Compte-Rendu de la réunion GFII-AFCDP du 16 décembre 2010 sur la Réutilisation des données publiques (Thème principal : Anonymisation)

De Bruno RASLE 06 1234 0884 charge-mission@afcdp.net et Ruth MARTINEZ 06 08 83 25 01 gfii@wanadoo.fr
Diffusion restreinte AFCDP et GFII Version 1.1 le 31 janvier 2011

Eléments de contexte

Le groupe de travail commun GFII (Groupement Français de l'Industrie de l'Information) et AFCDP (Association Française des Correspondants à la protection des Données à caractère Personnel) sur la réutilisation de données publiques contenant des données personnelles a tenu une deuxième réunion, le 16 décembre 2010.

Cette réunion s'est tenue dans le superbe salon Doré de l'hôtel de Mailly, mis à disposition par la DILA (Direction de l'Information Légale et Administrative - Services du Premier ministre), que nous remercions vivement.

Le thème principal était l'anonymisation des données personnelles. En préambule le groupe est revenu sur les points soulevés lors de la première réunion et a abordé deux points d'actualité (proposition contenue dans le projet LOPPSI 2 d'obliger les Producteurs de données publiques à diligenter une enquête administrative de « moralité » sur les Réutilisateurs, débats autour du traitement SIV et de l'utilisation à des fins de prospection commerciale des données à caractère personnel des propriétaires d'automobiles).

Les entités suivantes étaient représentées : Région Ile de France, la Chambre de Commerce de Paris, le Conseil Général de Seine-Maritime, la Ville de Vitry sur Seine, l'INTD, Canope, la DILA, Ancestry, NotreFamille.com, Reed Elsevier, Cabinet Gilles Vercken, GFII, AFCDP.

Etaient excusés: Conseil d'Etat, Agence d'Aide aux Collectivités Locales des Landes, Ucanss.

Retour sur la première réunion

Bruno Rasle fait part de la réponse reçue de l'association partenaire GDD (qui représente les « CIL » d'outre-Rhin) : pour le moment il ne semble pas que le sujet « réutilisation des données publiques » soit d'actualité. L'AFCDP attend des réponses de Suède et Hollande.

Le document « Les données publiques, guide juridique & pratique » publié par l'AEC (Agence des initiatives numériques) en décembre 2010 a été commenté. Plusieurs des questions soulevées par le groupe lors de sa première réunion sont abordés :

La démarche est-elle identique en présence d'une demande de transfert de données publiques émanant d'une autre collectivité et **une demande d'une société étrangère** (hors UE) n'ayant aucun établissement en Europe ou en France, d'une secte[1], d'un parti politique extrémiste ? → « Les droits d'accès et de réutilisation sont universels. Les données publiques détenues par les acteurs publics français peuvent être réutilisés par toute personne qui le souhaite, sans condition de nationalité » (p.16)

La loi exclut de la réutilisation les données publiques contenant des données à caractère personnel. Toutefois, les organismes publics peuvent décider de les mettre à disposition. Pour les données publiques contenant des données à caractère personnel, se posent alors toutes les questions que l'on retrouve dans les secteurs Vente-Marketing et Santé[2] : Quelle démarche opérationnelle adopter pour que **le consentement** soit indiscutable ? Quel niveau d'information délivrer ? Information vaut-elle consentement ou bien s'agit-il d'un consentement express ? Le consentement évoqué répond-t-il aux mêmes caractéristiques de celui défini par la Loi pour l'Economie Numérique (consentement libre, explicite, informé) ? Faut-il mentionner les catégories de destinataire – voir même citer les Réutilisateurs ? → « L'autorisation (consentement) des personnes concernées, explicite et signifiée par écrit, doit être préalable à la réutilisation d'un document qui n'a pas été rendu anonyme. Cette formalité permet de conserver une trace écrite du consentement, qui servira de preuve en cas de contestation » (p.13). L'auteur du document sera contacté, pour avoir connaissance du texte de référence qui fait état de ce consentement « explicite et signifié par écrit ».

Interconnexions. Une entité (laquelle ? CNIL ?) serait-elle en mesure d'être informée des réutilisations, afin de dresser un panorama (Observatoire) et de détecter les cas problématiques où des fichiers sont interconnectés, pouvant générer par leur croisement, des données identifiantes, donc à caractère personnel ? → « Au sein de son établissement, les missions de la PRADA…établir un bilan annuel des demandes d'accès et de réutilisation qui sera présenté à la CADA » (p.9)

Suite à la réunion le groupe a obtenu communication de la délibération n°2010-460 du 9 décembre 2010 « portant recommandations relative aux conditions de réutilisation des données à caractère personnel contenues dans des documents d'archives ». Ce texte, bien que focalisé sur les archives publiques, donne quelques précieuses indications sur d'autres points de questionnement soulevés lors de la réunion du 15 octobre 2010, tels que la durée de vie des personnes, la démarche par analyse de risques, le formalisme liée aux réutilisations de données publiques contenant des données à caractère personnel, etc. Le groupe y reviendra lors d'une prochaine réunion.

---

1 Plusieurs participants ont exprimé une inquiétude sur des risques de dérapages concernant certains fichiers, crainte concernant notamment la réutilisation par des sectes de certaines données ou informations.
2 Ces sujets sont traités par les groupes AFCDP « Données Clients et Prospects » et « Données de santé ».

Débats d'actualité

Ruth Martinez, Déléguée Générale du GFII, a commenté le point principal d'actualité, à savoir l'apparition d'un amendement au projet LOPPSI 2 (en cours de discussion) qui vise à subordonner les réutilisations à **une enquête administrative de moralité** sur les Réutilisateurs sous licence, amendement qui a suscité un vif émoi chez les acteurs concernés car il est en contradiction avec le contexte d'ouverture.

Cet amendement visait à encadrer la revente des données personnelles des « cartes grises » (traitement SIV). Le GFII a alerté les pouvoirs publics sur les effets induits.

Les participants ont rapproché ce débat sur une éventuelle enquête administrative sur le Réutilisateur de l'un des thèmes discuté lors de la première réunion (« *Questions relatives à la démarche poursuivie par le Réutilisateur et la « qualité » de celui-ci, aux risques et aux impacts pour les personnes concernées* », cf. compte-rendu de la réunion du 15 octobre 2010). Parmi les bonnes pratiques envisagées à ce stade figure « Réaliser une analyse de risques (au sens Informatique et Libertés) pour chaque projet de transfert - sous l'égide du CIL) ».

Ont également été évoquées :

- L'émoi de l'AAF (Association des Archivistes de France) sur les possibilités d'interconnexions de fichiers (possibilité de créer des mega bases de données, issues du rapprochement de plusieurs gisements de données obtenues dans le cadre de réutilisation de données publiques). On rappelle que l'interconnexion de fichiers de données personnelles **est soumise à autorisation préalable** auprès de la CNIL.
- Les caractéristiques du recueil de consentement dans les formulaires SIV (la personne doit cocher une case **pour s'opposer** à ce que ses données personnelles soient utilisées à des fins commerciales !) alors qu'il est souvent demandé de tendre vers un consentement actif (par défaut, si la personne concernée ne prend pas action, il n'y a pas consentement).
- Plusieurs points fondamentaux liés à la loi Informatique et Libertés : difficulté à définir clairement ce qu'est une donnée à caractère personnel (distinction d'une donnée nominative), identification du Responsable de traitement, définition d'une interconnexion, etc. (sur ces points, se reporter aux travaux des groupes AFCDP *ad hoc*), durée de conservation (une publication de la CNIL est attendue et l'AFCDP travaille sur ce sujet en partenariat avec l'AAF).
- Les archives départementales gèrent des documents qu'elles n'ont pas produits ou dont elles n'ont pas assuré la collecte (collecte par les mairies, l'INSEE…). Le Code du patrimoine définit le délai de reversement des archives des services déconcentrés de l'État, des établissements publics, dont le siège est situé dans le département (préfectures, rectorats, universités, agences de l'eau...), les archives notariales des notaires du département.
- L'anonymisation des délibérations des collectivités : les avis de la CADA et de la CNIL divergent.
- La confusion entre données nominatives et données à caractère personnel persiste : le trajet du domicile au lieu de travail est une donnée personnelle. Le nom n'est pas forcément une donnée personnelle.

Les participants sont également revenus sur la question du recueil du consentement: **que faire des données personnelles collectées sous l'empire de l'ancien texte ?** Il semble prohibitif de contacter toutes les personnes concernées.

L'anonymisation

Avant de procéder à l'audition de l'expert, les participants ont échangé sur ce thème.

Les nombreuses difficultés liées à cette approche (cf. Compte-rendu de la réunion du 15 octobre 2010) ont été à nouveau soulignées : sur qui porte la charge de l'anonymisation ?

Un participant est revenu sur **la question du vocabulaire** (« occultation », « disjonction », etc.). Sur ce point, les travaux de l'AFCDP ont été signalés (un document « Glossaire des termes utilisés dans le cadre d'Anonymisation de données personnelles » est accessible sur le site Web de l'association[3]).

Un participant a évoqué le risque, après anonymisation suivant une technique qui remplace un patronyme par un autre, de créer une situation à risque pour la personne concernée par le patronyme généré, avec un risque de préjudice (ex : si le nom attribué lors du procédé d'anonymisation correspond à un terroriste recherché par toutes les polices de la planète). Doit-on signaler que le jeu de données est anonymisé et non réel ?

Un participant a levé le point suivant : **anonymisation vaut souvent perte de sens** (il est très difficile d'obtenir une information qui soit à la fois très signifiante et totalement anonymisée – c'est souvent un compromis entre ces deux critères qui est obtenu). Dans certains cas et dans un cadre à définir strictement afin d'apporter toutes les garanties de confidentialité et de respect du droit des personnes, n'y aurait-il pas moyen d'autoriser le Réutilisateur à effectuer lui-même l'anonymisation, afin de conserver le maximum de pertinence et de sens aux documents ainsi traités ?

Le sujet de l'anonymisation des CV a été rapidement évoquée : Qui réalise l'anonymisation, et suivant quelles consignes ?

[3] *Disponible sur la page http://www.afcdp.net/-Referentiels-et-Labels-*

Audition d'un expert

Nous avons ensuite bénéficié de l'audition d'un représentant d'un grand Ministère qui nous a fait part de sa grande expérience opérationnelle et organisationnelle dans le domaine très particulier de l'anonymisation de données à caractère personnel.

Ses services ont notamment réalisé des traitements d'anonymisation portant sur des fichiers comportant plusieurs millions d'entrées.

Cet expert vise à systématiser l'anonymisation, par exemple pour tenir à disposition des développeurs des jeux de test (ces informaticiens ne devraient pas, en effet, travailler des données personnelles réelles).

Il confirme que chaque cas est spécifique. L'anonymisation d'une base de données peut prendre de quelques semaines à plusieurs mois, en fonction des contraintes (maîtrise-t-on, par exemple, totalement le schéma de la base de données et l'applicatif ? Présence de commentaires pouvant permettre d'identifier les personnes ?) et des objectifs visés (gestion des homonymies, respect de la logique de l'application, cohérence entre les données après anonymisation, chaînabilité, etc.). Ceci rend délicat, pour les Producteurs, la prévision du montant qu'ils doivent intégrer au titre des frais d'anonymisation au sein des licences et requiert une véritable expertise.

Les étapes de l'anonymisation ont été détaillées :

- Spécification de la demande
- Analyse des données et des règles de gestion (importance de la maîtrise du schéma des données)
- Choix de la technique d'anonymisation la plus pertinente
- Approbation par les utilisateurs
- Clonage de la base
- Réalisation de l'anonymisation
- Contrôle de l'anonymisation (comparaison source et clone)
- Livraison des données anonymisées.

Les palettes de techniques utilisées :

- Suppression,
- Masquage (substitution, mélange, chiffrement, hachage)
- Génération de données.

Il est signalé que l'anonymisation n'est pas seulement utilisée pour être conforme à la loi Informatique et Libertés : plusieurs sociétés y ont recours également dans une démarche de type Intelligence économique, pour protéger leurs actifs immatériels.

Les stratégies d'anonymisation doivent répondre à plusieurs critères : Chaînage (si demandé – il s'agit de pouvoir suivre les actions d'une personne dans le temps et/ou dans l'espace, sans jamais pouvoir l'identifier – faculté souvent souhaitée dans le domaine de la santé pour suivre un parcours de santé, ou dans le commerce pour analyser des habitudes de consommation), réversibilité (existe-t-il une possibilité pour le maître des données de procéder à un processus inverse ?) et la robustesse (la capacité, pour le corpus issu du procédé d'anonymisation, de résister à des tentatives frauduleuses de « ré-identification »).

L'expert a cité quelques rares outils du marché, qui nécessitent cependant une solide expérience pour aboutir à des résultats jugés conformes.

Il a été rappelé que les enregistrements audio et vidéo peuvent demander également à être anonymisés. C'est notamment le cas des enregistrements de vidéoprotection, dont les visages des tiers doivent être floutés avant de donner suite à une demande de droits d'accès d'une personne à ses données personnelles, au titre de la loi Informatique et Libertés.

Poursuite des travaux

Il est envisagé l'audition de représentants de la CADA et/ou de la CNIL. Cette audition serait l'occasion d'analyser la délibération de la CNIL sur la réutilisation des archives publiques contenant des données personnelles.

Nous rappelons que la participation suivie à ce groupe de travail commun au GFII et à l'AFCDP nécessite la qualité de Membres de l'une de ces deux entités.

Cependant le groupe est ravi d'accueillir :

- pour audition des représentants de Producteurs de données publiques (Collectivités territoriales, régies de transports, etc.) et de Réutilisateurs ;
- pour participation des représentants d'entités qui envisagent de rejoindre l'AFCDP ou le GFII.

Nous rappelons a) les règles de confidentialité qui s'appliquent à ces échanges, b) que les propos tenus en séance et consignés dans les comptes-rendus ne constituent en aucune manière des positions du GFII et de l'AFCDP, dans l'attente d'un éventuel positionnement des Conseils d'administration de ces associations.

Vous souhaitez participer à ces travaux ? Nul besoin d'être spécialiste et expérimenté ! Vos questionnements et cas pratiques vont nourrir les débats et les réflexions.

Merci de prendre contact avec : Bruno RASLE, Délégué Général de l'AFCDP - 06 1234 0884 charge-mission@afcdp.net et Ruth MARTINEZ, Déléguée générale du GFII – 06 08 83 25 01 gfii@gfii.asso.fr