Nations Unies A/72/540

Assemblée générale

Distr. générale
19 octobre 2017
Français
Original : anglais

Soixante-douzième session
Point 72 b) de l'ordre du jour
Promotion et protection des droits de l'homme : questions relatives aux droits de l'homme, y compris les divers moyens de mieux assurer l'exercice effectif des droits de l'homme et des libertés fondamentales

Droit à la vie privée*

Note du Secrétaire général

Le Secrétaire général a l'honneur de transmettre à l'Assemblée générale le rapport soumis par le Rapporteur spécial sur le droit à la vie privée, Joseph A. Cannataci, en application de la résolution 28/16 du Conseil des droits de l'homme.

* Le présent rapport a été soumis après la date limite afin qu'il reflète l'évolution récente de la situation.

Rapport du Rapporteur spécial du Conseil des droits de l'homme sur le droit à la vie privée

Résumé

Le présent rapport est constitué de deux parties : la première contient un résumé analytique des activités menées en 2016 et 2017, et la seconde le rapport d'activité de l'Équipe spéciale en charge des mégadonnées et des données ouvertes créée par le Rapporteur spécial du Conseil des droits de l'homme sur le droit à la vie privée.

Table des matières

I. Aperçu des activités menées par le Rapporteur spécial sur le droit à la vie privée en 2016 et 2017

1. Au cours de la période 2016-2017, particulièrement chargée pour le Rapporteur spécial, 26 manifestations ont été organisées dans 15 pays, sur quatre continents, notamment en vue d'établir des relations avec, entre autres, la société civile, les gouvernements, les forces de l'ordre, les services de renseignement, les autorités de protection des données, les autorités de surveillance du renseignement, le milieu universitaire et le milieu des entreprises. Dans ce cadre, le Rapporteur spécial s'est rendu dans plus de 30 villes, y compris en Asie, en Afrique du Nord et en Amérique centrale; 25 % des activités ont été menées aux États-Unis et plus de 50 % en Europe.

A. Projet d'instrument juridique international sur la surveillance et le droit à la vie privée

2. Les questions de sécurité et de surveillance ont joué un rôle important lors de l'élaboration du mandat du Rapporteur spécial sur le droit à la vie privée par le Conseil des droits de l'homme, en 2015.

3. Conformément à son mandat, énoncé dans la résolution 28/16 du Conseil des droits de l'homme, le Rapporteur spécial a pour attribution « d'identifier les obstacles qui peuvent se poser à la promotion et à la protection du droit à la vie privée, d'identifier, d'échanger et de promouvoir les principes et les pratiques optimales aux niveaux national, régional et international, et de soumettre au Conseil des droits de l'homme des propositions et des recommandations à cet égard, notamment dans l'optique des défis particuliers qui se posent à l'ère du numérique »[1].

4. Le Rapporteur spécial est principalement préoccupé par l'absence de règles de droit international sur la surveillance et la confidentialité dans le cyberespace (à la base des révélations d'Edward Snowden), qui constitue, selon lui, un obstacle majeur au droit à la vie privée. Il estime que ce n'est pas seulement l'absence de règles de fond qui nuit à la promotion et à la protection du droit à la vie privée, mais également l'absence de mécanismes adéquats[2].

5. Dans le cadre de son mandat, le Rapporteur spécial recommande vivement au Conseil des droits de l'homme d'appuyer l'examen et l'adoption, au sein des Nations Unies, d'un instrument juridique ayant deux objectifs principaux :

a) Fournir aux États Membres un ensemble de principes et de dispositions types qu'ils pourraient intégrer dans leur législation nationale pour consacrer et faire appliquer les principes fondamentaux du droit des droits de l'homme, et en particulier le droit à la vie privée, lorsqu'il s'agit de surveillance;

b) Proposer aux États Membres un certain nombre d'options qui pourraient contribuer à combler les lacunes du droit international, en particulier pour ce qui est du droit à la vie privée et de la surveillance dans le cyberespace.

6. Il est évident qu'un instrument juridique de ce type est nécessaire, mais sa portée et la forme qu'il prendrait ne sont pas encore bien définies. Par ailleurs, les recherches en cours et les consultations menées auprès des parties prenantes permettent de se

[1] A/70/53, sect. III, partie A, résolution 28/16, par. 4, al. c).

[2] Rapport du Rapporteur spécial sur le droit à la vie privée au Conseil des droits de l'homme, mars 2017 (version préliminaire non éditée, en anglais uniquement, consultable en ligne. Voir A/HRC/34/60).

faire une idée claire de la teneur du texte, mais il reste encore à déterminer quelle forme lui donner pour optimiser son efficacité.

7. Il est admis de longue date que deux des rares domaines où le droit à la vie privée ne peut pas être absolu sont la répression de la criminalité, qui inclut la détection et la prévention des infractions, les enquêtes et les poursuites, et la sécurité nationale. Toutefois, le fonctionnement des démocraties dépend de l'existence de contrôles et contrepoids garantissant que toute surveillance ait pour objet de protéger une société libre. L'obtention d'autorisations avant toute activité de surveillance et le contrôle desdites activités de surveillance sont des éléments essentiels des règles, des mesures de sauvegarde et des voies de recours dont une société démocratique a besoin pour protéger les libertés qui font d'elle ce qu'elle est.

8. Dans son rapport présenté au Conseil des droits de l'homme en mars 2017, le Rapporteur spécial a présenté ses conclusions intermédiaires relatives à un instrument juridique qui régirait la surveillance dans le cyberespace, en complément aux instruments de cyberdroit en vigueur, notamment la Convention sur la cybercriminalité adoptée en 2001 à Budapest par le Comité des ministres du Conseil de l'Europe. Un projet antérieur, mené avec le soutien de l'Union européenne et intitulé Projet Mapping (Gestion de solutions alternatives pour la protection de la vie privée, la propriété intellectuelle et la gouvernance d'Internet), explore les différentes options possibles concernant l'élaboration d'un tel instrument juridique. Un projet de texte a été soumis à la société civile et aux entreprises internationales pour examen et sera rendu public d'ici au printemps 2018.

9. Le processus est décrit de façon plus détaillée dans le document de référence V[3].

B. Lettres d'allégation

10. Certaines des lettres d'allégation envoyées par le Rapporteur spécial aux gouvernements au sujet de la surveillance seront prises en considération dans les rapports publiés par le Haut-Commissariat aux droits de l'homme sur les communications au titre des procédures spéciales.

C. Autres lettres : domaine public; Japon

11. Le 18 mai 2017, le Rapporteur spécial a publié une lettre adressée au Gouvernement japonais (voir document de référence III)[4]. Il y exprimait ses préoccupations relatives à un projet de législation qui autorise les activités de surveillance sans prévoir les mesures de sauvegarde nécessaires, officiellement afin de permettre au Japon de ratifier la Convention des Nations Unies contre la criminalité transnationale organisée adoptée en 2000. Les tentatives de dialogue à ce sujet se poursuivent et seront décrites dans le rapport que le Rapporteur spécial présentera au Conseil des droits de l'homme en mars 2018.

D. Autres initiatives en cours en matière de surveillance

12. Il existe d'autres initiatives ayant pour objet d'explorer les questions de surveillance, de sécurité et de confidentialité des données. Des détails à ce sujet seront rendus publics ultérieurement s'il y a lieu.

---

[3] Voir www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[4] Voir www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

E. Mieux comprendre le droit à la vie privée

13. Pour le Rapporteur spécial, le droit à la vie privée est, notamment, un droit essentiel nécessaire à l'exercice du droit fondamental général qu'a chacun au développement libre et sans entrave de sa personnalité. La Présidente de l'Équipe spéciale chargée du droit à la vie privée et de la personnalité, Elizabeth Coombs, ex-Commissaire au droit à la vie privée de l'État de Nouvelle-Galles du Sud (Australie), a bien voulu se pencher sur cette question, en se concentrant tout particulièrement sur la problématique hommes-femmes et le droit à la vie privée.

14. Vous trouverez plus de renseignements sur les activités de l'Équipe spéciale dans le document de référence IV[5].

F. Équipe spéciale chargée des données sur la santé

15. L'Équipe spéciale du Rapporteur spécial chargée des données sur la santé a débuté ses travaux sous la direction du docteur Steve Steffensen, des États-Unis d'Amérique. Elle devrait mener des consultations au cours du printemps et de l'été 2018.

G. Utilisation de données personnelles par les entreprises

16. Le Rapporteur spécial a continué de travailler à l'élaboration de modèles d'activité sur le droit à la vie privée dans le cadre de l'utilisation de données personnelles par les entreprises, de son côté et au sein du Projet Mapping, en vue de préparer la mise en place de son équipe spéciale chargée de la question. Le calendrier des activités prévues en la matière figure sur le site Web du Rapporteur spécial (http://www.ohchr.org/EN/Issues/Privacy/SR/Pages/ThematicReports.aspx).

H. Visites officielles de pays

17. Les visites de pays suivantes ont eu lieu ou sont prévues : États-Unis (du 19 au 28 juin 2017[6], France (visite confirmée du 13 au 17 novembre 2017); Royaume-Uni de Grande-Bretagne et d'Irlande du Nord (visite confirmée du 11 au 17 décembre 2017); Allemagne (visite confirmée du 29 janvier au 2 février 2018); République de Corée (visite confirmée du 3 au 15 juillet 2018).

I. Dotation en ressources

18. Seuls la visite officielle aux États-Unis et le voyage du Rapporteur spécial et d'autres orateurs à Hong Kong (Chine) pour la Conférence internationale des commissaires à la protection de la vie privée et des données personnelles et la série d'ateliers intitulée « Protection de la vie privée, personnalité et flux d'informations » ont été financés par le budget affecté au mandat du Rapporteur spécial, qui est géré par le Haut-Commissariat aux droits de l'homme. Les autres visites ont bénéficié de

[5] Voir www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[6] Le rapport final sur la visite officielle aux États-Unis devrait être publié au printemps 2018. La déclaration faite à l'issue de la mission peut être consultée à l'adresse suivante : http://www.ohchr.org/Documents/Issues/Privacy/SR_Privacy/VisitUSA_EndStatementJune2017.docx.

fonds de l'extérieur, provenant principalement des organisateurs des manifestations concernées.

II. Équipe spéciale en charge des mégadonnées et des données ouvertes

19. L'Équipe spéciale en charge des mégadonnées et des données ouvertes établie par le Rapporteur spécial est dirigée par David Watts[7]. Le présent rapport a principalement été élaboré par David Watts et Vanessa Teague[8] avec la contribution de nombreux autres membres de l'Équipe spéciale, dont la composition est la suivante : Christian d'Cunha (Contrôleur européen de la protection des données), Alex Hubbard (Commissariat à l'information du Royaume-Uni), le professeur Wolfgang Nejdl [Université de Hanovre (Allemagne)], Marty Abrams (Information Accountability Foundation, États-Unis) et Marie Georges (France). Sean McLaughlan, Elizabeth Coombs et Joe Cannataci ont également contribué à la rédaction du rapport.

20. Vous trouverez plus de renseignements sur le processus d'élaboration du rapport sur les mégadonnées et les données ouvertes dans le document de référence VII[9].

A. Définition des problèmes

21. Un des plus grands défis de la société de l'information en ce XXI^{e} siècle est de trouver un équilibre entre les avantages que la société peut retirer des nouvelles technologies de l'information et des communications d'une part, et la protection de droits fondamentaux tels que le droit à la vie privée d'autre part. Ces nouvelles technologies pourraient aider les États à garantir le respect et la protection des droits de l'homme, et à remplir leurs obligations à cet égard, mais elles risquent par ailleurs de faire obstacle à l'exercice de certains droits fondamentaux, notamment le droit à la vie privée.

22. Au vu de l'apparition de nouvelles méthodes de collecte et d'analyse des données – le phénomène des mégadonnées – et de la propension croissante des gouvernements du monde entier à divulguer les informations personnelles en leur possession, certes une fois anonymisées, en vue de favoriser la croissance économique et de stimuler la recherche scientifique – le phénomène des données ouvertes –, nombre des idées qui sous-tendent notre perception de la vie privée, de ce que cela suppose et du meilleur moyen de la protéger sont remises en cause.

23. En reconnaissant que le droit à la vie privée peut permettre l'exercice d'autres droits, et qu'il est notamment essentiel au droit à la dignité et au libre développement de la personnalité de chacun (voir la résolution 34/7 du Conseil des droits de l'homme en date du 23 mars 2017), le Conseil des droits de l'homme a étendu la portée du problème posé par les mégadonnées et les données ouvertes.

24. Certaines affirmations relatives aux mégadonnées et aux données ouvertes, qualifiées d'« utopiques »[10], consistent à soutenir que les mégadonnées apportent un

---

[7] David Watts est professeur de droit auxiliaire à l'Université La Trobe et à l'Université Deakin. Jusqu'au 31 août 2017, il était Commissaire au droit à la vie privée et à la protection des données de l'État de Victoria (Australie).

[8] Vanessa Teague est maître de conférences au département Informatique et systèmes d'informations de l'Université de Melbourne (Australie).

[9] Voir www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[10] Danah Boyd et Kate Crawford, « Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon », *Information, Communication and Society*, vol. 15, n^{o} 5.

nouvel éclairage sur des questions de politiques publiques extrêmement complexes telles que les changements climatiques, la menace terroriste et la santé publique. À l'extrême opposé, d'autres adoptent une perspective « dystopique », inquiets de la surveillance croissante exercée par les acteurs étatiques et non étatiques, des intrusions injustifiées dans la sphère privée et de l'effondrement des mesures de protection de la vie privée.

25. Une des plus grandes difficultés rencontrées pendant l'élaboration du présent rapport a été de trier et d'évaluer ces affirmations ainsi que toutes celles formulées par d'autres parties prenantes participant aux discussions complexes relatives aux mégadonnées et aux données ouvertes. Bien que ces deux questions aient fait l'objet de nombreux travaux de recherche et commentaires, nous ne comprenons pas encore complètement les technologies et leurs incidences pour l'avenir : paradoxalement, ce manque de données nous empêche également de comprendre les avantages et les dangers potentiels des mégadonnées et des données ouvertes.

B. Données

26. Chaque jour, nous générons par nos activités en ligne environ 2,5 quintillions d'octets de données[11] (2,5 suivi de 18 zéros[12]). Pour mettre ce chiffre en perspective, notons que le nombre d'octets de données d'un roman moyen de 300 pages est d'environ 3 suivi de cinq zéros, que 90 % de toutes les données du monde ont été générées ces deux dernières années[13], et que ce nombre augmente de façon exponentielle.

27. Dans notre monde connecté, les données sont omniprésentes. Des données sont produites à chaque fois que nous utilisons un ordinateur, un smartphone ou d'autres appareils courants contenant des capteurs capables d'enregistrer des informations. Elles prennent la forme de caractères ou de symboles que les appareils informatiques transforment en données binaires, lesquelles sont ensuite traitées, stockées et transmises sous forme de signaux électroniques.

28. Les mégadonnées proviennent de sources aussi diverses que les activités qui se déroulent en ligne :

> « Les données proviennent de nombreuses sources différentes, notamment d'instruments scientifiques, d'appareils médicaux, de télescopes, de microscopes ou de satellites; d'activités numériques telles que SMS, vidéos, messages audio, courriels, blogs, tweets, galeries d'images, flux de clics et transactions financières; de capteurs dynamiques et de réseaux sociaux ou autres; de simulations, modèles et enquêtes scientifiques; ou d'analyses informatiques de données observationnelles. Les données peuvent être d'ordre temporel, spatial ou dynamique, structurées ou non, et peuvent fournir des informations variées en termes de représentation, complexité, granularité, contexte, provenance, fiabilité ou domaine d'application. Elles peuvent en outre être générées et recueillies à des vitesses variables. »[14]

29. Certaines données ne se rapportent pas à des personnes. Elles sont issues d'activités telles que l'analyse des régimes météorologiques, l'exploration de

---

[11] Voir www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

[12] Ce mode de calcul est celui utilisé aux États-Unis. Au Royaume-Uni de Grande-Bretagne et d'Irlande du Nord, un quintillion est représenté par un 1 suivi de 30 zéros.

[13] Voir www-01.ibm.com/software/data/bigdata/what-is-big-data.html.

[14] États-Unis, National Science Foundation, « Critical techniques and technologies for advancing big data science and engineering (BIGDATA) », Program Solicitation NSF 14-543, p. 3, consultable à l'adresse suivante : www.nsf.gov/pubs/2014/nsf14543/nsf14543.pdf.

l'espace, les tests scientifiques de matériaux ou de projets, ou l'évaluation des risques associés aux opérations boursières sur les marchés financiers. Mais une grande partie des données est créée par nous ou à notre sujet. Le présent rapport porte sur cette catégorie de données – les données personnelles – qu'elles soient fournies, observées, dérivées ou déduites[15].

30. Les données personnelles reflètent l'individualité de chaque être humain. C'est cette capacité de définir chaque personne qui les rend si précieuses.

31. Les données que nous créons concernent notamment notre organisation. Elles comprennent nos courriels et nos SMS, ainsi que les images et les vidéos que nous créons et diffusons. D'autres données sont créées à notre sujet par des tiers mais avec notre participation, au moins dans une certaine mesure (dossiers médicaux électroniques ou achats en ligne par exemple).

32. Par contre, d'autres données sont générées à notre sujet sans qu'on en soit conscients, car cela se fait dans l'ombre, dans des circonstances obscures que, le plus souvent, nous ne connaissons pas – et ne pouvons pas connaître. Ce sont des « fils d'Ariane » (*digital breadcrumbs*)[16], c'est-à-dire des artéfacts numériques et autres traces laissées par nos activités en ligne et hors ligne. Ces données peuvent inclure les moments et les lieux où nos appareils mobiles se connectent aux antennes-relais de téléphonie mobile ou aux satellites du système de positionnement universel (GPS), les sites Web que nous consultons ou les images des systèmes de télévision en circuit fermé. Ces traces que nous laissons derrière nous et qui resteront probablement pour toujours sur des serveurs informatiques fournissent des indices sur ce que nous sommes, ce que nous faisons et ce que nous voulons. C'est ce qui rend les données personnelles – les données sur les personnes – incroyablement précieuses, tant pour le bien public que pour les entreprises privées[17].

33. Dans ce monde submergé par les données, l'informatique et les communications numériques instantanées, c'est à se demander comment les droits relatifs à la vie privée peuvent coexister avec les nouvelles technologies qui permettent de recueillir, traiter et analyser des données personnelles de manières inimaginables au moment de la rédaction de la Déclaration universelle des droits de l'homme (1948) et du Pacte international relatif aux droits civils et politiques (1966).

34. En conséquence de l'utilisation généralisée de l'informatique, presque chaque aspect du monde prend une dimension symbolique nouvelle à mesure que les événements, les objets, les processus et les personnes peuvent être présentés, connus et diffusés sous une autre forme. C'est en quelque sorte une nouvelle naissance pour le monde, car les données et les textes électroniques sont d'ampleur et de portée universelles[18].

35. Le procédé consistant à définir les personnes en analysant leurs données au moyen des technologies de l'information et des communications est basé sur l'idée qu'une

---

[15] Martin Abrams, « The origins of personal data and its implications for governance », consultable à l'adresse suivante : http://informationaccountability.org/wp-content/uploads/Data-Origins-Abrams.pdf.

[16] Evan Schwartz, « Finding our way with digital bread crumbs », *MIT Technology Review*, 18 août 2010, consultable à l'adresse suivante : www.technologyreview.com/s/420277/finding-our-way-with-digital-bread-crumbs/.

[17] Julie Lane *et al*., dir., *Privacy, Big Data, and the Public Good: Frameworks for Engagement,* New York, Cambridge University Press, 2014.

[18] Shoshana Zuboff, « Big other: surveillance capitalism and the prospects of an information civilization », *Journal of Information Technology*, vol. 30, nº 1, mars 2015.

personne est constituée par les informations qui la concernent[19]. Le phénomène qui rend ce procédé possible est généralement appelé « mégadonnées ».

C. Mégadonnées

36. Le terme « mégadonnées » est généralement utilisé pour décrire le volume important et toujours croissant de données et les techniques d'analyse sophistiquées utilisées pour les explorer, les analyser, établir des liens entre elles et en tirer des conclusions.

37. Il n'existe aucune définition établie des mégadonnées. D'après le National Institute of Standards and Technology des États-Unis, ce phénomène résulte de l'incapacité des architectures traditionnelles de données de gérer efficacement les nouveaux ensembles de données. Les mégadonnées ont des caractéristiques qui rendent obligatoire la création de nouvelles structures, à savoir :

a) Le volume (taille des ensembles de données);

b) La variété (données de différents répertoires, domaines ou types);

c) La vitesse (vitesse du flux de données);

d) La variabilité (variation d'autres caractéristiques).

38. Ces caractéristiques – volume, variété, vitesse et variabilité – sont couramment appelées les « V » des mégadonnées[20].

39. La description ci-dessus, élaborée par le National Institute, ainsi que beaucoup d'autres initiatives visant à définir le phénomène des mégadonnées, comme la déclaration de l'Union européenne selon laquelle les mégadonnées sont constituées de grandes quantités de données produites très rapidement par un grand nombre de sources diverses[21], attirent l'attention sur les technologies qui font de la collecte, du traitement et de l'analyse de grandes quantités de données une réalité courante. Cependant, compte tenu de leur caractère très général et du fait qu'elles mettent principalement l'accent sur les technologies, ces descriptions ne sont pas satisfaisantes face au phénomène des mégadonnées.

40. Plusieurs experts ont tenté de fournir une description plus complète des mégadonnées, qui ne se limite pas aux quatre « V ». D'après une description pertinente et plus détaillée, les mégadonnées ont les caractéristiques suivantes :

a) Un volume énorme, constitué de téraoctets ou de pétaoctets de données;

b) Une vitesse élevée, créées en temps réel, ou presque;

c) Une grande diversité, de nature structurée ou non;

d) Une portée exhaustive, visant à couvrir des populations ou des systèmes entiers;

e) Une granularité fine et une identification unique;

f) La capacité d'être liées à d'autres, avec des domaines communs permettant d'établir des liens entre différents ensembles de données;

---

[19] Luciano Floridi, « Four challenges for a theory of informational privacy », *Ethics and Information Technology*, vol. 8, nº 3, juillet 2006.

[20] D'autres caractéristiques leur sont attribuées, mais ce sont là les quatre principales. Voir http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1.pdf.

[21] Voir https://ec.europa.eu/digital-single-market/en/policies/big-data.

g) Un caractère flexible, rendant simple et rapide l'ajout de nouveaux domaines et les extensions[22].

41. Toutes les mégadonnées ne possèdent pas nécessairement chacune de ces caractéristiques.

42. D'autres approches consistent à présenter les mégadonnées comme un phénomène qui ne s'arrête pas à la technologie :

> « Nous définissons les mégadonnées comme un phénomène culturel, technologique et scientifique qui repose sur l'interconnexion des éléments suivants :
>
> a) La technologie, qui optimise la puissance informatique et la précision des algorithmes permettant de rassembler, analyser, lier et comparer de grands ensembles de données;
>
> b) L'analyse, qui, sur la base de grands ensembles de données, permet de définir des schémas et d'en tirer des conclusions sur les plans économique, social, technique et juridique;
>
> c) La mythologie, croyance très répandue selon laquelle de grands ensembles de données offrent une forme supérieure d'intelligence et de connaissance qui peut fournir un éclairage jusqu'alors impossible, ayant une apparence de vérité, d'objectivité et de précision. »[23]

43. Les partisans des mégadonnées affirment principalement qu'elles peuvent apporter une solution aux limites imposées à la recherche par le manque de preuves empiriques, c'est-à-dire le manque de données, et nous révéler la vérité objective au sujet de circonstances ou de phénomènes. Ces affirmations épistémologiques, qui tendent à faire des mégadonnées une nouvelle forme de méthode scientifique, sont au cœur du malaise que beaucoup ont exprimé au sujet des limites des mégadonnées et des risques qu'elles posent.

44. Il est généralement admis que la société peut retirer des avantages des mégadonnées, notamment des services personnalisés, un meilleur accès aux services, de meilleurs traitements médicaux, des progrès technologiques et une meilleure accessibilité[24]. La Commission européenne déclare que la nécessité de comprendre les mégadonnées entraîne des innovations technologiques, l'élaboration de nouveaux outils et l'acquisition de nouvelles compétences[25].

45. La Commission européenne voit l'information comme un atout économique, aussi important à la société que la main-d'œuvre et le capital[26]. Il est important de noter que ce marché est dominé par un petit nombre de géants de la technologie dont les parts de marché dépendent de l'utilisation des données.

---

[22] Rob Kitchin, « Big data, new epistemologies and paradigm shifts », *Big Data and Society*, vol. 1, nº 1, avril-juin 2014.

[23] Boyd et Crawford, « Critical questions for big data ».

[24] D'autres formulent des opinions contraires. Voir par exemple la déclaration du Groupe de travail, article 29 sur la protection des données de l'Union européenne, en date du 16 septembre 2014, sur l'impact des mégadonnées sur la protection des personnes pour ce qui est du traitement de leurs données personnelles dans l'Union européenne. Selon le Groupe de travail, les mégadonnées devraient avoir de nombreux avantages individuels et collectifs, mais leur valeur réelle reste à démontrer. Le Groupe de travail appuierait bien sûr tous les efforts, au niveau de l'Union européenne ou des États, visant réellement à concrétiser ces avantages pour les habitants de l'Union européenne, sur le plan individuel ou collectif. Voir http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf.

[25] Voir https://ec.europa.eu/digital-single-market/en/making-big-data-work-europe.

[26] Ibid.

D. Analyse sophistiquée

46. Le changement crucial est que les données sont utilisées de façon frénétique pour alimenter l'algorithme dont le comportement dépendra ensuite de ces mêmes données :

> « Le terme "apprentissage automatique" désigne le repérage automatique de schémas pertinents. En quelques décennies, cet outil est devenu essentiel dans presque toute tâche nécessitant l'extraction de données à partir de grands ensembles...
>
> Une caractéristique commune de toutes ces applications est que, contrairement à ce qui se passe lors d'utilisations plus traditionnelles de l'informatique, dans ces cas, compte tenu de la complexité des schémas qui doivent être repérés, un programmeur humain ne peut pas fournir d'instruction explicite et détaillée sur la façon dont ces tâches devraient être exécutées...
>
> Les outils d'apprentissage automatique consistent à doter les programmes de la capacité d'apprendre et de s'adapter. »[27]

47. Le principal aspect différenciant les techniques actuelles des techniques passées est leur nature autonome et semi-autonome.

48. Une des techniques d'analyse les plus couramment utilisées est appelée « exploration de données » (*data mining*). C'est un processus consistant à extraire des données à partir de grands ensembles pour ensuite les analyser en vue de repérer d'éventuels schémas ou liens. Avec l'exploration de données, il est devenu plus facile de simplifier de grandes quantités de données brutes[28] et d'en faire la synthèse, et de procéder à des déductions sur la base des schémas repérés.

49. Ces techniques et outils sont fondés sur des algorithmes.

E. Algorithmes

50. Les algorithmes n'ont rien de nouveau. Ils existent depuis l'aube des temps et étaient connus bien avant qu'un mot ne soit créé pour les désigner[29].

51. Les algorithmes ne se limitent pas qu'aux mathématiques. Les Babyloniens s'en servaient pour se prononcer sur des points de droit, les enseignants de latin les utilisent pour contrôler la grammaire d'un texte – et, dans toutes les cultures, ils ont été utilisés pour prédire l'avenir, décider d'un traitement médical ou faire la cuisine. Aujourd'hui, tout le monde se sert d'algorithmes d'une sorte ou d'une autre, souvent inconsciemment, en suivant une recette ou le patron d'un tricot, ou en se servant de gadgets courants[30].

52. Il est bien connu que les algorithmes, comme les autres éléments des mégadonnées, sont difficiles à définir avec précision[31]. Aux fins du présent rapport, on adoptera la définition suivante :

---

[27] Shai Shalev-Shwartz et Shai Ben-David, *Understanding Machine Learning,* New York, Cambridge University Press, 2014.

[28] Données ne portant que sur une personne.

[29] Jean-Luc Chabert, dir., *A History of Algorithms: From the Pebble to the Microchip,* Berlin, Springer-Verlag, Berlin, Heidelberg, 1999.

[30] Ibid.

[31] Felicitas Kraemer, Kees van Overveld and Martin Peterson, « Is there an ethics of algorithms? », *Ethics and Information Technology*, vol. 13, nº 3, septembre 2011.

> « Ensemble spécifique d'instructions suivies pour mener une procédure ou résoudre un problème, étant généralement entendu que la procédure doit prendre fin à un certain moment. Certains algorithmes sont également parfois appelés méthode, procédure ou technique... Le procédé consistant à appliquer un algorithme à une entrée pour obtenir une sortie s'appelle calcul. »[32]

53. Ce qui distingue l'algorithme utilisé pour faire un gâteau d'un algorithme utilisé pour évaluer la cote de crédit d'une personne est le degré d'automatisation, la nature autonome et non linéaire de l'algorithme et la quantité de données traitée.

54. De plus en plus, nous nous voyons, ainsi que notre relation au monde, au travers d'un algorithme. Les algorithmes sont aujourd'hui un élément crucial des sociétés de l'information, régissant de plus en plus d'opérations, de décisions et de choix qui étaient auparavant laissés aux humains[33] : recommandations sur les sites de rencontres[34], choix de la meilleure route à prendre[35] ou évaluation de la solvabilité d'une personne[36]. Ils sont utilisés pour le profilage – la définition de caractéristiques personnelles et de comportements systématiques en vue de pouvoir faire des prédictions personnalisées, notamment au sujet des biens ou des services que nous pourrions être tentés d'acheter. Ils déterminent comment les données devraient être interprétées et quelles mesures devraient être prises en conséquence. Ils font le lien entre les processus sociaux, les transactions commerciales, les décisions des gouvernements et la façon dont nous percevons et comprenons les autres et notre environnement et interagissons avec eux[37].

55. Pour les individus, les recommandations et les décisions qui résultent d'un calcul algorithmique paraissent sortir d'une boîte noire impénétrable et mystérieuse, une sorte d'oracle de Delphes du XXIe siècle qui semble rendre des décisions définitives et autoritaires, indépendantes de toute intervention humaine. Comprendre les mécanismes des procédés algorithmiques pour pouvoir ainsi évaluer les risques qu'ils posent est une entreprise complexe qui s'accompagne d'une multitude de questions. Cette complexité entrave notre capacité de comprendre le fonctionnement des algorithmes et leur influence sur notre vie.

56. De plus en plus de spécialistes s'intéressent aux problèmes que les algorithmes peuvent causer et préconisent de faire preuve de prudence au lieu de se précipiter dans un avenir régi par les algorithmes sans réfléchir aux mesures de précaution à prendre pour gérer les risques.

1. Les algorithmes sont fondés sur un système de valeurs

57. Malgré leur structure arithmétique, qui leur donne une apparence d'objectivité, les algorithmes sont inévitablement fondés sur un système de valeurs[38]. Ces valeurs reflètent souvent les préjugés culturels ou autres des concepteurs de logiciels, qui les intègrent dans la structure logique des algorithmes en tant qu'opinions implicites.

---

[32] Voir http://mathworld.wolfram.com/Algorithm.html.

[33] Brent Mittelstadt *et al*., « The ethics of algorithms: mapping the debate », *Big Data and Society*, vol. 3, nº 2, juillet-décembre 2016.

[34] Voir, par exemple, Rebecca Harrington, « Dating services tinker with the algorithms of love », *Scientific American*, 13 février 2015, consultable à l'adresse suivante : www.scientificamerican.com/article/dating-services-tinker-with-the-algorithms-of-love/.

[35] Voir https://motherboard.vice.com/en_us/article/4x3pp9/the-simple-elegant-algorithm-that-makes-google-maps-possible.

[36] Voir Michael Byrne, « The simple, elegant algorithm that makes Google Maps possible », 22 mars 2015, consultable à l'adresse suivante : http://mitsloan.mit.edu/media/Lo_ConsumerCreditRiskModels.pdf.

[37] Mittelstadt *et al*., « The ethics of algorithms ».

[38] Ibid.

58. Un algorithme utilisé pour calculer les cotes de crédit pourrait être conçu de telle manière qu'une personne doive indiquer son lieu de naissance, l'école qu'elle a fréquentée, son lieu de domicile et son statut professionnel. La sélection de ces facteurs d'approximation est fondée sur un jugement de valeur, à savoir que les réponses à ces questions permettent de déterminer si une personne peut recevoir un prêt et, si oui, à quelles conditions. Quel que soit le résultat, le demandeur de prêt n'a souvent aucun moyen de connaître la raison d'une décision et ne peut pas savoir quels jugements de valeur ont été appliqués.

59. Bien que ces facteurs d'approximation puissent être utiles pour décider de l'octroi de prêts dans certaines sociétés, ils seront au mieux des distractions inutiles ou, au pire, des dangers, dans d'autres. Par exemple, dans certains pays en développement, où une grande partie de la population n'a pas nécessairement de domicile fixe, a suivi un enseignement formel limité et travaille à son compte, l'application de ces facteurs empêcherait pour toujours ces personnes d'obtenir le moindre prêt.

60. D'autre part, les algorithmes qui analysent des formes non traditionnelles de données pourraient indiquer qu'une personne sans historique de crédit conventionnel serait néanmoins un bon risque – ce qui favoriserait le développement humain[39].

2. Le problème des données imparfaites

61. Les données sont la matière brute qui alimente les algorithmes, mais toutes ne sont pas exactes, suffisamment exhaustives, actuelles ou fiables[40]. La provenance de certaines données, par exemple les dossiers fiscaux, peut généralement être facilement établie, mais l'exactitude de ces données peut varier d'une autorité fiscale à l'autre, à l'intérieur d'un État et entre les États. D'autres données peuvent avoir été tirées de bases de données obsolètes qui n'ont jamais été bien nettoyées, de sources peu fiables, ou de bases dans lesquelles la saisie des données n'a pas été faite correctement et qui n'ont pas été correctement tenues.

62. Les algorithmes ayant pour fonction de traiter des données, ils sont limités, comme tous les outils de traitement de données, par le fait que le produit ne peut jamais être supérieur aux apports[41]. Le principe « telles entrées, telles sorties » s'applique.

3. Le choix des données

63. Le risque relatif au choix des données est semblable à celui évoqué au paragraphe 62 ci-dessus. Tout comme la qualité des données influe sur les résultats, la sélection de données inappropriées ou non pertinentes produit des résultats parfois peu fiables ou trompeurs.

64. Les traitements algorithmiques sont en grande partie fondés sur des déductions et sur l'établissement de liens entre des données apparemment disparates. Si les données utilisées ne sont pas les bonnes, toute recommandation ou décision sera biaisée.

---

[39] États-Unis, Commission fédérale du commerce, « Big data: a tool for inclusion or exclusion– understanding the issues », 2016, consultable à l'adresse suivante : www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[40] Par exemple, les intérêts de minorités qui ne sont pas suffisamment représentées dans un ensemble de données précis pourraient être desservis par les décisions et les prédictions découlant de ces informations.

[41] Mittelstadt *et al*., « The ethics of algorithms ».

4. Biais, discrimination et pérennisation de la discrimination

65. Bien que certains experts fassent une distinction entre les biais et la discrimination[42], les risques que posent ces derniers dans le contexte des mégadonnées sont suffisamment semblables pour justifier qu'on s'y intéresse en même temps.

66. Les algorithmes peuvent être utilisés à des fins de profilage, c'est-à-dire pour repérer des liens et faire des prédictions au sujet du comportement d'un groupe. Cependant, en présence de groupes (ou profiles) qui se transforment et que l'algorithme redéfinit constamment sur la base de l'apprentissage automatique, on constate ce qui suit :

> « Qu'elle soit dynamique ou statique, une personne est appréhendée en fonction de ses liens aux autres tels que définis par l'algorithme, et non sur la base de son comportement réel. Les choix des individus sont structurés en fonction d'informations relatives au groupe. Le profilage peut accidentellement créer une base de données d'expérience favorisant la discrimination. »[43]

67. Certains ont affirmé que les techniques d'analyse sophistiquées, telle que le profilage, accentuent la discrimination. Les activités de police, par exemple, sont menées sur la base de prédictions, elles-mêmes fondées sur des statistiques criminelles et des analyses algorithmiques visant à définir les quartiers « sensibles » et à en faire des priorités pour les forces de répression[44]. Comme la présence policière est renforcée dans ces quartiers, qui se situent souvent dans des zones socialement défavorisées plutôt que là où sont commis les crimes en col blanc, l'intensification des activités de police tend à concentrer les arrestations et les condamnations en un même endroit, ce qui conduit à un cercle vicieux, où l'identification régulière et renforcée des mêmes endroits sensibles expose les populations défavorisées à un risque plus élevé d'arrestation et de sanctions pénales.

68. La possibilité que certains gouvernements se servent de ces outils pour surveiller ou cibler certains groupes, ou leur porter atteinte d'une quelconque façon, a également fait naître des préoccupations[45].

5. Responsabilité et obligation de rendre compte

69. Le tort causé par le traitement algorithmique des données est largement imputable aux difficultés liées au traitement de grands jeux de données disparates et à la conception et à l'application des algorithmes mis en œuvre. Du fait de la multitude des variables en jeu, il est difficile de déterminer le responsable du dommage occasionné. Souvent, l'analyse des mégadonnées repose plus sur la découverte et la prospection que sur le test d'une hypothèse particulière. Par conséquent, il n'est pas facile de prédire (et, pour les individus, de définir) d'emblée la finalité de l'utilisation des données.

70. L'opacité des algorithmes n'est pas nécessairement « un donné ». Il est techniquement possible de conserver les données utilisées et le résultat de l'application de l'algorithme à chaque stade du traitement.

---

[42] Les biais sont considérés comme l'expression constante ou régulière d'une préférence, d'une valeur ou d'une croyance précise dans la prise de décisions. La discrimination est l'effet négatif disproportionné qui peut résulter de la prise de décisions fondées sur des algorithmes.

[43] Mittelstadt *et al*., « The ethics of algorithms ».

[44] Voir, par exemple, www.predpol.com/how-predictive-policing-works/.

[45] Lee Rainie et Janna Anderson, « Code-dependent: pros and cons of the algorithm age », Pew Research Center, 8 février 2017, consultable à l'adresse suivante : www.pewinternet.org/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/.

6. Enjeux relatifs à la protection de la vie privée

71. Les « Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel » de l'Organisation de coopération et de développement économiques (OCDE) ont été publiées en 1980[46]. Les huit principes énoncés dans les Lignes directrices, ainsi que les principes similaires définis dans la Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel (Convention pour la protection des données à caractère personnel) adoptée par le Conseil de l'Europe en 1981 et les Principes directeurs pour la réglementation des fichiers informatisés contenant des données à caractère personnel adoptés par l'Assemblée générale dans sa résolution 45/95 du 14 décembre 1990, ont inspiré les lois encadrant la protection des informations personnelles dans le monde entier.

72. Le principe fondateur que l'on retrouve aussi bien dans les Lignes directrices de l'OCDE que dans la Convention pour la protection des données, à savoir le « principe de la limitation en matière de collecte », pose que les données à caractère personnel doivent être collectées de manière licite et loyale et, quand il y a lieu, au su et avec le consentement de la personne concernée[47]. Le « principe de la limitation des finalités » prévoit que les données à caractère personnel ne doivent être collectées que pour des finalités déterminées, que le traitement ultérieur de ces données doit être compatible avec ces finalités et que les changements de finalité doivent être précisés[48]. Le « principe de la limitation de l'utilisation » dispose que les données à caractère personnel ne peuvent être divulguées pour des finalités incompatibles qu'avec le consentement de la personne concernée ou lorsqu'une règle de droit le permet[49]. Le « principe de la qualité des données » est remis en question par la collecte de grandes quantités de données et l'obligation de limiter le traitement aux données à caractère personnel adéquates, pertinentes et non excessives. Les Principes directeurs de 1990 pour la réglementation des fichiers informatisés contenant des données à caractère personnel posent le principe de la proportionnalité de la conservation des données à la finalité du traitement.

73. Les mégadonnées remettent en question ces principes tout en soulevant des questions éthiques et des dilemmes sociaux résultant de l'utilisation à mauvais escient des algorithmes. Plutôt que de résoudre des questions de politique publique, une telle utilisation comporte le risque de conséquences imprévues qui portent atteinte aux droits de l'homme, notamment à la protection contre toute forme de discrimination (femmes, personnes handicapées, etc.).

74. Parallèlement, on perçoit les signes d'un changement de mentalité dans la conception des algorithmes, qui se traduit par une amélioration des solutions algorithmiques de traitement des mégadonnées. On peut ainsi citer l'initiative lancée par l'Institute of Electrical and Eleconics Engineers Standards Association sur la prise en compte des considérations éthiques dans la conception des algorithmes[50].

---

[46] Organisation de coopération et de développement économiques (OCDE), « Lignes directrices régissant la protection de la vie privée et les flux transfrontières de données de caractère personnel ». Consultable à l'adresse : http://www.oecd.org/fr/sti/ieconomie/lignesdirectricesregissantlaprotectiondelaviepriveeetlesfluxtransfrontieresdedonneesdecaracterepersonnel.htm.

[47] Voir Principes de l'OCDE en matière de vie privée. Consultable à l'adresse : http://oecdprivacy.org/.

[48] Ibid.

[49] Ibid.

[50] Institute of Electrical and Electronics Engineers (IEEE), The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems, *Ethically Aligned Design : A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*,

75. S'agissant du respect de la vie privée, les instruments internationaux applicables étendent la portée du droit au respect de la vie privée au-delà du droit à la confidentialité des informations qui fait l'objet des Lignes directrices de l'OCDE et de la Convention pour la protection des données. Le droit au respect de la vie privée étant considéré comme un droit permettant l'exercice d'autres droits de l'homme mais aussi comme un droit étroitement lié aux notions de dignité humaine et de libre et plein développement de la personnalité (voir résolution 34/7 du Conseil des droits de l'homme), les enjeux que soulèvent les mégadonnées du point de vue de la vie privée s'étendent à toute une série de droits de l'homme. La tendance des mégadonnées à s'immiscer dans la vie des individus en révélant les moindres détails de leur soi virtuel à ceux qui collectent et analysent leurs données heurte de plein fouet le droit au respect de la vie privée et les principes consacrés pour protéger ce droit.

76. Les incidences réglementaires sont aussi profondes que les changements observables dans l'évolution des pratiques du secteur privé et des administrations publiques.

F. Données ouvertes

77. La notion de données ouvertes a gagné en popularité parallèlement à la complexification des outils d'analyse. L'objectif est d'encourager les secteurs public et privé à placer leurs données dans le domaine public en vue de renforcer la transparence et l'ouverture, en particulier celles de l'action publique.

78. Les données ouvertes peuvent se définir comme suit :

> « […] données pouvant être librement utilisées, réutilisées et redistribuées par quiconque – à la seule condition, au plus, d'en mentionner la source et de les partager à l'identique »[51].

79. Les données ouvertes peuvent relever de n'importe quelle catégorie de données. L'organisation Open Knowledge Foundation les récapitule comme suit :

a) Données culturelles : données sur les œuvres culturelles et les objets d'art – par exemple, les titres et les auteurs – généralement collectées et conservées par les galeries, les bibliothèques, les archives et les musées;

b) Données scientifiques : données produites dans le cadre de la recherche scientifique (de l'astronomie à la zoologie);

c) Données financières : données telles que les comptes des administrations publiques (dépenses et recettes) et les informations relatives aux marchés financiers (valeurs mobilières, actions, obligations, etc.);

d) Données statistiques : données produites par les bureaux de statistique (recensement, grands indicateurs socioéconomiques);

e) Données climatologiques : informations utilisées pour comprendre et prévoir les conditions météorologiques et climatiques;

f) Données environnementales : informations relatives à l'environnement naturel (présence et concentration des polluants, qualité de l'eau dans les cours d'eau et les mers, etc.)[52].

---

ver. 1, IEE Press, 2016. Consultable à l'adresse : http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf.

[51] Voir http://opendatahandbook.org/guide/en/what-is-open-data/.

[52] Voir https://okfn.org/opendata/.

80. Pour répondre aux critères de la définition, les données ouvertes sont souvent publiées sous licence Creative Commons. La licence CC-BY 4.0 autorise la copie, la redistribution et l'adaptation sans restriction (y compris à des fins commerciales) des informations sous licence à condition que la source en soit mentionnée[53].

81. Les données sur les citoyens que détiennent les autorités publiques ne se rangeraient dans aucune de ces catégories. À l'origine, l'ouverture des données publiques et de l'administration publique avait pour but de donner accès aux données sur l'administration et sur le monde. Elles n'avaient pas pour objet d'inclure les données que les pouvoirs publics recueillent sur les citoyens. C'est pourquoi, dans certains pays, les informations « à caractère personnel » et autres (notamment les données commerciales ou les données confidentielles) sont expressément exclues du champ des données ouvertes[54]. Il ne faut pas perdre de vue qu'avec l'apparition des termes tels que « partage » et « connexion », un renversement s'est opéré, à savoir qu'au lieu de publier des données sur le fonctionnement de l'administration pour permettre aux citoyens de contrôler son action, les autorités rendent publiques des données sur les citoyens.

G. Administration ouverte

82. L'une des premières mesures engagées par l'administration Obama a été de prendre un décret pour encourager la publication d'informations publiques afin de renforcer la confiance des citoyens et de promouvoir la transparence, la participation et la collaboration[55].

83. Par la suite a été créé le Partenariat pour le gouvernement ouvert. En septembre 2011, ce dernier a publié la Déclaration du gouvernement ouvert[56]. La Déclaration traduit la volonté de mieux informer les citoyens sur l'action des pouvoirs publics et souligne la nécessité de renforcer la participation des citoyens, la transparence de l'action publique, la lutte contre la corruption, l'autonomisation des citoyens et l'exploitation du potentiel des nouvelles technologies pour renforcer l'efficacité et la responsabilité des pouvoirs publics.

84. Par la Déclaration, les signataires se sont engagés, de manière volontaire et non contraignante :

a) À accroître la disponibilité des informations sur l'action des pouvoirs publics;

b) À promouvoir la participation civique;

c) À faire appliquer par les administrations les normes les plus strictes d'intégrité professionnelle;

d) À intensifier l'accès aux nouvelles technologies à des fins de transparence et de responsabilité[57].

85. Le premier décret présidentiel pris par l'administration Obama a été suivi d'un autre, adopté le 9 mai 2013, pour ouvrir et rendre exploitables informatiquement par

---

[53] Voir https://creativecommons.org/licenses/by/4.0/.

[54] Australie, gouvernement de Nouvelle-Galles du Sud, Department of Finance and Services, « Open Data Policy », 2013.

[55] Président Obama, « Transparency and Open Government », 21 janvier 2009, mémorandum à l'intention des chefs de ministères et d'agences de l'État. Consultable à l'adresse : https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

[56] Voir https://www.opengovpartnership.org/open-government-declaration.

[57] https://www.opengovpartnership.org/open-government-declaration.

défaut toutes les données produites par l'administration fédérale américaine[58]. Ainsi, un changement d'orientation est intervenu entre les deux textes. Aux termes du second décret, l'ouverture des données publiques favorise la prestation efficiente et efficace de services au public et contribue à la croissance économique. Avantage essentiel d'une administration ouverte, rendre les informations faciles à trouver, accessibles et utilisables peut encourager l'esprit d'entreprise, l'innovation et les découvertes scientifiques qui améliorent la vie des Américains et contribuent fortement à la création d'emplois[59].

86. Au cours des années qui ont suivi, les données ouvertes ont tellement évolué qu'en 2017, l'ambition est de rendre publiques, non plus seulement des données qui ne sont pas ou n'ont jamais été dérivées d'informations personnelles, mais également des données à caractère personnel anonymisées. Les partisans de cette évolution affirment que les bases de données publiques et autres archives renferment des informations d'une grande « valeur » et que la publication de ces données favorisera la recherche et stimulera la croissance de l'économie de l'information.

87. Les données ouvertes tirées des informations à caractère personnel reposent ainsi entièrement sur l'efficacité des techniques de « désidentification » pour empêcher la « réidentification » des données et, partant, l'association entre les données et la personne concernée. La question de savoir si la désidentification permet de protéger la vie privée et de fournir des données « utiles pour la recherche » suscite des débats très vifs.

H. Complexité des mégadonnées

88. En 2015, le journaliste australien Will Ockenden a mis en ligne ses métadonnées téléphoniques et demandé aux internautes de lui dire ce qu'ils pouvaient en déduire sur sa vie. Ces métadonnées indiquaient l'heure exacte de tous ses appels et minimessages ainsi que l'antenne-relais la plus proche. Même s'il a remplacé les numéros de téléphone par des pseudonymes, les internautes ont pu répondre facilement et correctement à des questions comme « où vit ma mère » à partir de ses habitudes de communication et de déplacement. La réponse n'était pas bien compliquée à trouver. Les internautes ont simplement deviné que la mère du journaliste vivait à l'endroit où il s'est rendu le jour de Noël.

89. C'est là un thème central des travaux de recherche sur la vie privée : l'idée que les éléments récurrents dans les données, dépourvues de noms, de numéros de téléphone ou d'autres éléments d'identification évidents, peuvent être utilisés pour identifier une personne et, partant, pour extraire des informations supplémentaires la concernant. Ce mode d'utilisation est particulièrement efficace lorsque ces éléments récurrents peuvent être exploités pour relier de nombreux jeux de données différents et ainsi dresser un portrait complexe d'une personne.

90. Certaines données doivent inévitablement être révélées. Les opérateurs de téléphonie connaissent les numéros composés par leurs clients et les médecins sont informés des résultats d'analyse de leurs patients. Ce sont la divulgation de ces données à des tiers (entreprises, chercheurs, etc.), l'utilisation faite par les autorités de ces informations et l'incidence de ces pratiques sur l'exercice des droits de l'homme qui prêtent à controverse.

---

[58] Président Obama, décret du 9 mai 2013 sur l'ouverture et l'exploitation informatique par défaut des données de l'administration fédérale. Consultable à l'adresse : https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-.

[59] Ibid.

91. D'autres données sont délibérément récoltées, souvent à l'insu ou sans le consentement des intéressés. Des chercheurs de l'organisation Electronic Frontier Foundation ont publié les résultats de « Panopticlick », une expérience qui a montré qu'il était possible de prendre les « empreintes digitales » du navigateur Web d'un internaute à partir de caractéristiques simples, telles que les modules d'extension et les polices de caractère utilisés[60]. Ils ont mis en garde contre le fait que la confidentialité de la navigation sur Internet risquait d'être compromise si des limites n'étaient pas fixées au stockage de ces empreintes digitales et de leur association avec l'historique de navigation. Aucun changement de politique important n'est intervenu depuis. Aujourd'hui, en 2017, la confidentialité de la navigation sur le Web a disparu. De nombreuses sociétés pistent systématiquement et délibérément les internautes, et ce, généralement à des fins commerciales. Le pistage des utilisateurs sur le Web est désormais quasi omniprésent et il n'est possible d'y échapper qu'au prix d'un grand effort.

92. Une grande partie de l'économie moderne d'Internet repose sur la collecte de données complexes sur des clients potentiels dans la perspective de leur vendre des produits, pratique connue sous le nom de « capitalisme de surveillance »[61]. Toutefois, la surveillance ne paraît pas plus justifiée pour l'économie des données que le travail des enfants ne l'est pour l'économie industrielle. C'est seulement le moyen le plus commode et le plus facile d'exploiter l'information. Il ne s'agit nullement d'un droit fondamental comme le droit au respect de la vie privée. En effet, l'économie axée sur les données pourrait survivre et prospérer si des normes minimales et des technologies améliorées contraignaient les entreprises et les administrations publiques à opérer dans un monde où les citoyens ordinaires exercent un contrôle bien plus grand sur leurs données personnelles[62].

93. Les pouvoirs publics pourraient également innover avec une licence plus légitime. Le degré de confiance des citoyens dans les pouvoirs publics conditionne fortement leur perception de l'incidence des initiatives d'ouverture des données et de l'administration. Ceux qui ont confiance dans les autorités sont beaucoup plus enclins à voir un avantage dans les données ouvertes[63]. Les recherches indiquent que, si les gens sont globalement ouverts à l'idée que les pouvoirs publics publient des informations en ligne sur leurs collectivités, ils se montrent plus réservés lorsque ces données les concernent plus directement. Le degré d'aise des citoyens varie en fonction de la nature des données collectées[64].

94. La plupart des lois encadrant la protection des informations personnelles régissent la collecte et le traitement des données à caractère personnel. Autrement dit, les informations qui n'ont pas un « caractère personnel » ne sont pas visées par ces textes. Nombre de ces lois prévoient que les informations personnelles peuvent être « désidentifiées » pour que les données puissent être utilisées ou traitées aux fins de l'intérêt public d'une manière qui ne porte pas atteinte au droit à la protection des

---

[60] Peter Eckersley, « How unique is your web browser? », *in* Mikhail Atallah et Nicholas Hopper (dir.), *Privacy Enhancing Technologies*, Berlin, Springer-Verlag, 2010.

[61] Shoshana Zuboff, « Big other : surveillance capitalism and the prospects of an information civilization », *Journal of Information Technology*, vol. 30, nº 1, mars 2015.

[62] Il n'est pas forcément nécessaire de contraindre les entreprises et les administrations publiques à assurer la protection de la vie privée. Pour des exemples de démarches éthiques adoptées par les entreprises, voir Information Commissioner's Office, « Big data, artificial intelligence, machine learning and data protection », ver. 2.2 (2017). Consultable à l'adresse : https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

[63] John Horrigan et Lee Rainie, « Americans' views on open Government data », Pew Research Center, 21 avril 2015.

[64] Ibid.

données personnelles. Les autorités et autres acteurs ont cherché à conserver la confiance de ceux dont ils recueillent les données en offrant des assurances de désidentification.

95. Se pose alors l'importante question suivante : « Les procédés de désidentification permettent-ils de mettre à disposition des données qui ne portent pas atteinte au droit des individus à la protection des données personnelles? »

96. Les types de données simples, comme les statistiques agrégées, se prêtent à un traitement véritablement respectueux de la vie privée comme la « confidentialité différentielle ». Les algorithmes de confidentialité différentielle fonctionnent mieux à grande échelle et sont actuellement intégrés à l'analyse des données commerciales. Les algorithmes randomisés qui permettent la confidentialité différentielle sont un outil précieux dans l'arsenal de protection de la vie privée, mais ils ne permettent pas la désidentification générale de jeux très complexes de données unitaires sur les individus[65]. L'utilisation par la société Apple de ces techniques en 2016 est un exemple d'application à grande échelle des techniques de confidentialité différentielle[66].

97. Il n'est pas possible de « désidentifier » les données unitaires de grande dimension en toute sécurité sans en réduire considérablement l'utilité. C'est le type de données produites par l'historique longitudinal des données d'une personne en matière de santé, de mobilité, de recherche sur le Web, etc. On trouvera dans le document de référence I[67] un récapitulatif des outils de désidentification et des débats suscités par ces techniques.

Données publiques ouvertes

98. Il existe de nombreux cas où des personnes ont pu être « réidentifiées » dans des données publiées par les pouvoirs publics[68]. Cette « réidentification publique » est publique à double titre : d'une part, les résultats sont rendus publics; d'autre part, la réidentification se fait au moyen d'informations auxiliaires publiques.

99. Plus les informations auxiliaires sont nombreuses, plus il est facile de réidentifier un grand nombre de personnes. À mesure que le nombre des ensembles de données reliés augmente, le nombre d'informations auxiliaires nécessaires pour la réidentification diminue. La publication et la mise en relation des jeux de données ont pour effet de réunir au même endroit un grand nombre d'informations auxiliaires sur

---

[65] Données ne se rapportant qu'à une seule personne.

[66] Andy Greeberg, « Apple's "differential privacy" is about collecting your data – but not your data », 13 juin 2016. Consultable à l'adresse : https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/https://techcrunch.com/2016/06/14/differential-privacy/ https://arxiv.org/abs/1709.02753.

[67] Voir www.ohchr.org/Documents/Issues/Privacy/A-72-Slot-43103.docx.

[68] Lors d'une audition devant le Comité consultatif sur la vie privée et l'intégrité du Département de la sécurité intérieure, le 15 juin 2005, L. Sweeney a déclaré que c'était en 1997 qu'elle avait pu montrer comment le dossier médical de William Weld, alors Gouverneur du Massachusetts, pouvait être réidentifié uniquement grâce à sa date de naissance, son sexe et son code postal. En fait, 87 % de la population américaine est identifiable par la date de naissance (mois, jour et année), le sexe et le code postal. Le fait est que les données qui peuvent paraître anonymes ne le sont pas nécessairement. Voir www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_06-2005_testimony_sweeney.pdf; voir également Latanya Sweeney, « Matching known patients to health records in Washington State data », Harvard University, 2012. Consultable à l'adresse : http://dataprivacylab.org/projects/wa/1089-1.pdf et http://dataprivacylab.org/index.html; Latanya Sweeney, « Achieving k-anonymity privacy protection using generalization and suppression », *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, nº 5, 2002.

les individus, ce qui rend qui d'autant plus facile la réidentification de toute donnée les concernant.

100. La « réidentifiabilité » des données ouvertes n'est que la partie émergée de l'iceberg : la réidentifiabilité d'ensembles de données commerciales désidentifiées, qui sont régulièrement vendues, partagées et échangés.

101. Les forces qui se dressent face au droit au respect de la vie privée à l'ère des mégadonnées et des données ouvertes sont puissantes. Il est fort probable que le plus faible procédé de désidentification permis sera privilégié pour des raisons financières par tous ceux qui traitent les données, que ce soit à des fins commerciales ou autres, et les pouvoirs publics sont pressés non seulement d'ouvrir l'accès aux données sur les individus mais également de ne pas réglementer cet accès.

102. Les organisations non gouvernementales se sont inquiétées de l'essor de l'utilisation des mégadonnées sans que soient dûment prises en compte la participation des intéressés, les considérations éthiques et juridiques découlant de la mauvaise gestion des informations à caractère personnel ou la nécessité d'une réglementation adéquate[69]. Ces organisations continueront à militer pour qu'une protection adaptée soit assurée et que des mesures suffisantes soient prises.

I. État des lieux du présent : mégadonnées commerciales et vie privée

103. La croissance exponentielle de la collecte des données et le mouvement accéléré tendant à connecter tous les objets à Internet sans qu'il soit suffisamment tenu compte de la sécurité des données sont une source de risques pour les individus et pour les groupes. Dans le souci d'assurer les consommateurs et les individus de la sécurité des informations les identifiant, un certain nombre d'idées ont été distillées dans les esprits. Ainsi, par exemple, la notion de données « anonymisées » extrêmement complexes est alimentée par un secteur qui profite du sentiment d'anonymat illusoire des utilisateurs[70].

104. Un grand nombre de données sont recueillies auprès des utilisateurs ordinaires à leur insu ou sans leur consentement. Ces données peuvent être vendues et reliées à des données provenant d'autres sources pour dresser un tableau complexe de nombreux aspects de la vie d'une personne. Ces informations sont utilisées à de nombreuses fins, y compris dans un but de contrôle politique, comme l'a montré la divulgation involontaire d'un jeu de données par une organisation politique américaine[71]. Cet ensemble de données comprenait des renseignements personnels sur près de 200 millions d'électeurs américains, ainsi qu'un nombre étonnant de détails recueillis (ou devinés) sur leurs convictions politiques. En Chine, un projet de « crédit social » a été envisagé non seulement pour noter la solvabilité financière des citoyens mais également pour évaluer leur comportement social, voire politique. Ce système repose sur des données provenant de diverses sources, principalement électroniques, recueillies au fil du temps[72].

---

[69] Voir www.privacyinternational.org/node/8.

[70] L'anonymisation ne fait pas disparaître la pertinence des principes relatifs à la protection de la vie privée et des considérations telles que le « consentement ».

[71] Sam Biddle, « Republican data-mining firm exposed personal information for virtually every American voter », *The Intercept*, 20 juin 2017. Consultable à l'adresse : https://theintercept.com/2017/06/19/republican-data-mining-firm-exposed-personal-information-for-virtually-every-american-voter/.

[72] « China invents the digital totalitarian state », *The Economist*, 17 décembre 2016. Consultable à l'adresse : https://www.economist.com/news/briefing/21711902-worrying-implications-its-social-

105. Les courtiers de données – entreprises qui collectent des informations personnelles sur les consommateurs pour les partager ou les revendre à des tiers – sont des acteurs importants de l'économie des mégadonnées. Lors de l'élaboration de leurs produits, ces courtiers recueillent un grand éventail d'informations détaillées et précises sur les consommateurs auprès de sources diverses[73], puis les analysent pour tirer des conclusions – dont certaines peuvent être sensibles – avant de les partager avec des clients de divers secteurs. Toute cette activité se déroule à l'insu des consommateurs[74].

106. Si les produits des courtiers de données contribuent à prévenir la fraude, à améliorer les offres de produits et à fournir des services personnalisés, nombre des usages pour lesquels les courtiers recueillent et utilisent des données présentent des risques pour les consommateurs. Le manque de transparence, la collecte de données sur les jeunes, la conservation indéfinie des données et l'utilisation de ces données à des fins de sélection ou à des fins discriminatoires ou illicites sont autant de sujets de préoccupation[75].

107. Dans le récent projet de rapport du Parlement européen sur la réglementation européenne en matière de respect de la vie privée, il est recommandé ce qui suit : « Les utilisateurs finaux devraient disposer d'un éventail de réglages de confidentialité, depuis les plus restrictifs (par exemple, "ne jamais accepter les cookies") jusqu'aux plus permissifs (par exemple, "toujours accepter les cookies"), en passant par des options intermédiaires (par exemple, "rejeter les cookies de tiers" ou "accepter uniquement les cookies propres"). »[76]

108. La nécessité de renforcer le contrôle exercé par les individus sur la confidentialité de leur activité en ligne suscite aujourd'hui de vastes débats. Les individus utilisent leurs appareils et leurs données pour obtenir les informations dont ils ont besoin (cartes, indications) et pour consulter les publicités qui les intéressent. À cet égard, il est essentiel de se poser la question suivante : si les technologies facilitant le contrôle des utilisateurs finaux sont importantes, dans quelle mesure les individus peuvent-ils exercer un contrôle suffisamment exhaustif? L'adoption de ces instruments va à l'encontre des forces économiques actuellement à l'œuvre sur Internet[77]. Les pouvoirs publics ont-ils un rôle à jouer dans l'élaboration et l'adoption de ces outils?

---

credit-project-china-invents-digital-totalitarian; Lucy Hornby, « China changes tack on "social credit" scheme plan », *Financial Times*, 4 juillet 2017. Consultable à l'adresse : https://www.ft.com/content/f772a9ce-60c4-11e7-91a7-502f7ee26895.

[73] Les exemples sont nombreux d'acquisition de données commerciales à grande échelle à partir de dispositifs tels que les télévisions, les « appareils intimes », les jouets et les applications de covoiturage dans les « voitures connectées ».

[74] Sénat des États-Unis, Commission du commerce, des sciences et des transports, « A review of the data broker industry: collection, use, and sale of consumer data for marketing purposes », rapport, 18 décembre 2013. Consultable à l'adresse : http://educationnewyork.com/files/rockefeller_databroker.pdf.

[75] Commission fédérale du commerce des États-Unis, « Data Brokers: a call for transparency and accountability », mai 2014. Consultable à l'adresse : https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

[76] Marju Lauristin, « Projet de rapport sur la proposition de règlement du Parlement européen et du Conseil concernant le respect de la vie privée et la protection des données à caractère personnel dans les communications électroniques et abrogeant la directive 2002/58/CE », Parlement européen, Commission des libertés civiles, de la justice et des affaires intérieures, 2017.

[77] Par exemple, le module AdNauseam neutralise le pistage en cliquant sur toutes les publicités présentées à l'utilisateur afin de masquer celles que ce dernier consulte véritablement. Cette application a été bloquée par le navigateur Google Chrome. D'autres sites détectent et bloquent les individus qui les consultent en utilisant des dispositifs de blocage de la publicité. Voir Daniel

Technologies permettant de contrôler la collecte de données

109. Le contrôle (y compris le blocage) de la collecte des données est utile pour protéger les données que l'utilisateur ne souhaite pas partager. Avec les « anciennes » technologies, la question ne se posait pas. En effet, l'utilisateur exerçait inévitablement un contrôle, car les technologies dépendaient entièrement de sa volonté. Ainsi, les appareils photographiques étaient dotés de caches physiques et étaient raccordés à Internet par un câble Ethernet qui pouvait être débranché manuellement. Désormais, ces dispositifs disposent de connexions Wi-Fi internes et n'ont pas de cache. Les téléviseurs disposent de microphones qui ne peuvent pas être éteints. Si les fonctionnalités de désactivation manuelle ont disparu, il existe toutefois des technologies permettant de bloquer la collecte de données[78]. Grâce au succès de la campagne « TLS Everywhere », la majorité du trafic Internet est désormais crypté et beaucoup moins susceptible d'être intercepté en cours de route par une entité inconnue de l'utilisateur. Ces technologies ont des avantages qui doivent être examinés plus avant et encouragés.

110. L'idée de masquer son identité et son activité n'est pas neuve non plus. Il suffit de regarder la bataille opposant certains réseaux sociaux qui obligent à utiliser un « vrai nom » à ceux qui défendent le droit de créer un compte sous un pseudonyme. Le masquage suppose des outils qui permettent aux utilisateurs de présenter un profil « réservé » et de le séparer de leurs autres profils.

111. Toutes les recherches montrent que, lorsque les individus ont des réserves quant aux pratiques en matière de protection des informations personnelles des organisations auxquelles ils ont affaire, ils sont plus enclins à fournir des informations inexactes ou incomplètes[79]. Parce qu'elle produit de la confiance, la protection de la vie privée et des données personnelles a une incidence positive sur la qualité des données et, partant, sur les analyses qui en sont faites. La confiance des utilisateurs est également importante pour la stabilité et l'exactitude des algorithmes d'apprentissage automatique. L'apprentissage automatique classique peut être très vulnérable aux entrées délibérément confuses et fabriquées[80]. Que se passerait-il si un grand nombre de personnes adoptaient délibérément des outils permettant de masquer leur identité en raison de leurs inquiétudes en matière de confidentialité?

112. Une conception simpliste des mégadonnées et des données ouvertes qui ne tient pas compte de l'interaction complexe entre les pratiques des entreprises en matière de gestion des données personnelles, le sentiment que la protection de la vie privée est assurée et les comportements des individus ne favorisera pas les « mégadonnées » mais conduira au contraire à une prise de décision inexacte et de mauvaise qualité.

Howe et Helen Nissenbaum, « Engineering privacy and protest: a case study of AdNauseam ». Consultable à l'adresse : https://adnauseam.io/.

[78] Le routeur du réseau anonyme TOR masque qui communique avec qui (c'est-à-dire les métadonnées), mais son utilisation n'est pas très répandue. Certains navigateurs (comme Firefox et Brave) propose un mode de « navigation privée » qui empêche la collecte des données. Les modules « Privacy Badger » de l'Electronic Frontier Foundation et « TrackMeNot » de la New York University sont très efficaces mais ne sont pas beaucoup utilisés.

[79] Office of the Australian Information Commissioner, Attitudes des Australiens face aux enquêtes sur le respect de la vie privée, mai 2017 et octobre 2013; Deloitte, « Trust starts from within : Deloitte Australian privacy index 2017 », 2017.

[80] Ian Goodfellow, Johnathon Shlens et Christian Szegedy, « Explaining and harnessing adversarial examples », ArXiv preprint, 2014.

J. Principes pour l'avenir : le contrôle de la communication des données

113. Le droit relatif à la protection de la vie privée tend à reposer sur des principes offrant une souplesse suffisante face à l'évolution des risques d'atteinte. Il serait intéressant de voir si des principes supplémentaires sont nécessaires pour compléter les principes existants afin de protéger les données à caractère personnel contre les atteintes à la vie privée permises par les technologies.

114. Certains proposent d'encadrer le partage des données par les sept principes suivants[81] :

1. Exécuter les algorithmes à l'endroit où se trouvent les données : partager les résultats et non directement les données;

2. Ouvrir les algorithmes : assurer l'examen ouvert et le contrôle public de tous les algorithmes utilisés pour le partage des données et la protection de la vie privée afin de permettre la détection et la correction des erreurs ou des failles;

3. Respecter les usages permis : respecter les autorisations (expresses ou implicites) d'utilisation des données ou l'« intégrité contextuelle »[82]. Dans le domaine médical, l'octroi et le retrait exprès du consentement ont été mis en pratique au moyen de l'interface de consentement dynamique[83];

4. Toujours donner des « réponses sûres » : mettre la confidentialité différentielle en pratique;

5. Assurer le cryptage permanent des données : veiller à ce que les données ne puissent être lues que par ceux qui possèdent la clef de déchiffrement correspondante[84];

6. Prévoir des environnements de collaboration en réseau et des chaînes de blocs pour l'audit et la responsabilité;

7. Instaurer des incitations sociales et économiques.

115. Ces principes ne sont pas nécessairement des solutions complètes en elles-mêmes dans la mesure où ils soulèvent à leur tour de nouvelles questions. Ainsi, par

---

[81] Alex Pentland *et al*., « Towards an Internet of trusted data: a new framework for identity and data sharing », 2016.

[82] Le respect de la vie privée se définit comme « le principe selon lequel les informations concernant les personnes ("informations à caractère personnel") doivent circuler à bon escient, c'est-à-dire conformément aux normes en matière d'information. […] Les contextes sociaux forment l'arrière-plan de cette conception de la vie privée ». Voir Solon Barocas et Helen Nissenbaum, « Big data's end run around anonymity and consent », in Julian Lane *et al*. (dir.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press, 2014.

[83] Jane Kaye *et al*., « Dynamic consent: a patient interface for twenty-first century research networks », *European Journal of Human Genetics*, vol. 23, nº 2, 2014.

[84] Les progrès récents de la cryptographie permettent à de multiples parties de calculer ensemble une fonction sur leurs données privées et de ne révéler ensuite que les résultats bien délimités. Il existe des outils très généraux reposant sur le calcul multipartite (voir, par exemple, Ivan Damgård *et al*., « Multiparty computation from somewhat homomorphic encryption », *Advances in Cryptology – CRYPTO*, vol. 7417, 2012) et sur le cryptage homomorphique (voir www.microsoft.com/en-us/research/project/homomorphic-encryption/#). La plupart de ces outils ne sont actuellement pas suffisamment rapides pour traiter de grands ensembles de données, mais des variantes simplifiées le pourront peut-être à l'avenir. Il existe de nombreux protocoles spécifiques qui permettent de résoudre des problèmes précis sur de grands ensembles de données. Le calcul sur des données cryptées fonctionne très bien pour des calculs simples portant sur un seul ensemble de données mais peut s'avérer impossible pour des calculs complexes ou des jeux de données répartis en plusieurs endroits.

exemple, la transparence est particulièrement problématique lorsque les techniques utilisées pour garantir le respect de la vie privée sont si complexes que seule un petit nombre de personnes sont en mesure de les comprendre. Le principe de l'ouverture des algorithmes est un premier pas essentiel, mais les algorithmes précis utilisés et leurs incidences continueront de poser des problèmes dans la pratique.

116. D'autres « principes » ont été proposés, comme ceux du « pouvoir d'agir » et de la « transparence », le « pouvoir d'agir » comprenant notamment le droit de modifier les données, de brouiller les données, d'expérimenter avec les raffineries de données[85]. L'idée sous-jacente est d'autonomiser l'individu et d'égaliser le rapport de force entre les entreprises/détenteurs de données et les utilisateurs. D'autres proposent de donner à l'utilisateur la faculté de compliquer et d'empêcher la collecte des données ou de refuser de participer à une telle activité.

117. Dans l'ensemble, les principes de transparence et de maîtrise de l'utilisateur sont importants pour permettre aux usagers de choisir les données qu'ils souhaitent révéler sans avoir à subir de perte déraisonnable de services.

118. Surtout, les tentatives visant à poser des principes en matière de mégadonnées et de données ouvertes qui respectent la vie privée constituent un bon point de départ pour la discussion. Quels que soient les principes retenus, toutes les parties prenantes, y compris les organisations de la société civile, devraient être consultées pour en garantir la bonne adéquation.

119. La mise en œuvre de ces principes soulève la question du rôle des pouvoirs publics et du type d'incitations et de réglementation susceptibles de favoriser la protection de la vie privée et des autres droits de l'homme ainsi que l'évaluation de leurs effets relatifs sur les valeurs éthiques et politiques que sont la loyauté, la justice, la liberté, l'autonomie, le bien-être et autres plus spécifiques au contexte en question[86].

120. Une économie de l'information innovante trouverait sans doute un plus grand soutien dans la population si les pouvoirs publics et les entreprises adhéraient ouvertement à une réglementation solide en matière d'acquisition, de partage et de contrôle des données des citoyens.

III. Documentation de référence

121. Les documents de référence suivants, qui ont servi à l'élaboration du présent rapport, sont consultables sur le site Web du Rapporteur spécial[87] :

I. Comprendre l'histoire : procédés de désidentification et controverses;

II. Démarches du Rapporteur spécial en Afrique, en Amérique, en Asie et en Europe;

III. Contexte de la lettre ouverte adressée au Gouvernement du Japon;

IV. Activités de l'équipe spéciale sur la protection de la vie privée et la personnalité;

---

[85] Andreas Weigend, *Data for the people: How to Make our Post-Privacy Economy Work for You*, New York, Basic Books, 2017.

[86] Solon Barocas et Helen Nissenbaum, « Big data's end run around anonymity and consent », in Julia Lane *et al*. (dir.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement*, Cambridge University Press, 2014.

[87] Voir www.ohchr.org/EN/Issues/Privacy/SR/Pages/SRPrivacyIndex.aspx; voir également www.ohchr.org/Documents/Issues/Privacy/A-72-slot-43103.docx.

V. Description du processus d'élaboration d'un projet d'instrument juridique sur la surveillance;

VI. Remerciements pour l'assistance reçue;

VII. Clarifications d'ordre procédural sur le rapport thématique consacré aux mégadonnées et aux données ouvertes.

IV. Conclusion

122. **Les questions soulevées dans le présent rapport ne se limitent pas à quelques pays. L'apparition de vastes collections de données permet aux individus, aux entreprises et aux pouvoirs publics dans le monde entier de prendre des décisions de plus en plus et de mieux en mieux éclairées. Toutefois, la mauvaise gestion de la vie privée risque de compromettre leur utilité potentielle.**

123. **Il est nécessaire de bien comprendre et de pouvoir atténuer les risques d'atteinte à la vie privée, aux autres droits de l'homme connexes et aux valeurs éthiques et politiques d'autonomie et de loyauté.**

124. **Les données sont et resteront une ressource économique clef, au même titre que le capital et le travail. Le respect de la vie privée et l'innovation peuvent aller et vont de pair. Comprendre comment utiliser efficacement les mégadonnées et comment en partager les avantages loyalement sans rogner sur la protection des droits de l'homme sera certes difficile mais utile à terme.**

V. Recommandations

125. **Dans l'attente des observations devant être formulées pendant la période de consultation jusqu'en mars 2018 et du résultat des enquêtes en cours et des lettres d'allégation adressées aux gouvernements, le Rapporteur spécial examine actuellement les recommandations suivantes dans la perspective de les faire figurer dans une nouvelle version du présent rapport devant être publiée en 2018 ou après.**

126. **Les politiques d'ouverture des données doivent indiquer clairement les limites posées à l'utilisation des données à caractère personnel, sur le fondement des normes et des principes internationaux. Il conviendrait notamment d'instituer une exemption pour les informations personnelles avec une obligation contraignante de garantir la fiabilité des procédés de désidentification destinés à permettre la publication sous forme de données ouvertes, et de prévoir des mécanismes d'application solides.**

127. **Toute initiative d'ouverture de l'administration publique mettant en jeu des informations à caractère personnel, que ces données soient désidentifiées ou non, suppose une analyse scientifique, publique et rigoureuse des mesures de protection de la confidentialité des données personnelles, notamment une évaluation de l'incidence sur le respect de la vie privée.**

128. **Les données unitaires sensibles de grande dimension sur les individus ne devraient être publiées en ligne ou échangées que s'il est solidement démontré qu'elles ont été bien désidentifiées et sont protégées contre toute réidentification ultérieure.**

129. **Des cadres devraient être mis en place pour gérer le risque que des données sensibles soient mises à la disposition des chercheurs.**

130. **Les pouvoirs publics et les entreprises devraient soutenir activement l'élaboration et l'utilisation des technologies visant à renforcer la protection de la vie privée.**

131. **Les éléments suivants devraient être pris en considération lors de l'exploitation des mégadonnées :**

Gouvernance

a) Responsabilité : définition des responsabilités, précision des modes de décision et, s'il y a lieu, identification des décideurs;

b) Transparence : indication de la nature, du moment et des modalités de traitement des données personnelles avant publication, et indication de l'utilisation qui en est faite, y compris les « algorithmes ouverts »;

c) Qualité : garanties minimales de qualité des données et de traitement;

d) Prévisibilité : prévisibilité des résultats en cas d'apprentissage automatisé;

e) Sécurité : prise de mesures adaptées pour empêcher que les données et les algorithmes utilisés soient modifiés sans autorisation;

f) Mise au point de nouveaux outils pour détecter les risques et définir les mesures d'atténuation correspondantes;

g) Appui : formation (et, si nécessaire, accréditation) des employés aux obligations juridiques, politiques et administratives relatives aux informations à caractère personnel;

Environnement réglementaire

h) Des dispositions devraient être prises pour doter les autorités chargées de protéger les données des citoyens d'un mandat, de responsabilités et de pouvoirs clairs;

i) Les pouvoirs de réglementation devraient être à la hauteur des nouveaux enjeux posés par les mégadonnées, et les autorités de réglementation devraient notamment pouvoir examiner les procédés et les résultats d'analyse;

j) Les lois sur la protection de la vie privée devraient être examinées pour veiller à ce qu'elles soient adaptées au regard des enjeux soulevés par les avancées technologiques, telles que les données personnelles générées par la machine, et par les outils d'analyse des données, tels que les procédés de désidentification;

Mise en place de mécanismes de retour d'information

k) Il conviendrait d'instituer des mécanismes de consultation, notamment des comités d'éthique, avec les organisations professionnelles, associatives et autres ainsi qu'avec les citoyens, en vue de prévenir l'érosion des droits et de recenser les pratiques solides;

l) Une vaste consultation devrait être menée sur les recommandations et les questions figurant dans le présent rapport, notamment le souhait d'interdire la mise à disposition de jeux de données publiques;

Recherche

m) Technique : les techniques relativement nouvelles que sont notamment la confidentialité différentielle et le cryptage homomorphique devraient être étudiées pour déterminer si elles permettent d'assurer des processus et des résultats respectueux de la vie privée;

n) Il faudrait étudier la sensibilisation des citoyens aux activités des pouvoirs publics et des entreprises en matière de données, à l'utilisation des informations à caractère personnel, notamment pour la recherche, ainsi qu'aux moyens technologiques destinés à renforcer la maîtrise exercée par les individus sur les données les concernant et à accroître leur capacité de les utiliser pour leurs besoins personnels.