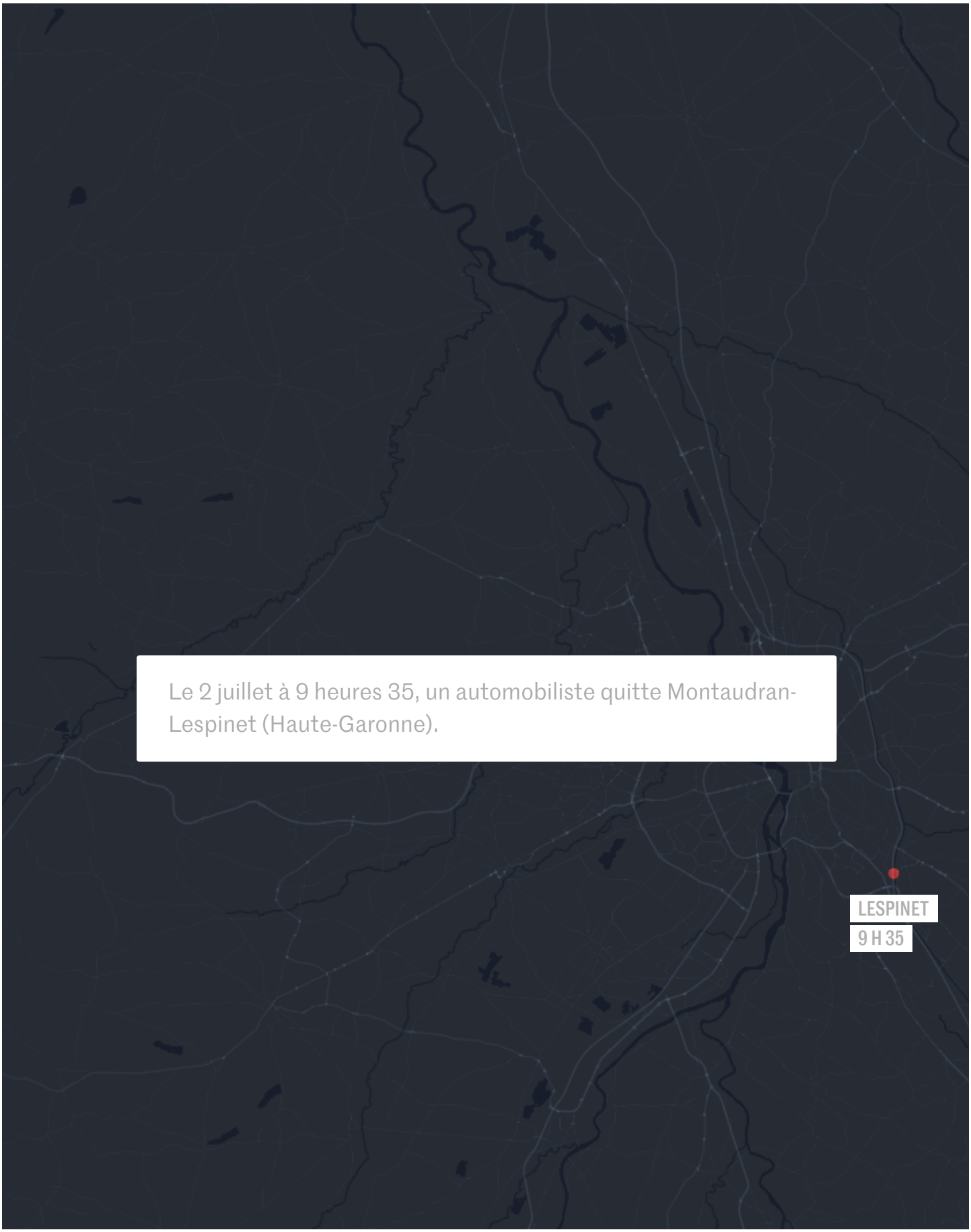

Le 2 juillet à 9 heures 35, un automobiliste quitte Montaudran-Lespinet (Haute-Garonne).
LESPINET
9 H 35
Protomaps © OpenStreetMap

Il contourne Toulouse par l’est et se gare, à 10 h 16, sur le parking du bowling de Saint-Jory.

LESPINET
9 H 35

Si nous savons tout cela, ce n’est pas parce que cette personne nous l’a dit. C’est un gigantesque fichier marketing qui nous l’a appris.

LESPINET
9 H 35

Dans cette base données, on trouve des millions de coordonnées géolocalisées, comme celles récoltées par l'application WordBit, que notre conducteur utilise pour parfaire son anglais.

LESPINET
9 H 35

Le Monde et ses partenaires ont enquêté sur ces super fichiers, souvent constitués par des courtiers en données au mépris de la protection de la vie privée. Avec des risques pour la sécurité des individus, voire des Etats.

LESPINET
9 H 35

• VIE PRIVÉE

Données personnelles en vente libre : les « data brokers », une industrie hors de contrôle

Par Adrien Sénécat, Martin Untersinger, Elsa Delmas (développement), Léa Girardot (design) et Anne Morel (design)

Publié aujourd'hui à 05h00, modifié à 09h09

Lecture 8 min.

ENQUÊTE | Les données personnelles géolocalisées de millions d'utilisateurs sont agrégées par des courtiers en données, dans des conditions douteuses. Une menace pour la vie privée des personnes, révèle une enquête du « Monde », menée en partenariat avec plusieurs médias internationaux.

Quand elle veut s'occuper, Marie-Claire sort son téléphone et s'installe dans le fauteuil du salon de sa petite maison de Pleubian (Côtes-d'Armor). La retraitée de 66 ans peut passer des heures sur *Candy Crush*, *Farm Heroes* et bien d'autres jeux. Elle ne se doutait pas, en revanche, que ses distractions préférées pouvaient collecter et revendre des informations sur elle : ses heures de connexion, son modèle de smartphone ou sa position géographique.

Ces éléments figurent dans un gigantesque fichier vendu par un courtier en données américain, Datastream Group. *Le Monde* et huit médias partenaires

ont eu accès à un échantillon commercial de cette base de données, dans laquelle figurent déjà plus de 47 millions de personnes.

A première vue, ces mobinautes semblent anonymes, car ils apparaissent sous l'identifiant publicitaire (unique) associé à leur téléphone. Cependant, il ne nous a fallu que quelques minutes pour remonter à Marie-Claire en suivant la trace d'un joueur de *Words of Wonders* en Bretagne, ses coordonnées GPS permettant d'identifier son domicile.

Les « Data Brokers Files », une enquête internationale

Cette enquête a été lancée à l'initiative de nos partenaires de *Netzpolitik.org*, qui ont obtenu un échantillon de la base de données de l'entreprise Datastream Group, partagé avec de potentiels clients. On y trouve des informations sur plus de 47 millions d'appareils mobiles dans 137 pays du monde pour la journée du 2 juillet 2024, dont un million en France.

Au total, 380 millions de coordonnées géographiques, d'un degré de précision variable, collectées par 39 499 applications, figurent dans ce fichier. Les plus fines, fiables à quelques mètres près, s'appuient sur le GPS des appareils. D'autres se fondent sur l'adresse IP des smartphones et donnent une indication plus approximative – à l'échelle d'un quartier ou d'une région.

Neuf médias ont pris part à l'enquête sur ces « Data Brokers Files » : *Wired* et *404 Media* (Etats-Unis), *Le Monde* (France), SRF et RTS (Suisse), NRK Beta (Norvège), BNR Nieuwsradio (Pays-Bas), Bayerischer Rundfunk et *Netzpolitik.org* (Allemagne). Un premier volet de cette série d'articles, qui explique comment des applications

Voir plus

9 H 35

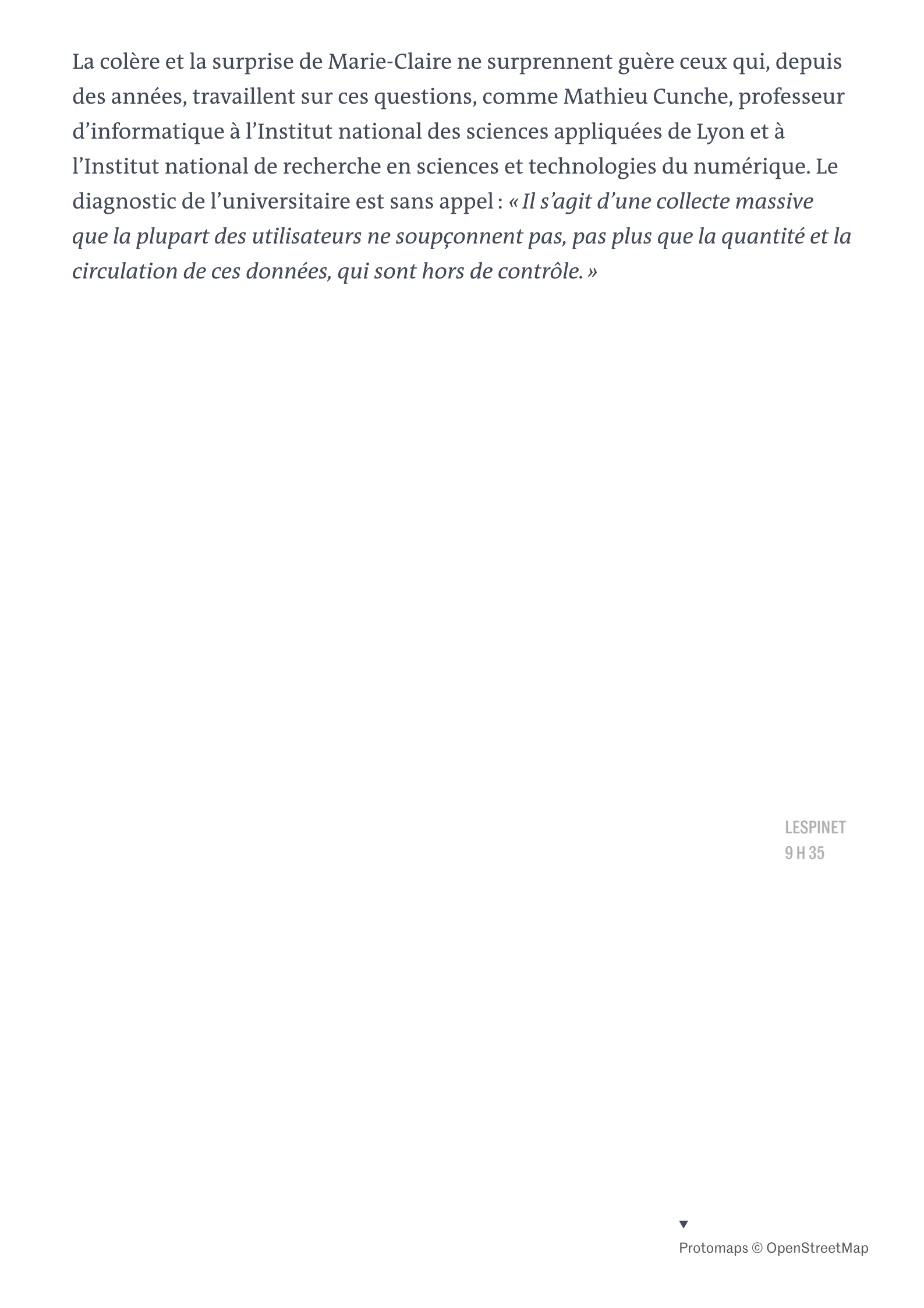
Etonnée de notre appel, elle nous a confirmé s'adonner régulièrement à ce jeu, sans savoir que son créateur tirait profit de ses données. *« Pour moi, ce n'est pas clair du tout. Maintenant que vous me l'apprenez, je vais le désinstaller ! »*, peste-t-elle.

De gigantesques fichiers constitués de manière opaque

▼

Protomaps © OpenStreetMap

La colère et la surprise de Marie-Claire ne surprennent guère ceux qui, depuis des années, travaillent sur ces questions, comme Mathieu Cunche, professeur d'informatique à l'Institut national des sciences appliquées de Lyon et à l'Institut national de recherche en sciences et technologies du numérique. Le diagnostic de l'universitaire est sans appel : *« Il s'agit d'une collecte massive que la plupart des utilisateurs ne soupçonnent pas, pas plus que la quantité et la circulation de ces données, qui sont hors de contrôle. »*



UN COMMERCE AVEC DE MULTIPLES INTERMÉDIAIRES

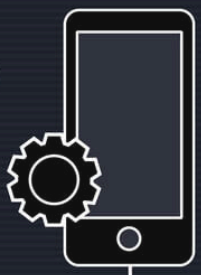
De nombreux intervenants peuvent collecter et transmettre les données, dans le respect des règles ou de manière frauduleuse.

① UTILISATEUR

L'application collecte un certain nombre de données personnelles : l'utilisateur doit donner son consentement explicite, en particulier lorsqu'elles sont utilisées à des fins publicitaires.

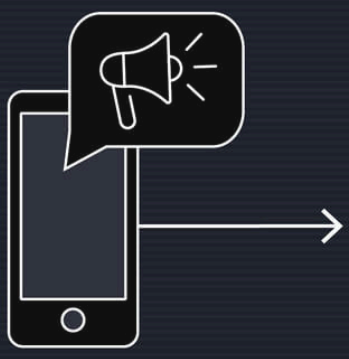
② L'ÉDITEUR DE L'APPLICATION

Fait appel à des services tiers à des fins publicitaires, marketing ou techniques.



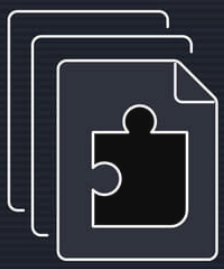
③ LES SERVICES TIERS

Peuvent accéder à tout ou partie des données et les transmettre à d'autres intermédiaires.



④ LE COURTIER DE DONNÉES

Agrège des données qu'il se procure auprès d'applications ou de services tiers.



Les *data brokers* eux-mêmes restent discrets sur l'origine de leurs données. Plusieurs circuits existent. Elles peuvent d'abord être collectées et partagées directement par les applications. La transmission peut aussi se faire par le biais d'une de leurs briques logicielles – les « SDK » (pour « Software Development Kit »). Ces outils, gratuits, sont utilisés par de nombreux éditeurs pour faciliter le développement et la monétisation de leurs applications. Les entreprises tierces qui les proposent se rémunèrent, quant à elles, grâce aux données personnelles des mobinautes.

Des données s'échangent également par le biais du marché publicitaire à enchères en temps réel (dit RTB, pour « real-time bidding »), très répandu dans la réclame en ligne. Concrètement, à chaque fois qu'un encart publicitaire va être disponible sur l'écran d'un internaute, l'application diffuse son profil auprès d'une myriade d'annonceurs potentiels (ou de leurs représentants) : « Qui veut envoyer une publicité à un utilisateur du Boncoin à Lille ? » L'encart est ensuite attribué au plus offrant dans le cadre de transactions ultrarapides (la durée des échanges se chiffre en millisecondes). Mais ce processus a un effet de bord : les enchérisseurs potentiels ont accès aux (nombreuses) données d'internautes en circulation. Dès lors, certains acteurs peuven détourner ce système afin d'alimenter leurs fichiers.

Pour ne rien simplifier, il peut aussi y avoir des incohérences dans les fichiers des data brokers : mauvais horaires, mauvais nom d'application… Ces irrégularités peuvent être le fruit d'erreurs techniques apparues au fil de la constitution du fichier ou de tromperies sur l'origine de certaines données.

Plusieurs éléments laissent penser que tout ou partie des données de Datastream pourrait provenir d'Eskimi, une plateforme publicitaire lituanienne. Le data broker américain l'identifie comme la source d'un de ses

autres fichiers, très proche de celui sur lequel nous avons enquêté, dans une lettre à un sénateur américain enquêtant sur les données publicitaires.

Eskimi est par ailleurs référencé comme partenaire d'une grande part des applications de notre base de données. L'entreprise a également des liens étroits avec Redmob, un courtier sis à Dubaï, qui propose justement des fichiers issus d'enchères publicitaires : son fondateur, Vytautas Paukstys, est aussi le patron d'Eskimi. Interrogé, celui-ci affirme qu'Eskimi n'a aucune relation commerciale avec Datastream, qu'il n'a pas d'activité de courtage en données et que Redmob, de son côté, respecte toutes les lois en vigueur.

Dans un communiqué, Datastream nous assure, pour sa part, se procurer des données de manière légitime, *« par le truchement d'une entreprise tierce qui a bonne réputation »*. Elle ajoute estimer que *« ces données anonymisées ne peuvent, à elles seules, révéler des informations sensibles »*. Nos interrogations sur les risques de ce pistage ne seraient, par conséquent, qu'une *« présentation trompeuse des faits »*.

Des risques pour les individus, voire les Etats

Explorer ce fichier, c'est aussi mesurer la quantité d'informations qu'il agrège sur un individu précis. Les identifiants publicitaires associés aux smartphones permettent de croiser les données glanées sur un mobinaute, application par application, facilitant un profilage assez sophistiqué.

LESPINET
9 H 35

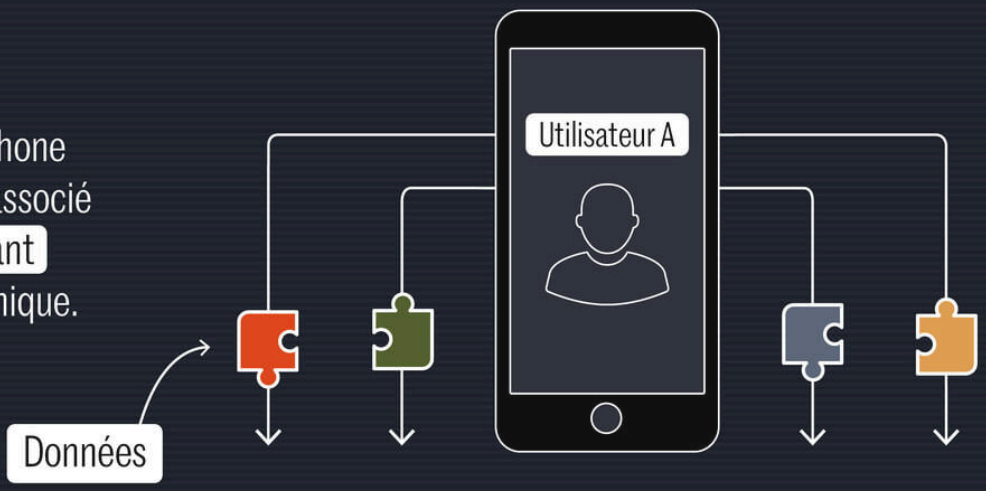
Plusieurs entreprises vendent d'ailleurs des études qui s'appuient sur ces fichiers. En France, c'est par exemple le cas du groupe CBRE, qui propose des études de fréquentation de zones urbaines, vendues notamment aux collectivités territoriales, à partir des données géolocalisées de millions d'habitants achetées à un data broker. L'entreprise utilise notamment un algorithme qui déduit automatiquement l'adresse des utilisateurs à partir de leurs déplacements.

CROISER LES DONNÉES POUR PROFILER LES UTILISATEURS

Les super fichiers agrègent des données issues de plusieurs applications pour une même personne.

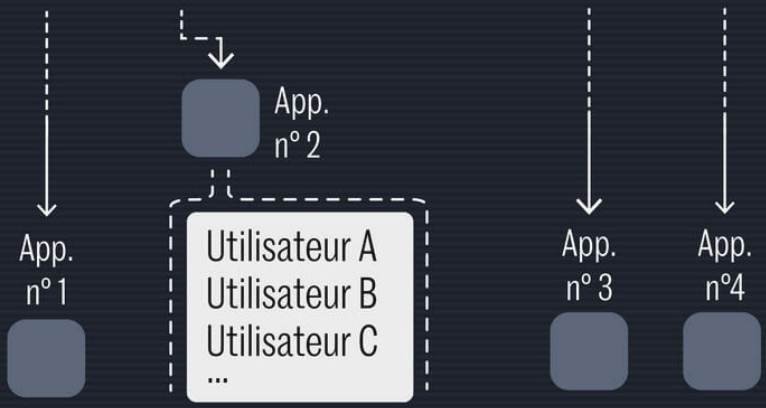
1

Chaque téléphone portable est associé à un identifiant publicitaire unique.



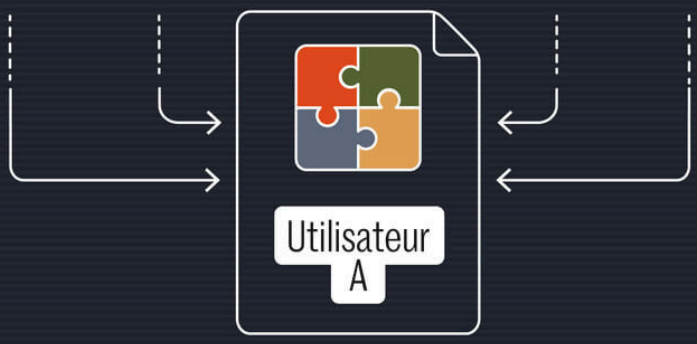
2

Chacune des applications installées sur le portable récolte des données auxquelles est associé l'identifiant du téléphone.



3

Grâce à cet identifiant, les courtiers de données sont en mesure de constituer un profil précis de l'utilisateur.

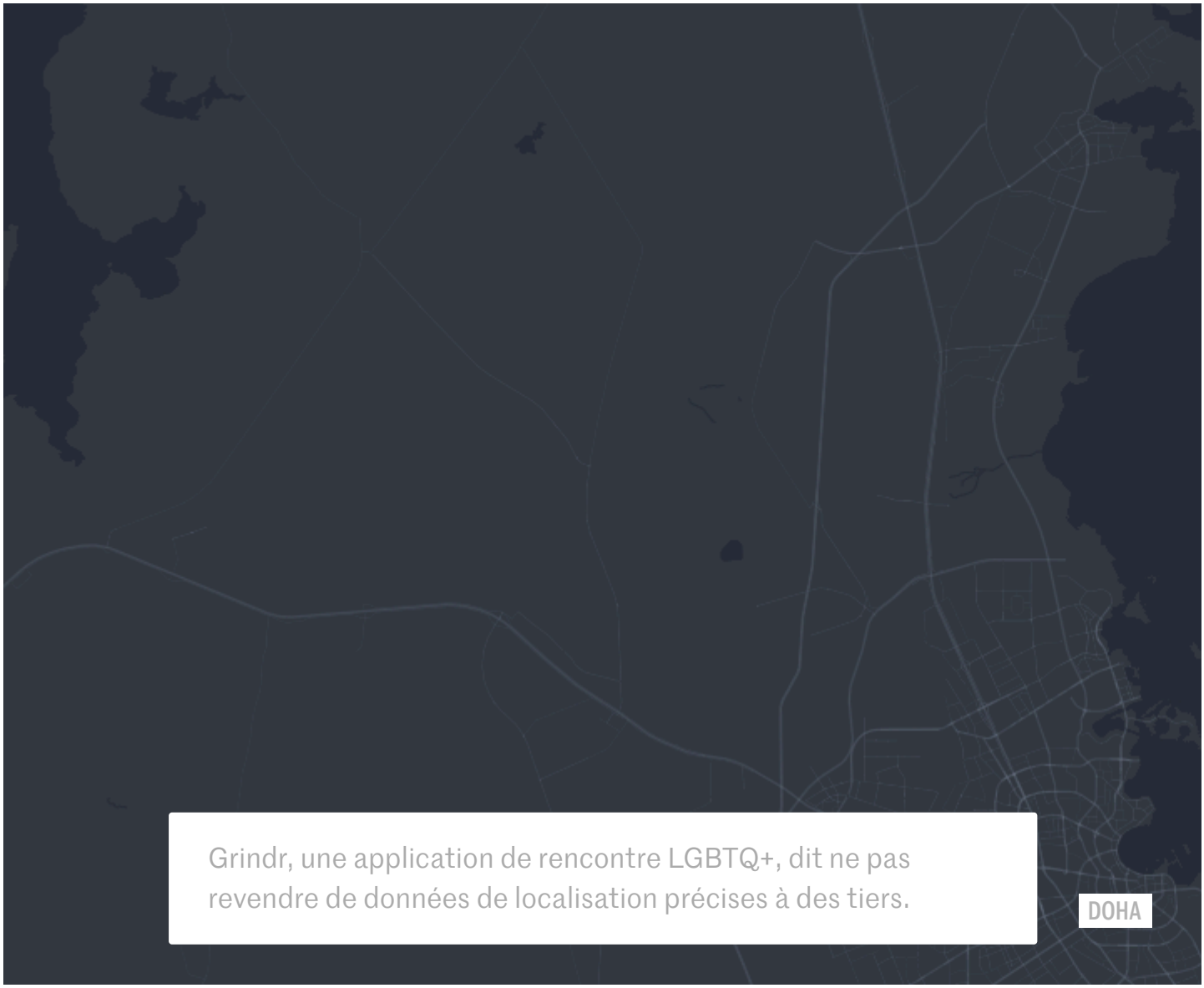


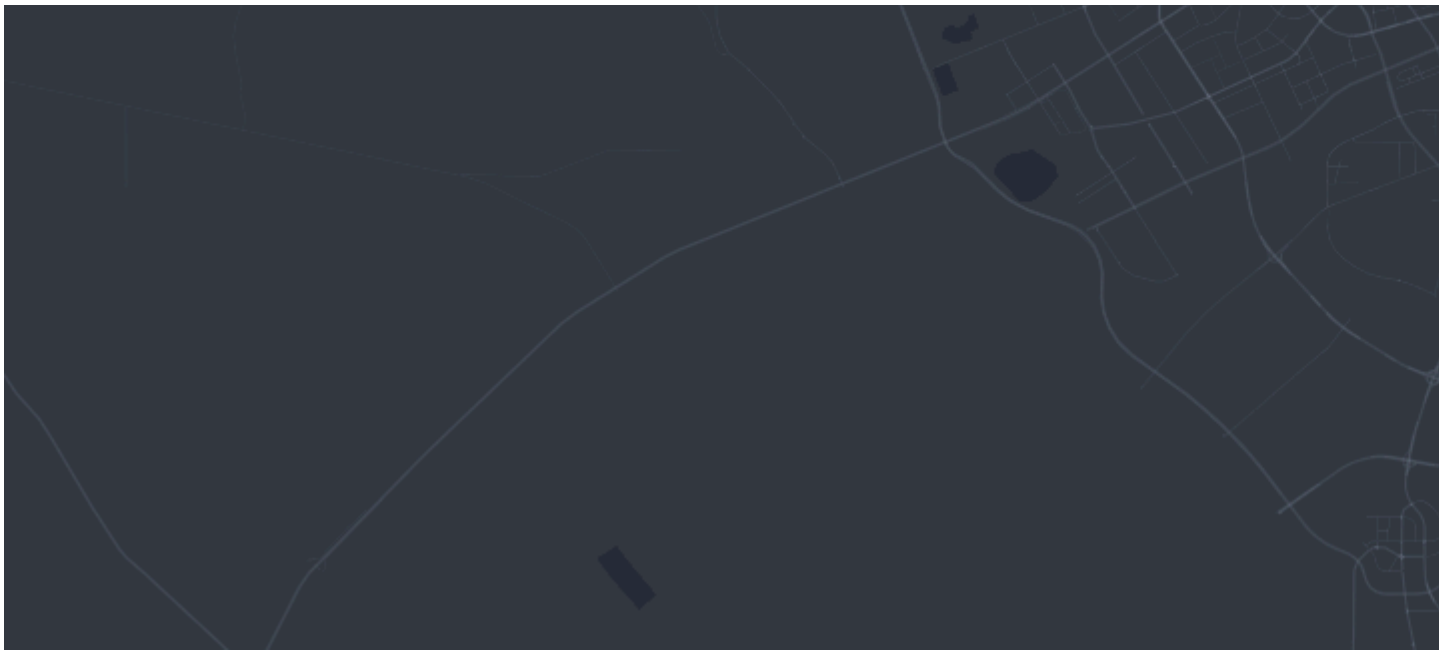
Protomaps © OpenStreetMap

Un autre problème est apparu au fil de nos recherches : parmi les milliers de services qui font commerce des données personnelles de leurs utilisateurs, bon nombre portent sur des sujets délicats. Dans l'échantillon de Datastream, on trouve ainsi des applications confessionnelles, permettant de lire des pages de la Bible ou du Coran, d'autres dans le domaine de la santé, utilisées par exemple pour surveiller sa pression artérielle, ou encore des sites de rencontres, notamment réservés à la communauté LGBTQ+.

Lire aussi | Comment des applications grand public facilitent le pistage des utilisateurs à leur insu

De telles informations peuvent causer un préjudice si elles tombent entre de mauvaises mains et leur partage est strictement encadré par le droit. En janvier, des millions de données géolocalisées provenant du courtier en données Gravy Analytics ont ainsi été mises en ligne par des hackeurs, rappelant le caractère sensible de telles bases de données.





Protomaps © OpenStreetMap

Dans notre échantillon, on ne trouve que des positions approximatives d’utilisateurs de l’application, comme pour cet utilisateur à Doha (Qatar).

Mais pour les mêmes personnes, on trouve parfois d'autres coordonnées GPS précises, issues d'autres applications, dans la même base de données. Si bien que l'on peut pister des utilisateurs de Grindr de cette manière, y compris dans des

pays où les relations homosexuelles sont réprimées, comme le Qatar.

Interrogé par *Le Monde*, Grindr reconnaît que *« les piratages et les combinaisons de données de différentes sources par des acteurs peu scrupuleux sont une préoccupation pour tous les développeurs d'applications mobiles financées par la publicité »*. L'application estime ne travailler qu'avec des partenaires fiables et précise ne pas proposer de publicités

issues de services tiers *« dans les pays où il est illégal ou dangereux d'être homosexuel »*.

Grâce à ces fichiers, certaines sociétés vendent directement à des Etats – notamment à des services d'enquête américains, – des outils de géolocalisation et de surveillance permettant, par exemple, de savoir qui a fréquenté tel lieu ou d'identifier les participants à une manifestation. Des questions de sécurité nationale se posent également, le même procédé pouvant servir à surveiller les entrées et sorties dans des sites sécurisés – agences de renseignement, institutions, centrales nucléaires… Au cours [illegible] précédente enquête, nos partenaires de Bayerischer Rundfunk et *Netzpolitik.org* avaient démontré que des agents du Bundesnachrichtendienst, le service de renseignement extérieur allemand, pouvaient être pistés de la sorte.

LESPINET
9 H 35

Un cadre juridique en principe strict

Les data brokers seraient-ils sans foi ni loi ? Un cadre juridique rigoureux et bien défini s'impose pourtant à eux : le Règlement général sur la protection des données personnelles (RGPD), le texte de référence au niveau européen sur cette question, ainsi que la directive ePrivacy. La Commis[illegible]

l'informatique et des libertés (CNIL), le « gendarme » français chargé du respect de ces règles, peut aussi bien sanctionner des acteurs à l'étranger manipulant indûment des données d'utilisateurs en France que des sociétés ayant leur siège dans l'Hexagone.

Les données géolocalisées manipulées par les data brokers ne peuvent être considérées comme strictement « anonymes », au sens juridique du terme. Même en l'absence du nom ou du prénom de l'utilisateur, elles *« restent des données à caractère personnel, dans la mesure où elles sont indirectement identifiantes »*, rappelle Nacera Bekhat, cheffe du service de l'économie numérique et du secteur financier à la CNIL. Il en va de l'IP comme de toute information permettant, par croisement, de retrouver l'identité d'une personne.

Le RGPD s'applique donc, conférant des droits aux citoyens concernés (notamment la possibilité de demander la suppression de leurs données) et imposant des obligations aux entreprises, comme stocker les informations de manière sécurisée ou ne pas les conserver trop longtemps. Surtout, l'utilisation des données à des fins publicitaires ne peut se faire qu'avec le consentement de l'utilisateur, qui doit être recueilli dans des conditions précises : être libre, informé, explicite et spécifique.

Or, l'information fournie aux internautes au moment de l'installation d'une application est souvent vague. L'application *Words of Wonders* utilisée par Marie-Claire, par exemple, demande à ses utilisateurs de pouvoir *« suivre* [leurs] *activités dans les apps et sur les sites Web d'autres sociétés »*, un accès présenté comme *« nécessaire pour éviter les publicités indésirables »*. Cela peut-il suffire à légitimer la revente de ces données à des tiers dans des fichiers commerciaux comme ceux de Datastream ? Contacté, son éditeur, Fugo Games, n'a pas donné suite à nos sollicitations.

Plus généralement, les éléments soumis aux utilisateurs n'éclairent guère leur choix. Peu lisent vraiment les conditions d'utilisation d'une application ou consultent la liste des prestataires, notamment publicitaires, à qui les données peuvent être envoyées. Et pour cause : elle compte souvent plusieurs centaines de noms.

Des dérives difficiles à juguler

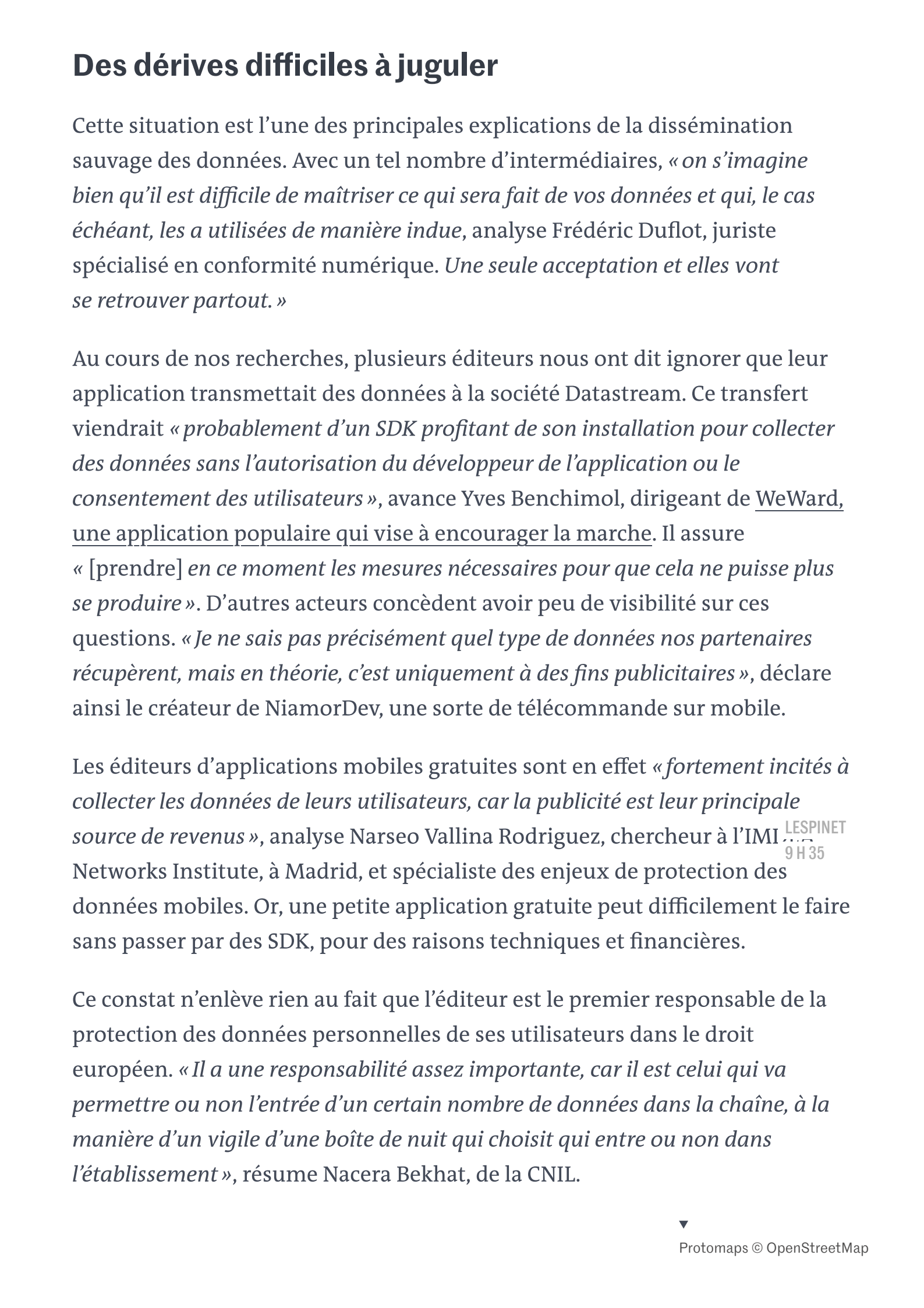
Cette situation est l'une des principales explications de la dissémination sauvage des données. Avec un tel nombre d'intermédiaires, *« on s'imagine bien qu'il est difficile de maîtriser ce qui sera fait de vos données et qui, le cas échéant, les a utilisées de manière indue*, analyse Frédéric Duflot, juriste spécialisé en conformité numérique. *Une seule acceptation et elles vont se retrouver partout. »*

Au cours de nos recherches, plusieurs éditeurs nous ont dit ignorer que leur application transmettait des données à la société Datastream. Ce transfert viendrait *« probablement d'un SDK profitant de son installation pour collecter des données sans l'autorisation du développeur de l'application ou le consentement des utilisateurs »*, avance Yves Benchimol, dirigeant de WeWard, une application populaire qui vise à encourager la marche. Il assure « [prendre] *en ce moment les mesures nécessaires pour que cela ne puisse plus se produire »*. D'autres acteurs concèdent avoir peu de visibilité sur ces questions. *« Je ne sais pas précisément quel type de données nos partenaires récupèrent, mais en théorie, c'est uniquement à des fins publicitaires »*, déclare ainsi le créateur de NiamorDev, une sorte de télécommande sur mobile.

Les éditeurs d'applications mobiles gratuites sont en effet *« fortement incités à collecter les données de leurs utilisateurs, car la publicité est leur principale source de revenus »*, analyse Narseo Vallina Rodriguez, chercheur à l'IMI[illegible] Networks Institute, à Madrid, et spécialiste des enjeux de protection des données mobiles. Or, une petite application gratuite peut difficilement le faire sans passer par des SDK, pour des raisons techniques et financières.

LESPINET
9 H 35

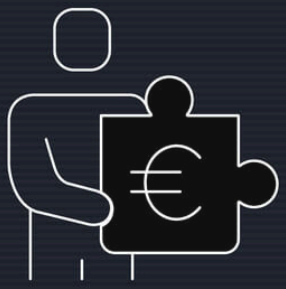
Ce constat n'enlève rien au fait que l'éditeur est le premier responsable de la protection des données personnelles de ses utilisateurs dans le droit européen. *« Il a une responsabilité assez importante, car il est celui qui va permettre ou non l'entrée d'un certain nombre de données dans la chaîne, à la manière d'un vigile d'une boîte de nuit qui choisit qui entre ou non dans l'établissement »*, résume Nacera Bekhat, de la CNIL.



UN COMMERCE OPAQUE ET SOUVENT PROBLÉMATIQUE

Une application doit déclarer quelles données elle récolte, dans quel but et avec qui elle les partage. Mais ce n'est pas toujours le cas...

Certaines applications ne disent pas clairement quels types de données elles collectent.



D'autres ne sont pas claires sur la destination des données qu'elles récupèrent. Par exemple, elles ne disent pas qu'il y aura revente à des tiers.

D'autres encore ne sont pas conscientes du commerce de données, qui ont été revendues à leur insu par un service tiers, notamment par un système d'enchères publicitaires.

Certaines données peuvent être erronées : fraude sur le nom de l'application, la date, la localisation...

Des décisions et sanctions ont déjà visé des acteurs de cet écosystème, mais les autorités peinent à en juguler vraiment les dérives. *« Le circuit peut être très long entre la source et l'utilisateur final de ces données à caractère personnel. S'il y a un acteur qui présente des défauts de conformité n'imp*

chaîne, alors le ver est dans le fruit et il n'est pas toujours évident de retrouver le "péché originel"», analyse Tony Martin, chef du service des contrôles de la CNIL. Si bien que les enquêtes sont longues et complexes.

Le cas de l'application Androïd du *Monde* nous en a apporté une preuve saisissante. Dans l'échantillon de Datastream, des données de géolocalisation approximatives concernant 4 581 identifiants publicitaires sont présentées comme issues de notre édition mobile. Renseignements pris, *Le Monde* n'a aucun contact, juridique ou technique, avec le data broker américain. Dans le cas où de telles données auraient bien été partagées avec lui, ce serait sans notre accord. A ce jour, l'entreprise n'a pas répondu à la mise en demeure qui lui a été adressée.

Explications | Le pistage des internautes par des applications mobiles (et comment s'en protéger) en quatre questions

Adrien Sénécat
Martin Untersinger
Elsa Delmas
développement
Léa Girardot
design
Anne Morel
design

LESPINET
9 H 35