N° 393714
Société JCDecaux France

9ème et 10ème chambres réunies
Séance du 13 janvier 2017
Lecture du 8 février 2017

CONCLUSIONS

Mme Aurélie BRETONNEAU, rapporteur public

L'architecture du quartier de la Défense lui donne déjà des petits airs de Gattaca ; équiper l'ensemble des équipements publicitaires qui en jalonnent la grande dalle de dispositifs de mesure d'audience ou d'analyse des comportements y ajouterait une touche *Minority report*, certaines de ces technologies n'ayant rien à envier aux scanners rétiniens permettant, dans le film de Steven Spielberg, d'adresser des publicités personnalisées aux passants qui y jettent un œil.

Ce n'est pas le plus intrusif de ces dispositifs répertoriés que la société JC Decaux a souhaité mettre en œuvre, en sollicitant l'autorisation d'équiper à titre expérimental six des panneaux lumineux qu'elle exploite sur la grande dalle d'un dispositif de comptage des piétons destiné à mesurer efficacement l'audience des publicités qui y sont projetées. Toujours est-il que la CNIL, compétente en vertu de l'article L. 581-9 du code de l'environnement pour connaître de l'installation de « Tout système de mesure automatique de l'audience d'un dispositif publicitaire ou d'analyse de la typologie ou du comportement des personnes passant à proximité », a refusé d'autoriser ce dispositif. La société JC Decaux vous demande d'annuler pour excès de pouvoir la délibération n° 2015-255 du 16 juillet 2015 refusant l'autorisation sollicitée et constituant accessoirement la première prise de position de la CNIL sur un dispositif de mesure d'audience de médias publicitaires.

La demande de JC Decaux trouve son origine dans les difficultés de mesure d'audience des supports publicitaires implantés sur la dalle piétonne de la Défense. En effet, les procédés classiques, c'est-à-dire les mesures d'audience des réseaux de publicité extérieure produits par l'institut Affimétrie, reposent sur des enquêtes réalisées par téléphone au sein d'une unité urbaine, auprès d'échantillons de résidents déclarant leurs trajets ; cette façon de faire ne rend qu'imparfaitement compte de la fréquentation de la dalle qui, en sa double qualité de quartier d'affaires et de zone touristique, est fréquentée par un grand nombre de non résidents.

Sans être unique[1], cette caractéristique a semblé suffisamment singulière à la société pour qu'elle envisage de mettre en place, à titre expérimental pour une durée de quatre mois, un dispositif de comptage spécifique. Celui-ci reposerait sur l'installation, sur les panneaux numériques exploités notamment à des fins publicitaires, de six boîtiers wi-fi capables de

[1] D'où l'importance de votre position, la requérante admettant elle-même que si l'expérimentation était autorisée et s'avérait satisfaisante, elle pourrait être reproduite dans d'autres espaces touristiques.

capter les identifiants réseaux autrement dénommés adresses MAC, des appareils mobiles circulant dans un rayon de 25 mètres, à la condition que leur connectivité wi-fi soit activée. Toutes les deux minutes, ces boîtiers transmettraient à un serveur hébergé à Francfort les données collectées au cours des deux minutes écoulées : identifiant réseau, horaire exact de détection de cette adresse et puissance d'émission du signal wi-fi, permettant d'extrapoler la distance approximative séparant le boîtier de l'appareil mobile. Grâce à l'identifiant réseau dont on peut raisonnablement supposer qu'il se rattache à une personne donnée, il serait ainsi possible de déterminer non pas seulement le nombre de passages devant les panneaux, mais le nombre d'individus qui y sont exposés, tout en mesurant des taux de répétition et en modélisant des schémas de mobilité.

La CNIL a, pour apprécier la demande d'autorisation, passé ce dispositif au crible de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, alors même que la demande d'autorisation puisait, nous l'avons dit, sa base légale dans l'article L. 581-9 du code de l'environnement. Il nous semble qu'en l'absence de toute disposition spéciale, autre que celle d'attribution de compétence, dans le code de l'environnement, il s'agissait de la bonne manière de procéder. Les travaux préparatoires de la loi n° 2010-788 du 12 juillet 2010 dite Grenelle II dont sont issues ces dispositions[2] confirment d'ailleurs que l'intention du législateur était de doter la CNIL d'un droit de regard sur les dispositifs de comptage au regard de sa grille habituelle d'examen des demandes relatives aux traitements de données à caractère personnel.

A l'aune de cette grille, la CNIL est d'abord partie du principe que les adresses MAC, bien qu'elles se rattachent à un appareil et non à un individu, constituent des données identifiantes de nature à conférer dispositif le caractère d'un traitement de données personnelles. Ce point ne fait pas débat entre les parties et il nous semble juridiquement incontestable : la Cour de justice de l'Union européenne a déjà jugé que les adresses IP attribuées à chaque branchement d'un appareil à un réseau informatique utilisant le protocole internet constituent des « données protégées à caractère personnel, car elles permettent l'identification précise des utilisateurs » (CJUE, 24 novembre 2011, *Scarlet Extended*, aff. C-70/10, point 51), position qui est également celle du Conseil constitutionnel (décision n° 2009-580 DC du 10 juin 2009, loi favorisant la diffusion et la protection de la création sur internet, considérant 27) et de la 1ère chambre civile de la Cour de cassation (Cass. 1ère civ., 3 novembre 2016, n° 15-22.595)[3]. Nous ne voyons aucune raison justifiant de raisonner, pour « l'adresse physique » que constitue l'adresse MAC, autrement que pour l'« adresse logique » que constitue l'adresse IP.

La CNIL a ensuite estimé que le traitement envisagé était susceptible de trouver une base légale dans le 5° de l'article 7 de la loi informatique et libertés qui, par dérogation aux dispositions qui imposent le consentement des personnes pour la collecte de données personnelles, permet une collecte sans consentement lorsqu'elle satisfait : « La réalisation de l'intérêt légitime poursuivi par le responsable du traitement ou par le destinataire, sous réserve de ne pas méconnaître l'intérêt ou les droits et libertés fondamentaux de la personne concernée. » Ce point, sur lequel pour le coup, nous nourrissons des doutes que nous vous exposerons tout à l'heure, n'est absolument pas débattu devant vous, de sorte que vous ne pourrez pas vous en saisir pour en faire un motif déterminant de votre décision.

[2] Introduites par amendement parlementaire en première lecture à l'Assemblée nationale.
[3] La chambre criminelle avait retenu une analyse différente avant la décision du Conseil constitutionnel : Cass. crim., 13 janvier 2009, 08-84088, Bull.

La CNIL a ensuite examiné si l'intérêt et les droits des personnes étaient respectés, en procédant en deux temps. Premièrement, elle a estimé que le traitement ne ferait pas l'objet d'une anonymisation suffisante pour entrer dans les prévisions du IV de l'article 32 de la loi informatique et libertés, dispositions qui allègent considérablement les obligations d'information des usagers pesant sur le responsable d'un traitement dans le cas où « les données à caractère personnel recueillies sont appelées à faire l'objet à bref délai d'un procédé d'anonymisation ». Deuxièmement, elle a estimé que les obligations d'information (non allégées) des personnes dont les donnés personnelles seraient traitées ne seraient pas suffisamment remplies.

Commençons par évacuer les deux moyens de légalité externe qui ne sont pas fondés : la délibération de la CNIL est suffisamment motivée, et la commission n'avait pas à alerter la société sur l'importance de la question de l'information des personnes de sorte qu'elle ne l'a pas induite en erreur en ne lui signalant pas en cours d'instruction ses attentes précises sur ce point.

Au fond, l'essentiel du débat contentieux porte sur l'existence ou non d'un procédé d'anonymisation suffisant pour faire échapper le traitement aux dispositions les plus contraignantes de la loi informatique et libertés. Vous ne pourrez donc pas venir à bout du débat juridique sans vous plonger des les détails techniques du procédé imaginé par la société JC Decaux pour brouiller les adresses MAC. Nous allons vous le décrire par le menu, menu dans lequel il sera question de hachage et de salage, termes dont nous espérons qu'ils vous mettent en appétit.

Ledit procédé consiste à faire subir à l'adresse MAC, transmise au serveur par connexion sécurisée, une série de trois opérations préalables à son enregistrement. Première étape, l'adresse MAC ferait l'objet d'une opération de troncage, consistant à en retirer le dernier demi-octet (les x derniers chiffres). Deuxième étape, elle ferait l'objet d'une opération de salage, procédé qui consiste à ajouter une chaîne de caractères aléatoire à l'information concernée pour, en quelque sorte, la travestir. Une fois salée, l'adresse MAC ferait – c'est la dernière étape, l'objet d'une opération de hachage, consistant cette fois à calculer, à partir de la donnée de base, une empreinte de taille fixe la rendant méconnaissable. Ce n'est que la version découpée, salée puis hachée de l'adresse MAC qui serait enregistrée par le serveur et conservée pendant la durée de l'expérimentation. Si vous nous permettez de filer la métaphore culinaire dont ces termes techniques sont directement inspirés, l'effet recherché est de rendre aussi difficile la reconstitution de l'adresse MAC à partir des données sorties de cette moulinette informatique que celle d'un morceau de viande à partir de la farce sortie de la moulinette du boucher.

La CNIL a estimé que cet objectif n'était pas atteint, la technique retenue constituant un procédé de pseudonymisation et non d'anonymisation au sens de l'article 32 de la loi. Cette distinction, que la CNIL emprunte au G29 dans son avis 05/2014 du 10 avril 2014 sur les techniques d'anonymisation, donne prise à deux groupes de moyens.

D'abord, la CNIL aurait commis une double erreur de droit en se fondant l'avis du G29 alors même qu'il comporterait des erreurs et sans appliquer en outre ses préconisations tendant à apprécier le risque de ré-identification des personnes à la lumière du contexte et non de façon abstraite. Pris en tant que tel, le moyen est inopérant, la CNIL n'ayant pas fait application de l'avis 10 avril 2014, dépourvu de valeur juridique, qu'elle n'a ni visé, ni cité et dont elle s'est contentée de s'inspirer.

Ensuite, la CNIL se serait trompée en estimant que le hachage après salage et troncage constituerait une pseudonymisation insatisfaisante. L'erreur serait à la fois de droit, à avoir estimé que l'article 32 impose de rendre impossible l'individualisation d'une personne même non identifiée et d'appréciation, à avoir estimé pour ce motif et alors que les risques de réidentification sont minimes, que le dispositif envisagé ne remplit pas la condition d'anonymisation. Nous n'avons pas de réticence à exercer un contrôle normal, malgré la technicité de la matière, qui est l'un des motifs traditionnels (discutable) de dérogation au contrôle normal que vous êtes en principe supposés assurer. Nous sommes en effet convaincue qu'il est impossible d'exercer le moindre contrôle, fût-il restreint, sans se plonger dans les détails techniques. L'investissement étant inévitable, il nous semble du coup indolore d'exercer un plein contrôle, opportun compte tenu des droits et libertés en jeu.

En l'espèce, la question est délicate, mais nous pensons que le moyen doit être écarté.

Le IV de l'article 32 mobilise la notion de « procédé d'anonymisation » sans la définir. Sans doute faut-il la lire à la lumière du considérant 26 de la directive 95/46/CE du 24 octobre 1995[4], qui pose qu'il y a anonymisation dès lors que « la personne concernée n'est plus identifiable ». Ce considérant précise encore que « pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre, soit par le responsable du traitement, soit par une autre personne, pour identifier ladite personne », exigence qui, contrairement à ce que soutient la société dans une esquisse de moyen d'inconventionnalité de la loi informatique et libertés au regard de la directive, est traduite à l'article 2 de cette loi, qui prescrit « pour déterminer si une personne est identifiable, (…) de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne. » Quant à l'identification, elle est définie par la directive et la loi comme la possibilité de remonter à l'identité de la personne « directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres ».

A cet égard, la CNIL, reprenant à son compte la position du G29, qualifie de procédés d'anonymisation les techniques ayant pour effet une dilution des informations propres à chaque individu de sorte que ses caractéristiques ne puissent plus être isolées en tant que telles. Deux familles de techniques peuvent prétendre à cette qualification : les techniques dites de randomisation (par injection de bruit ou permutation de données), consistant, pour faire simple, à introduire des anomalies dans les valeurs propres à un individu ; et les techniques dites de généralisation (agrégation, k-anonymat, l-diverté et t-proximité), consistant à noyer certaines de ces valeurs dans un ensemble plus large. Dans les deux cas, le résultat est un brouillage de l'identité virtuelle, qui ne correspond plus à un individu réel donné, de sorte qu'il est impossible d'isoler les données relatives à un individu réel dans le traitement anonymisé.

En revanche, la CNIL est plus circonspecte quant aux procédés de pseudonymisation, qui ont en commun de remplacer l'identité de la personne réelle par un nom de code plus ou moins sophistiqué. Il en va ainsi des techniques de chiffrement ou cryptage comme des différentes techniques de hachage. Dans ces systèmes, les individualités virtuelles conservées

[4] directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données.

continuent de correspondre à des individualités réelles ; il est simplement difficile de recouper les deux individualités.

La société soutient qu'il est artificiel de poser que l'anonymisation satisfait par principe à la loi et que la pseudonymisation n'y satisfait pas, dans la mesure où toute technique d'anonymisation est potentiellement vulnérable et qu'en matière de ré-identification, le risque zéro n'existe pas. Il est vrai que la lecture de l'avis du G29 est édifiante quant à la sophistication des techniques de contournement de l'anonymisation. Il reste que la façon de voir de la société est assez artificielle. Le fait que les dispositifs d'anonymisation soient plus ou moins robustes et qu'il faille au cas par cas que la CNIL examine si un procédé d'anonymisation donné atteint véritablement son but n'enlève rien au fait que la pseudonymisation puisse, pour sa part, présenter une fragilité structurelle, tendant à ce qu'elle n'est résistante ni l'individualisation, ni à la corrélation. La véritable difficulté pour disqualifier par principe toute forme de pseudonymisation se nicherait plutôt, et seulement pour l'avenir, dans le règlement du parlement européen et du conseil 2016/679 du 27 avril 2016, qui fait une place aux dispositifs de pseudonymisation pourvu qu'ils soient d'une sophistication telle que la reconstitution de l'identité ne soit pas raisonnablement envisageable. Même dans ce cadre, la mise en place d'un système de pseudonymisation ne vaudrait pas blanc-seign.

La seule question est donc de savoir si le type de traitement pseudonymisé refusé par la CNIL dans sa délibération présente une fragilité structurelle emportant un risque de réidentification.

La réponse nous semblerait sans aucun doute positive s'il reposait sur une technique de chiffrement, dont le propre est d'être réversible à la seule condition d'avoir connaissance du code. Dès lors que par construction, ce code est au moins connu du responsable du traitement, la ré-identification par ce dernier est structurellement possible et il n'y a pas anonymisation au sens des exigences légales.

Le doute est davantage permis pour une technique de hachage après troncage et salage.

Le troncage en lui-même n'apporte pas grand-chose, du moins en tant qu'il porte sur une fraction de données réduite : les parties s'accordent à dire qu'il aboutit seulement à ce que 16 adresses IP possibles donnent la même empreinte. Sur 281 475 milliards d'adresses MAC possibles correspondant à 2 milliards de *smartphones* dans le monde dont 27,7 millions en France, un « taux de collision » aussi faible n'est certainement pas de nature à empêcher des recoupements frôlant la certitude.

Le hachage présente pour sa part l'avantage d'être irréversible, ni les tiers ni le responsable du traitement ne pouvant remonter directement à l'information initiale à partir de son empreinte hachée. Il n'en reste pas moins vulnérable aux attaques dites par force brute, consistant à tester toutes les entrées possibles, dès lors que leur fourchette de valeur est connue, afin de constituer des tableaux de correspondance permettant de relier les valeurs d'entrée à leur empreinte. Il peut certes sembler difficile de lancer une attaque par force brute à partir de 280 000 milliards d'adresses MAC de 48 bits possibles : mais en réalité, un ordinateur personnel performant permet déjà de tester quelques millions de combinaisons par seconde et l'on place généralement autour de 128 bits la limite technique de réidentification par attaque brute.

Le salage enfin permet de renforcer significativement l'efficacité du hachage en brouillant les correspondances entre empreinte et donnée initiale. Il reste que dès lors que le « sel » est connu du responsable du traitement, il peut être neutralisé par lui aussi bien que découvert par un tiers. En tout état de cause, la littérature scientifique estime qu'il reste possible, avec des moyens raisonnables, de calculer la valeur originale de l'attribut qui se cache derrière le résultat d'une fonction de hachage avec salage, dès lors du moins que le type d'attribut et son format sont- connus, ce qui est le cas pour les adresses MAC dont le format est stable et les 280 000 milliards de combinaisons mathématiquement possibles à reconstituer.

Plus généralement, un dispositif pseudonymisé, c'est-à-dire qui conserve la possibilité d'individualiser une personne, nous semble structurellement vulnérable à la réidentification dès lors que les données non altérées comportent, par elles-mêmes, des indices donnant prise à une réidentification. Or tel est assurément le cas des données de géolocalisation collectées en l'espèce. Une étude du MIT qui a fait sensation et dont les deux parties se prévalent a ainsi démontré, à partir de l'analyse d'un ensemble de données pseudonymisées couvrant 15 mois de coordonnées mobiles de 1,5 millions de personnes dans un rayon de 100 km, que quatre points de localisation permettaient d'isoler 95% des personnes et que deux points suffisaient à en isoler 50%[5]. Dès lors et indépendamment même de la question de savoir s'il est possible de remonter directement de l'empreinte à l'identité, on comprend qu'il suffit qu'une personne connaisse avec certitude certaines habitudes de déplacement propres à personne donnée pour qu'il puisse la retrouver aisément parmi les identités virtuelles et accéder ainsi au détail de ses trajets quotidiens, voire de ses fréquentations. Dans ces conditions, il ne nous semble pas déraisonnable d'estimer que ce dispositif pseudonymisé d'enregistrement de données indirectement identifiantes, parce qu'il n'interdit pas d'individualiser une personne ou de relier entre elles des données résultant de deux enregistrements qui la concernent, n'est pas anonymisé au point d'échapper aux obligations d'information posées par la loi informatique et libertés. C'est cette fragilité, qui s'ajoute au risque résiduel d'un décryptage massif du traitement, qui nous convainc du bien-fondé de la position de la CNIL.

A ce postulat, la société oppose encore deux arguments qui ne nous semblent pas porter. D'abord, elle soutient qu'il suffirait, pour juguler cette fragilité structurelle, que la CNIL interdise dans sa délibération le croisement des dispositifs pseudonymisés avec tout autre ensemble de données. A l'argument selon lequel un tel interdit ne protègerait pas des attaques des tiers, la société répond qu'on ne peut pas miser par principe sur la réalisation de comportements pénalement réprimés. Nous somme sensible à cette logique et il est vrai qu'en matière de fichiers, la protection des données repose en grande partie sur le pari que les règles d'accès et de conservation, techniquement contournables, seront effectivement respectées ; sans ce pari, l'exigence de protection des données conduirait vraisemblablement à paralyser toute exploitation de données personnelles, ce qui n'est pas l'objectif de la législation qui recherche un juste équilibre. Il reste que ce pari n'est pas total, comme en témoigne le fait que les règles d'accès sont généralement assorties d'obligations de traçabilité rendant possibles des contrôles, qui n'auraient pas lieu d'être si les obligations juridiques se suffisaient à elles-mêmes (v., sur ce point, le règlement de 2016). Du reste, la condition d'anonymisation en cause ne conditionne pas la possibilité de traiter ou non les données, mais la faculté de se dispenser ou non d'une information complète des personnes. S'agissant d'ouvrir une faculté

[5] Y.-A de Montjoye, C. Hidalgo, M. Verleysen et V. Blondel, "Unique in the crowd: The privacy bounds of human mobility", *Nature*, n° 1376, 2013.

dérogatoire, le fait que la CNIL s'assure de l'impossibilité technique de réidentifier ne nous choque pas.

Ensuite, la société soutient que soumettre le dispositif aux droits d'opposition et de rectification revient à exiger qu'elle se mette en mesure de réidentifier les personnes et renonce donc aux mesures d'anonymisation prises pour limiter les risques d'attaques : c'est faux, car une personne souhaitant exercer ses droits n'aura qu'à fournir son adresse MAC pour que soit rejouée la fonction de hachage et que les données la concernant puissent lui être communiquées.

Dans ces conditions, dès lors que nous sommes bien en présence d'un dispositif impliquant par nature l'individualisation et la corrélation, nous ne voyons ni erreur de droit, ni erreur d'appréciation à avoir refusé de dispenser le dispositif des obligations d'information.

Une fois franchi ce pas, vous ne pourrez qu'écarter les deux autres moyens.

L'un est tiré de ce que la CNIL aurait fait peser à tort sur le traitement les obligations d'information relatives aux collectes directes de données alors qu'il s'agissait de collecte indirecte. Mais d'une part, nous ne croyons pas que le fait que les données soient collectées par une machine et non directement par l'exploitant du traitement lui confère le caractère d'une collecte indirecte. Cette expression renvoie au cas où les données ont été récupérées auprès d'une autre base de données (par exemple un achat de fichier) et non aux cas où la collecte a été effectuée par une machine exploitée par le responsable du traitement (contrairement à ce qui est soutenu, els textes européens ne disent pas le contraire). D'autre part et en tout état de cause, la qualification de collecte indirecte ne change pas la teneur des informations à fournir, mais simplement le moment de la première information. Il n'y a dérogation totale, aux termes du III de l'article 32, que dans des hypothèses très particulières dont la société ne soutient même qu'elle en relèverait.

L'autre moyen est tiré de l'erreur de droit et de dénaturation à avoir estimé les modalités d'information mises en place insuffisantes. Il découle de ce que nous avons dit que la société était tenue aux obligations d'information maximales. Il n'y a donc pas d'erreur de droit sur ce point, pas plus qu'il n'y a de dénaturation à avoir estimé que des écriteaux succincts posés sur les panneaux lumineux seraient insuffisants.

Pour finir, la société fait valoir qu'à ce compte, aucune information suffisante des personnes ne sera jamais possible dans le cadre d'un fonctionnement normal du traitement et que les exigences de la CNIL reviennent purement et simplement à l'interdire dans son principe. C'est peut-être vrai, mais cela nous effraie d'autant moins que, contrairement à la CNIL, nous ne pensons pas que la finalité d'un tel traitement soit légitime au point de permettre un recueil sans consentement préalable de données personnelles, dans des conditions de collecte d'autant plus intrusives qu'elles concernent des personnes qui n'ont aucun intérêt au traitement et qui devraient, si elles souhaitaient s'en prémunir, désactiver une fonctionnalité de leurs moyens de communication qui leur est par ailleurs utile pour de toutes autres raisons. La société fait valoir que pourtant, le législateur n'a pas souhaité interdire les systèmes de mesure automatique d'audience publicitaire ou d'analyse de la typologie ou du comportement des passants, puisqu'il a confié à la CNIL le soin d'examiner les demandes. Cela ne nous semble pas décisif, d'une part parce que le législateur a surtout entendu éviter que ces dispositifs se développent sans que leur conformité au droit des traitements soit

vérifiée par quiconque, d'autre part parce que reste possible la mise en place de dispositifs moins intrusifs, même si cela implique une moins grande efficacité.

PCMNC – Rejet.