre), il faut réaliser un envoi d'un mail en masse à tous les utilisateurs desquels vous souhaiteriez le consentement. Notez ici que cette autorisation peut être ajoutée dynamiquement.
- **Article 6 de la RGPD**.

Ne pas loguer les données personnelles

- Se débarrasser des données personnelles depuis des fichiers de log (spécialement si elles sont fournies par un service tiers) peut être fastidieux ou même impossible.
- Loguez donc juste les identifiants si nécessaire, et assurez-vous que les anciens fichiers de log sont nettoyés, juste au cas où.

Ne mettez pas de champs sur l'enregistrement/le formulaire de profil dont vous n'avez pas besoin

- Il est toujours tentant de placer autant de champs que le designer le souhaite, mais sauf si vous avez absolument besoin des données pour fournir votre service, vous ne devriez pas les collecter.
- Vous devrez probablement toujours collecter les noms, mais à moins de livrer quelque chose, une adresse personnelle ou un numéro de téléphone est inutile.

Ne supposez pas que les tiers sont conformes

- Vous êtes responsable s'il y a une fuite de données chez un tiers (c’est-à-dire le tiers effectuant le traitement) à qui vous avez envoyé des données personnelles.
- Donc avant d'envoyer des données à un autre service via une API, assurez-vous d'au moins un niveau de protection basique. Si ce n'est pas le cas, signalez-le à la direction.

Ne supposez pas que d'avoir une norme ISO vous rend conforme

- Les informations standards de sécurité et même les standards de données personnelles sont un bon point de départ et représenteront probablement 70 % des éléments requis par la réglementation, mais ne sont pas suffisants. La plupart des choses listées ci-dessus ne sont couvertes par aucun de ces standards.

Conclusion

- Globalement, le but de la réglementation est de vous faire prendre des décisions conscientes lors de traitements de données personnelles. Elle impose les meilleures pratiques de manière légale. Si vous suivez les conseils présentés ci-dessus et concevez vos modèles de données, vos stockages, vos flux de données, et vos appels d'API avec la protection des données à l'esprit, vous ne devriez alors pas vous inquiéter des amendes énormes préconisées, elles sont pour les cas extrêmes, comme **Equifax** par exemple. Les régulateurs (les autorités de protection de données, comme la CNIL en France) auront très probablement des listes de contrôle auxquelles vous devrez d'une manière ou d'une autre répondre, mais si vous suivez les meilleures pratiques, cela ne devrait pas être un problème.
- Les fonctionnalités ci-dessus peuvent être implémentées en quelques semaines par une petite équipe. Soyez méfiant quand un gros vendeur vous offre une solution générique et *plug and play* de « conformité RGPD ». RGPD ne concerne pas uniquement les aspects techniques évoqués, il a des implications de traitements et d'organisation. Mais soyez aussi méfiant si un consultant prétend que RGPD est compliquée. Ça ne l'est pas. Ça repose sur quelques principes de base qui sont en tout cas de bonnes pratiques. Donc, ne les ignorez pas.