La Privacy by design : une fausse bonne solution aux problèmes de protection des données personnelles soulevés par l'Open data et les objets connectés ?

Philippe Pucheral, Alain Rallet, Fabrice Rochelandet, Célia Zolynski

La Privacy by design : une fausse bonne solution aux problèmes de protection des données personnelles soulevés par l'*Open data* et les objets connectés ?

Philippe Pucheral, Professeur d'informatique,
Université Versailles Saint Quentin, INRIA - Paris Saclay
Alain Rallet, Professeur d'économie, Université Paris Sud - Paris Saclay
Fabrice Rochelandet, Professeur en sciences de la communication,
Université Sorbonne Nouvelle
Célia Zolynski, Professeur de droit privé,
Université Versailles Saint Quentin – Paris Saclay

La régulation de la collecte et de l'exploitation des données à caractère personnel est confrontée à des problèmes de plus en plus complexes avec la multiplication incessante des sources de collecte, la capacité technologique de les traiter et d'individualiser ce traitement et l'extension de leur exploitation aux domaines les plus divers. Le développement de l'*Open data* et des objets connectés sont emblématiques de cette complexité.

L'*Open data* vise à ouvrir les bases de données publiques à des fins de transparence démocratique et d'innovation économique ou sociale, ces données étant la source potentielle de nouveaux services. Or une telle ouverture peut soulever d'importantes difficultés concernant la protection des données à caractère personnel, directement ou indirectement par recoupement des bases[1]. Les objets connectés posent avec encore plus d'acuité ces problèmes puisqu'ils enregistrent nos activités personnelles au travers de leur diversité (du nombre de pas effectués chaque jour à l'utilisation de nos appareils ménagers), de leurs localisations et de leurs temporalités[2].

Open data et objets connectés sont ainsi la source d'innovations prometteuses mais prennent dans le même temps en étau l'individu confronté aux flux d'informations personnelles descendants de l'*Open data* et ascendants des objets connectés. Comment sortir l'individu de cet étau, source potentielle de préjudices, sans le priver du bénéfice des innovations exploitant le potentiel de l'*Open data* et des objets connectés ? Tel est le dilemme général de la régulation de la *privacy*, dilemme singulièrement relancé par ces deux sources de données.

La *Privacy by design* (PbD) est à cet égard un mode de régulation séduisant puisqu'il intègre la protection des données à caractère personnel dès la conception des outils de collecte, de traitement ou d'exploitation des données. La PbD agit comme un filtre, à la charge du fabriquant voire de l'exploitant des objets connectés ou du diffuseur de bases de données dans le cas de l'*Open data*, minimisant les risques d'une exploitation

[1] Groupe de l'article 19, avis 6/2013 sur la réutilisation des informations du secteur public et des données ouvertes, 5 juin 2013, WP 207, p. 3&s. ou encore G. Force et F. Pillet, Rapport d'information au Sénat sur *l'Open data* et la protection de la vie privée, 16 avril 2014, p. 24 &s.

[2] Groupe de l'article 29, avis 8/2014 sur le récent développement de l'internet des objets, 16 sept. 2014, WP 223, p. 6&s. - *Adde*, T. Piette-Coudol, *Les objets connectés – Sécurité juridique et technique*, LexisNexis, 2015.

préjudiciable aux individus sans fermer la porte aux innovations, même si elle peut en restreindre le champ. Cela explique qu'elle soit désormais plébiscitée par les autorités de régulation pour promouvoir le développement de ces nouveaux usages[3].
Il s'agit dès lors d'interroger la pertinence de cette solution, son intérêt et ses limites au regard des problèmes soulevés par *l'Open data* et les objets connectés grâce à une étude multidisciplinaire croisant analyses juridiques, techniques et économiques afin de prendre la mesure de l'efficacité de ce nouveau principe. Un constat peut alors être dressé : la PbD est l'objet d'un paradoxe dans la mesure où, prétendant régler les problèmes de protection en amont, elle en fait naître en aval. C'est pourquoi une autre logique de protection peut être plébiscitée, la *Privacy by Using* (PbU), qui est tout à la fois le complément et l'inversion de la PbD.

Pour le comprendre, il convient tout d'abord de présenter la PbD et sa consécration juridique (I). Il s'agit ensuite d'affiner son rôle selon que ce concept est appliqué à l'Open data ou aux objets connectés (II). Il est enfin nécessaire d'en souligner les limites techniques, juridiques et économiques, ce qui conduit à énoncer le paradoxe de la PbD (III). Il en découlera la nécessité de lui adjoindre une autre logique, celle de la *Privacy by Using.*

I- Définition et consécration juridique de la PbD

Afin d'analyser la démarche des autorités de régulation et des opérateurs en faveur de la promotion de la PbD, il convient de revenir sur la définition de ce concept (A) ainsi que sur sa consécration juridique (B).

A- Définition de la PbD

Le principe de *Privacy by Design* vise l'intégration de la protection de la vie privée dès la conception du traitement des données à caractère personnel à un nouvel outil, procédure ou service qui devra s'y conformer tout au long de sa vie : « *il s'agit de faire ab initio de la garantie de la vie privée une cellule de veille placée au sein de la technologie en phase de conception* »[4]. La *PbD* est donc une modalité, parmi d'autres, de la régulation de la vie privée dans l'univers numérique.

Plus précisément, le concept de *Privacy by Design* se structure autour de 7 principes fondateurs, popularisés dès la fin des années 1990 par Ann Cavoukian, Commissaire à la protection des données personnelle de l'Etat d'Ontario. Si ces principes prennent leur racine dans l'analyse des PETs (*Privacy Enhancing Technologies*), leur spectre est significativement plus large et englobe des aspects organisationnels et opérationnels. Il s'agit donc d'abord d'une philosophie de conception et de construction de systèmes d'information traitant des données personnelles, voire, au-delà, d'une méthodologie ou d'un processus, qui suppose, outre l'assemblage de procédés techniques, de penser des

[3] Concernant l'*Open data*, v. l'avis 6/2013 du Groupe de l'article 29, préc. p. 7&s. et le rapport au Sénat préc. p. 66 - Quant à l'internet des objets v. déjà la communication de la Commission européenne « L'internet des objets. Un plan d'action pour l'Europe », 18 juin 2009, COM(2009) 278 final, p. 7 ou encore l'avis 8/2014 du G29 préc., spéc. p. 19 – *Adde*, N. Weinbaum, « Les données personnelles confrontées aux objets connectés », *Comm. comm. électr*. 2014, étude 22, , n°7&s.
[4] G. Loiseau, « De la protection intégrée de la vie privée (*privacy by design*) à l'intégration d'une culture de la vie privée », *Légipresse* 2012/300, p. 712.

mesures organisationnelles et, plus généralement, de mettre en oeuvre une politique *data responsable*.

Ces 7 principes[5] se résument ainsi : (1) *Proactivité*, afin de prévenir les risques d'atteinte à la vie privée plutôt que d'essayer d'en corriger les conséquences *a posteriori* ; (2) *Protection par défaut*, de sorte à protéger la vie privée de l'individu en toutes circonstances, même sans action préalable de sa part ; (3) *Protection par construction*, intégrant le respect de la vie privée dès la conception du système plutôt que d'apporter des correctifs a un système conçu sans prise en compte de cette dimension ; (4) *Somme positive*, garantissant un intérêt partagé entre l'individu, qui bénéficie d'un service non altéré par la protection de sa vie privée, et le prestataire de ce service qui en tire un avantage concurrentiel ; (5) *Protection de bout en bout*, pendant toute la durée de vie d'une information, jusqu'à sa destruction physique ; (6) *Visibilité et transparence*, permettant de contrôler l'exactitude des informations stockées ainsi que d'auditer le comportement du système ; et enfin (7) *Souveraineté de l'individu*, reconnu comme le chef d'orchestre autour duquel s'organisent tous les échanges d'information dans ce qu'il convient d'appeler son éco-système de données personnelles.

Si ces principes sont simples à énoncer, leur mise en œuvre effective introduit des verrous technologiques dont certains s'avèrent particulièrement ardus à lever. Ces principes rentrent ainsi en résonance avec des mécanismes techniques tels que la *minimisation* (réduction de la collecte de données au strict minimum requis pour l'accomplissement d'un objectif), *l'anonymisation* (rendre une donnée nominative indistinguable de celle d'un groupe d'individus en interdisant toute forme de ré-identification), la *destruction physique de données* (qui doit être irréversible et effective quel que soit le nombre de copies effectuées de cette donnée), *l'auditabilité* (qui doit présenter toutes les garanties attestant de sa fiabilité). On conçoit aisément la difficulté de mettre en œuvre de tels mécanismes rentrant frontalement en opposition avec des contextes de plus en plus ouverts (*Open data*, objets connectés, hébergement sur le Cloud) et l'impérieuse nécessité de faire des traitements analytiques sur ces nouveaux gisements de données (*Big Data*).

B- Consécration juridique de la PbD

Les autorités européennes ont entendu promouvoir le principe de Privacy by design en lui conférant une réalité juridique[6]. Le règlement sur la protection et le libre échange des données à caractère personnel dans l'Union européenne lui consacre une disposition[7]. Qualifiée de principe fondamental de la protection des données personnelles, la PbD permet, dans le même temps, de porter les solutions innovantes des entreprises, particulièrement dans le domaine des objets connectés et dans le cadre des politiques d'Open data, comme celles préconisées par le groupe de l'article 29 dans ses avis rendus

[5] A. Cavoukian, « Privacy by design - The 7 foundational Principles » https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

[6] F. Musiani, « La notion de Privacy by design », in *Numérique et libertés : un nouvel âge démocratique*, Rapport à l'Assemblée nationale n°3119 présenté par C. Paul et C. Féral-Schuhl 2015, p. 282, spéc. p. 283.

[7] Article 23, Règlement sur la protection et le libre échange des données personnelles dans l'Union européenne.

en la matière[8]. Elle serait ainsi l'instrument idoine pour opérer la balance des intérêts entre protection et innovation[9].

Une des caractéristiques centrales de la PbD est de faire peser la charge de la régulation de la *privacy* sur les opérateurs exploitant les données. Pour les utilisateurs, la PbD entend alléger leur charge cognitive et résoudre le fameux « Privacy Paradox », autrement dit la discordance entre l'affirmation des principes de protection et les pratiques réelles de divulgation. Avec la PbD, l'individu est protégé de l'opérateur ainsi que de lui-même.

En outre, la PbD doit permettre de dépasser les limites que rencontrent aujourd'hui les autorités de régulation dans leur pouvoir d'intervention. Ces autorités promeuvent désormais, au sein de l'Union européenne, une logique de responsabilisation des professionnels, voire la diffusion d'une culture de l'éthique des données au sein de l'entreprise. C'est ce que font aujourd'hui la CNIL française et le groupe de l'article 29 et ce que préconise pour demain le règlement européen : la nécessité est impérieuse pour ces « éducateurs prescripteurs » de s'appuyer sur les acteurs du secteur pour promouvoir les pratiques « data-responsables » dans une démarche de co-régulation[10]. Il convient donc d'encourager le recours à l'auto-régulation, sous le contrôle de la co-régulation dans la mesure où l'Europe ne semble pas encourager la règle « *Let's the market talk* ! »

La PbD s'inscrit dès lors dans la logique d'analyse de risque portée par le règlement, prolongée par une consécration de l'*accountability*, *i.e.* de l'obligation de rendre des comptes et de justifier des garanties mises en œuvre pour prévenir tout risque ou y remédier en cas de survenance (par ex. en cas de faille de sécurité)[11]. Les exigences conduisant à imposer cet ensemble d'obligations au responsable de traitement pour attester de son niveau de conformité à la réglementation dépendront toutefois du degré de risque généré. Le principe de proportionnalité impose en effet d'écarter toute approche englobante au profit d'une analyse pragmatique qui commandera de s'adapter au traitement opéré, aux techniques disponibles, ainsi qu'à la nature, à la portée, au contexte ou encore aux finalités du traitement. La PbD traduit ainsi la balance des intérêts qu'il convient d'opérer au cas par cas.

II- Application de la PbD à l'*Open data* et aux objets connectés : principe de prévention et principe de protection

La PbD répond à une même idée (intervenir en amont pour limiter les risques de dommages liés à l'exploitation de données personnelles sans en même temps empêcher cette exploitation) mais s'exerce différemment selon qu'elle s'applique à l'*Open data* ou aux objets connectés.

[8] V. les avis 6/2013 et 8/2014, préc.
[9] I. Falque-Pierrotin, M. Griguer et M. Mossé, « Comment gagner la confiance des individus à l'ère du Big data ? », *Cahiers droit de l'entreprise* 2014, entretien n°6.
[10] I. Falque-Pierrotin, M. Griguer et M. Mossé, art. préc.
[11] Conseil d'Etat, *Le numérique et les droits fondamentaux*, La documentation française, 2014, p. 191 - *Adde*, A. Debet, J. Massot et N. Métalinnos, *Informatique et libertés, La protection des données personnelles en droit français et européen*, Lextenso éd. 2015, n°974&s.

L'*Open data* s'apparente à une descente d'informations (d'origine publique) sur les individus tandis que les objets connectés procèdent à une remontée d'informations à partir des individus. Les filtres opérés par la PbD sont différents dans les deux cas et conduisent à promouvoir la PbD soit comme principe de prévention (A), soit comme principe de protection (B).

A- La PbD comme principe de prévention

Quand l'information descend sur l'individu, la problématique est celle d'un *principe de prévention* dans le prolongement des deux premiers principes identifiés par Ann Cavoukian : empêcher que les données transmises par les administrations aux exploitants puissent être utilisées ou réutilisées de manière préjudiciable pour les individus qu'elles permettent d'identifier ou de ré-identifier, quel que soit l'usage qui sera fait des données, sauf si ceux-ci y ont consenti ou si une disposition législative ou réglementaire le permet[12].

L'individu ne joue là aucun rôle. Il ne subit que les bénéfices ou les préjudices finals. Le problème se situe entre l'administration et l'exploitant : quelles précautions doit prendre l'administration pour que l'ouverture des données ne puisse conduire à une exploitation préjudiciable pour les individus sans constituer pour autant "un obstacle excessif au marché de la réutilisation des données"[13] ? Quels engagements l'exploitant doit-il prendre pour avoir accès à ces données ? Sur le premier point, il s'agit de prévenir deux risques, l'un *ex ante*, l'autre *ex post*.

Le risque ex ante est de laisser la porte ouverte à une exploitation préjudiciable des données. Il est d'autant plus grand que les bases de données n'ont pas été construites pour être ouvertes et que les administrations ne sont pas toujours préparées, et encore moins organisées, pour prévenir les risques associés à l'ouverture de leurs données[14]. Un ensemble de mesures organisationnelles et techniques est proposé à cette fin. Les administrations sont ainsi incitées à développer une culture de la gestion du risque consistant à mener des études d'impact en associant le délégué aux données personnelles à leur politique d'*Open data* [15] et à mettre en réseau leurs chefs de projet pour bâtir, aux côtés des autorités de régulation (CNIL, CADA en France), des bonnes pratiques à l'échelle nationale et européenne[16].

Les techniques d'anonymisation des données sont également au cœur de ce dispositif de prévention[17]. Le problème est alors celui de la garantie effective qu'elles apportent. En

12 Loi n°78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal modifiée par l'ordonnance n°2015-1341 du 23 octobre 2015 transposant la directive 2013/37, article 13.

13 Groupe de l'article 29, avis 6/2013 préc., p. 3.

14 L'enquête de la CNIL réalisée en 2014 a révélé que les pratiques d'anonymisation demeuraient parfois assez rudimentaires (ex. simple retrait des données les plus identifiantes telles que le nom et l'adresse des personnes).

15 En ce sens, v. le rapport au Sénat présenté par G. Force et F. Pillet, préc. p. 55 et la recommandation n°18 ainsi que l'avis 6/2013 du Groupe de 29, préc. p. 32.

16 V. sur ce point les propositions formulées par le Conseil national du Numérique dans son rapport au Premier Ministre, *Ambition numérique – Pour une politique française et européenne de la transition numérique* , juin 2015, p. 147 &148.

17 En ce sens, l'analyse du Groupe de l'article 29 dans son avis 05/2014 du 6 avril 2014 sur les techniques d'anonymisation, p. 3.

effet, comme l'ont démontré de multiples études [18] et attaques médiatisées, l'anonymisation n'est pas un mécanisme binaire rendant une donnée anonyme subitement inoffensive. Cette technique introduit toujours un compromis entre risque de réidentification et utilité de la donnée anonymisée. Par exemple, un des modèles fondateurs de l'anonymisation est la *k-anonymité*, consistant à réduire la précision, et donc l'utilité, de certaines données personnelles de sorte à masquer un individu dans une foule de k-1 autres individus[19]. L'objectif est qu'après dégradation, les k individus d'un même groupe partagent ces mêmes caractéristiques, les rendant indistinguables (ainsi, une adresse exacte peut être dégradée en code postal, une date de naissance en classe d'âge, etc). Mais il a été montré que ce modèle ne masque pas en toutes circonstances le lien entre un individu précis et certaines de ses données que l'on voulait justement garder privées. Il en résulte donc une succession d'autres modèles (*l-diversité*, *t-closeness*, *m-invariance*), de plus en plus protecteurs de la vie privée au prix d'une dégradation de plus en plus forte de l'utilité des données. Des modèles alternatifs existent, tels que la *differential privacy*[20] bruitant les données tout en préservant certaines propriétés statistiques, pour des cas d'usage, hélas, très limités[21]. Transparait donc un conflit clair entre les principes de PbD reposant sur l'anonymisation et le principe de *Somme Positive* (4ème principe énoncé par Ann Cavoukian) qui voudrait que le service obtenu ne soit en rien dégradé. L'équation semble impossible à résoudre. Des alternatives émergent, consistant à anonymiser les données par rapport à un objectif précis connu au préalable, permettant ainsi de trouver le meilleur compromis entre protection et utilité des données[22]. Mais une solution généraliste parait encore hors de portée. Dès lors, pour que l'anonymisation soit correctement conçue, il faut en définir les conditions et l'objectif afin de choisir la technique idoine au cas par cas[23]. Les nouvelles politiques de diffusion ouverte devront à l'avenir prendre en compte cette dimension afin de mieux gérer le risque *ex ante*.

Le risque ex post vient du caractère dynamique des données et de la possibilité de ré-identifier des personnes à partir des données anonymisées publiées. La ré-identification peut survenir par rapprochement avec d'autres sources d'information[24] ou par suite de l'évolution des méthodes de traitement,, voire par application d'algorithmes statistiques

[18] A. Narayanan, V. Shmatikov, Robust De-anonymization of Large Sparse Datasets, *IEEE Symposium on Security and Privacy*, 2008, pp111-125..

[19] L. Sweeney, « k-anonymity: a model for protecting privacy», *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 2002, 10 (5), pp. 557-570,.

[20] Cynthia Dwork and Aaron Roth, «The Algorithmic Foundations of Differential Privacy», Theoretical Computer Science: Vol. 9: No. 3–4, pp 211-407.

[21] V. l'analyse de ces techniques et de leurs avantages et inconvénients par le Groupe de l'article 29 dans son avis 05/2014, préc, p. 12&s.

[22] A. Basu et al., « A Privacy Risk Model for Trajectory Data», *IFIP Advances in Information and Communication Technology*, 2014, pp 125-140.

[23] Le groupe de l'article 29 préconise ainsi de prendre en compte des éléments contextuels « par exemple, la nature des données originales, les mécanismes de contrôle en place (y compris les mesures de sécurité restreignant l'accès aux ensembles de données), la taille de l'échantillon (aspects quantitatifs), la disponibilité de ressources d'informations publiques (sur lesquelles peuvent s'appuyer les destinataires), la communication envisagée de données à des tiers (limitée ou illimitée, par exemple sur l'internet, etc. ») » (avis 5/2014 sur les techniques d'anonymisation préc., p. 27-28).

[24] Dans son avis sur la réutilisation des informations du secteur public, le groupe de l'article recommande de prendre en compte les autres données disponibles et la probabilité d'une tentative de ré-identification à partir d'elles (avis 6/2013 préc. p. 17).

pouvant conduire à une « fausse anonymisation »[25] . Le risque de ré-identification est d'autant plus grand qu'il dépend de l'usage fait des données alors que cet usage n'est pas prédéfini.

Les risques d'exploitation préjudiciable des données peuvent être limités par les engagements des exploitants au regard de l'utilisation qu'ils font des données. Pour une part, ces engagements ressortent de l'application des dispositions de la loi Informatique et Libertés, ce que rappelle l'article 13 de la Loi CADA modifiée par l'ordonnance du 23 octobre 2015. Pour une autre part, ces engagements peuvent être matérialisés par le recours aux licences d'utilisation des données. Le Conseil national du numérique propose ainsi de garantir le respect de l'anonymisation des données placées en Open data lors de leur réutilisation par le recours à une licence spécifique qui devrait être généralisée[26]. Celle-ci pourrait imposer à toute personne désirant réutiliser les données une obligation de mentionner le fait que le jeu de données a fait l'objet d'une anonymisation, ou encore stipuler l'interdiction des recoupements d'informations anonymisées ou toute autre pratique qui permettrait de rendre les données identifiantes[27]. Instruments de pédagogie, ces licences auraient ainsi pour intérêt de borner les possibilités de ré-utilisation, notamment celles susceptibles de conduire à des exploitations préjudiciables.

Enfin, sur un plan économique, se pose la question de l'incitation des exploitants à adopter un comportement de prévention optimal. Cela revient à s'interroger sur le régime de responsabilité (hors contractualisation) incitant l'exploitant à minimiser le préjudice potentiel causé aux individus par le traitement de données ouvertes. Or, dans le contexte de l'Open data, l'exploitant ne dispose pas toujours des informations sur les coûts en termes de préjudices (les données personnelles pouvant circuler et causer des préjudices en cascade), ni anticiper quelles données (sensibles) peuvent apparaître à l'occasion d'un type de traitement donné (par exemple, des données sur la religion de la personne par le croisement de bases de données[28]). Dans ce cas, les exploitants peuvent investir excessivement (surévaluant le risque de devoir compenser les victimes) ou au contraire insuffisamment (étant dans l'incapacité d'apprécier les risques face à des profits et des coûts certains) dans la conception de services conformes aux règles de PbD.

B- La PbD comme principe de protection

Lorsque l'information remonte à partir des individus, ce qui est le cas des objets connectés, le filtre constitué par la PbD n'est plus le même. C'est un filtre entre l'individu et l'exploitant, l'individu étant la source directe des informations à caractère personnel via les objets qu'il manie. Le risque n'est plus de même nature. Dans le cas de l'*Open data*, il s'agissait de prévenir un risque potentiel lié à l'ensemble des utilisations et ré-utilisations possibles de données publiques n'ayant pas été conçues pour une exploitation particulière. Il fallait alors énoncer les règles d'une protection générale

[25] W. Maxwell, « Données publiques et données personnelles : un 'mariage pour tous' aux limites de la légalité », *Editions Multimédi@,* 22 juil. 2013, n°84, p. 8.
[26] Conseil national du Numérique, Rapport « Ambition numérique », préc., p. 141.
[27] Egalement en ce sens v. l'avis 6/2013 du Groupe d l'article 29 préc., p. 29.
[28] V. C. Su et al., « Privacy-Preservation techniques in data mining », in A. Acquisti et al., *Digital Privacy. Theory, Technologies, and Practices*, Auerbach Publications, 2008, pp.187-226.

minimisant les risques d'un mésusage, quel que soit l'usage. C'est pourquoi la qualification de *principe de prévention* peut être alors retenue.

Dans le cas des objets connectés, il s'agit de filtrer les données qui remontent de l'individu vers le collecteur des données sachant que ces données peuvent avoir *ipso facto* un caractère personnel (enregistrement d'un comportement lié à une personne identifiée via l'individualisation connue de l'objet connecté). On pourrait soutenir que cette situation n'est pas sur le fond différente de celle des données personnelles collectées au moyen des cartes de fidélité ou de transactions sur le Web. Il suffirait alors d'appliquer la réglementation qui régit toute exploitation de données à caractère personnel quel que soit le moyen de collecte.

Or deux facteurs modifient sensiblement la problématique de protection inscrite dans la réglementation existante, du moins du côté de l'individu qui est appelé à exprimer un consentement et exercer les droits qui lui sont conférés. D'abord, les données collectées sont potentiellement plus sensibles. Elles le sont non seulement parce qu'elles s'étendent à des domaines par nature sensibles (localisation, santé…), mais aussi parce qu'il y a un lien très étroit, consubstantiel, entre la donnée collectée par l'objet connecté et l'intimité de la personne car le service fourni repose sur la consubstantialité de ce lien (je révèle ma localisation, mon rythme cardiaque ou ma consommation électrique via l'objet connecté car le service qu'il me rend le présuppose). Les objets connectés fonctionnent à cet égard comme les réseaux sociaux : l'exposition de soi est un élément intrinsèque du service obtenu. Ils ne sont pas dissociables. Contrôler la divulgation des données à caractère personnel devient un problème redoutable, bousculant la notion même de consentement.

L'autre nouveauté est la gigantesque accumulation de données qu'annonce le développement des objets connectés. Selon une étude de l'institut IDATE, 42 milliards d'objets sont déjà connectés aujourd'hui et plus de 80 milliards le seront en 2020, date à laquelle le volume de données annuel produit sera de l'ordre de 44 Zettaoctets (soit 44 000 milliards de gigaoctets). Les algorithmes de Big Data portent la promesse d'analyser ces nouveaux déluges de données et d'en extraire une connaissance enfouie échappant à l'expertise humaine. Mais ils placent en porte à faux la réglementation fondée à l'inverse sur des principes de limitation : minimisation des données collectées, finalité définie de la collecte, proportionnalité de la collecte et du traitement à l'égard de cette finalité…

L'exposition de soi via les objets connectés et l'accumulation de données qu'ils produisent altèrent fortement la capacité d'attention rationnelle des individus[29], rendant très problématique l'application d'une réglementation qui repose sur cette capacité. Cette capacité d'attention est en outre altérée par la mise en réseau consubstantielle au fonctionnement des objets connectés dont beaucoup reposent sur

[29] L'individu utilisateur d'objets connectés peut être victime de biais cognitifs tout en étant informé sur les petits risques associés à chaque divulgation, même automatisée, de ses données (mes performances sportives enregistrées automatiquement par une application de training), mais étant en même temps incapable de sommer l'ensemble de ces petits risques (mes performances sportives et les commandes automatiques d'aliments gras par mon réfrigérateur connecté, etc.).

des communications *Machin to Machin* qui peuvent transformer un traitement routinier en un traitement particulièrement sensible *in fine*[30].

La PbD présente alors un intérêt déterminant : protéger en amont l'individu (les principes 3 et 4 d'Ann Cavoukian) qui n'est plus en situation de le faire lui-même. Elle s'apparente à un *principe de protection* par défaut. Les techniques employées ont des similitudes avec celles utiles au principe de prévention mais leur contexte d'utilisation est différent. Pour l'Open data, il s'agit de rendre difficile l'identification ou la ré-identification des individus. Concernant les objets connectés, il s'agit au contraire de convertir une donnée brute en une donnée agrégée ou brouillée, donc moins intrusive même si elle reste potentiellement identifiante.

L'avantage de la PbD est d'inscrire la protection des données à caractère personnel dans la conception des dispositifs techniques utilisés (moyens de collecte, bases de données, stockage des données, outils de traitement, outils de publication, etc.). Or cette remontée ouvre toute une série de problèmes qui ne peuvent être réglés qu'en aval. C'est le paradoxe de la PbD.

III- Le paradoxe de la PbD : pour la promotion de nouveaux modes de gestion de la privacy

L'existence de limites techniques, l'incertitude et l'arbitrage entre protection et innovation expliquent la nécessité d'ajustements aval (A) ainsi que la recherche de solutions permettant de dépasser le paradoxe de la PbD (B).

A - La nécessité d'ajustements aval

La nécessité de ces ajustements s'explique par les limites techniques, l'incertitude inhérente aux risques encourus et la délicate balance des intérêts entre protection et innovation.

1- Les limites techniques

Nombreux sont ceux qui constatent l'efficacité limitée des techniques de la PbD[31]. Sa mise en œuvre pratique se heurte en effet à une première gageure : traduire en concepts algorithmiques des principes dont tous s'accordent à reconnaître la généralité et le flou. Il en résulte une réelle difficulté à estimer si une solution technique répond en pratique à un principe de PbD et, dans l'affirmative, dans quelle proportion. Le risque devient alors que des systèmes se fassent chantres de la PbD sans qu'aucune métrique ne puisse l'attester de façon tangible. Pour autant, on ne peut pas en conclure que les principes de PbD sont inimplémentables. Les techniques d'anonymisation, de minimisation, de chiffrement de données, de contrôle d'accès et d'usage notamment, sont des briques de base indispensables à l'établissement de chaque principe de PbD.

Il serait cependant dangereux d'attendre de la technologie une protection plus forte qu'elle ne peut offrir. Prenons à titre illustratif le principe de *Protection de bout en bout*.

[30] V. l'avis 8/2014 sur le récent développement de l'internet des objets rendu par le groupe de l'article 29, préc., p. 6.
[31] Sur ce constat, v. notamment le rapport de G. Force et F. Pillet au Sénat, préc. p. 43 &s.

Cette protection inclut le concept de rétention limitée des données conduisant à détruire une donnée dès lors que l'objectif pour lequel elle a été collectée est totalement accompli. La destruction physique de données, bien que de perception très intuitive, s'avère quasiment impossible à mettre en œuvre de façon absolue dans la pratique ; car on attend d'une destruction de données qu'elle soit irréversible. Or de nombreux facteurs compromettent cette irréversibilité. Il est en effet très difficile de contrôler le nombre de copies qui ont pu être faites d'une donnée par des mécanismes de caches sur Internet (qui copient la donnée pour la resservir plus rapidement) ou de reprise après panne (qui copient la donnée pour restaurer un état cohérent d'un système après une panne). Les traces laissées par une donnée dans les index peuvent également être combinées pour reconstruire tout ou partie d'une donnée. Ceci ne doit pas remettre en cause le principe même de rétention limitée des données, outil puissant de protection de la vie privée, mais faire prendre conscience que la seule protection ultime d'une donnée serait de ne jamais la produire !

Une autre difficulté majeure réside dans la composition de principes PbD parfois antagonistes, rendant la construction d'une solution cohérente hasardeuse. Prenons pour exemple le principe de *Minimisation* déjà évoqué. Ce principe a pour objectif de limiter la collecte d'information au minimum requis pour l'accomplissement d'un objectif. Mais les objets connectés ont radicalement changé la donne en terme de collecte et les algorithmes de *Big data* bousculent l'idée même d'un objectif prédéfini. En effet, le *Big data* se nourrit de ces gisements de données pour en extraire de nouvelles connaissances cachées. Comment alors définir le minimum d'information requis pour accomplir un objectif que le *Big data* est censé découvrir par lui-même ? Le principe de *Proactivité* pourrait se traduire ici par la nécessité de stocker les données à la source, c'est à dire dans les objets qui les ont captées, pour éviter une centralisation de données brutes très intrusives sur des serveurs où elles seront difficiles à protéger *a posteriori*[32]. Mais alors, le principe de *Somme Positive* appellerait à un déploiement d'algorithmes *Big data* massivement décentralisés, pour préserver l'intérêt de cette analyse de données pour la communauté. De tels algorithmes restent aujourd'hui grandement du ressort de la recherche du fait du facteur d'échelle envisagé.

2 - Les ajustements *ex post* liés à l'incertitude

Les techniques d'anonymisation n'offrant pas de garantie certaine, l'emploi de ces techniques assorties de contraintes juridiques ne suffira pas à éliminer tout risque résiduel de réidentification des personnes dans le cas de *l'Open data* ou d'exploitation préjudiciable dans le cas des objets connectés.

Au delà des risques résiduels qui sont des risques connus à l'instant *t* mais contre lesquels on ne peut totalement se prémunir, il existe des occurrences inconnues ou difficilement probabilisables. De même qu'il était difficile de prévoir, avant que l'événement ne survienne, qu'un employeur puisse exploiter des conversations entre amis sur Facebook, de même est-il difficile d'imaginer tous les risques associés à l'exploitation des données de notre consommation électrique ou de notre bien-être.

[32] A titre d'exemple, l'analyse de la trame sortant de certains compteurs électriques intelligents permet d'identifier tous les équipements électriques domestiques et donc de tracer toute l'activité du foyer au cours d'une journée. Des chercheurs allemands ont même montré qu'il était possible d'identifier la chaîne de télévision regardée à un instant *t* en analysant le spectre de consommation de l'écran de TV.

C'est pourquoi il ne suffit pas de mener une analyse de risque préalable à l'ouverture des données ou à l'exploitation des données issues des objets connectés. Cette analyse doit être prolongée en aval, périodiquement, afin de prendre en compte la révélation *ex post* de nouvelles occurrences de risque et le développement des techniques. Plus généralement, il convient de procéder à une veille tout au long de la ré-utilisation potentielle des données. A cette fin, certains préconisent de prévoir des mesures de surveillance des mésusages et d'adopter une attitude proactive pour minimiser une éventuelle atteinte aux données personnelles[33].

3 - Les ajustements aval nécessaires à la balance des intérêts entre protection et innovation

La nécessité d'ajustements *ex post* n'est pas uniquement liée à l'incertitude résiduelle ou structurelle. Elle s'explique aussi par les compromis qui doivent s'établir entre la protection des données personnelles et leur exploitation à des fins économiques. La question se posera par exemple pour déterminer, nonobstant la protection de l'individu, comment garantir la diffusion de données demeurant utiles[34]. La dynamique d'innovation de services serait bridée si le curseur était poussé trop loin du côté de la protection de ces données. Ce, sans évoquer les risques d'atteintes aux libertés individuelles évoqués par ceux dénonçant la surveillance « pervasive » qui pourrait résulter du développement d'architectures techniques présentées comme protectrices de la *privacy*, dans le prolongement des débats précédemment suscités par le recours aux mesures techniques de protection des œuvres de l'esprit[35].

Or il est très difficile de régler *ex ante* le compromis, et de l'incorporer dans la conception du dispositif informatique, dans la mesure où il dépend beaucoup du contexte d'utilisation, du type de données, du type d'usage des données, etc. Des bornes peuvent cependant être posées par la PbD : une trop grande rigidité en amont limiterait l'innovation et une trop grande flexibilité laisserait trop de champ à des exploitations préjudiciables. L'ajustement fin réalisant le compromis *ad hoc* entre protection et exploitation est nécessairement *ex post* car dépendant du contexte d'utilisation.

B- Comment gérer le paradoxe ?

La PbD crée une situation paradoxale : focalisée sur l'incorporation en amont d'une prévention des risques dans des dispositifs techniques, elle rend nécessaire de nombreux ajustements aval. Il faut procéder à une veille constante des techniques et des risques nouveaux, effectuer des analyses périodiques de risque, examiner les arbitrages protection/innovation, établir les compromis nécessaires… C'est une situation difficile à gérer pour deux raisons. Tout d'abord, au plan organisationnel, qui va être en charge de ces ajustements et en supporter le coût : les exploitants, les administrations, les autorités de régulation ? Ensuite, que faut-il inclure dans la PbD : comment délimiter les traitements amont et aval ?

[33] V. les analyses de G. Force et F. Pillet dans leur rapport au Sénat préc., p. 64 et les recommandations 9&10.
[34] G. Force et F. Pillet, rapport préc., recommandation n°11, p. 65 - également en ce sens l'avis 6/2013 du groupe de l'article 29, préc., p. 20.
[35] F. Musiani, art. préc., p. 283&s.

Par exemple, dans le cas de l'*Open data*, on pourrait être tenté de radicaliser le principe de prévention, d'en pousser la logique jusqu'au bout. Mais cela limiterait fortement la dynamique d'innovation. Il a ainsi été proposé de prévoir une « *stratégie de rapatriement ou de suppression des jeux de données compromis* », mais ces pratiques peuvent poser difficulté compte tenu du caractère circulant de la donnée[36]. Une « accessibilité modulée » peut aussi être préconisée lorsque les informations publiques comportent des données personnelles : l'accessibilité serait admise proportionnellement au risque pouvant résulter de l'ouverture des différents jeux de données parmi lesquels il conviendrait d'opérer un tri pour admettre un accès total, partiel ou interdit[37].

Dans le cas des objets connectés, une solution envisageable consisterait, pour prévenir au maximum les risques, à développer une protection décentralisée, sur un modèle « User centrics ». La meilleure garantie de protection des données personnelles serait en effet que l'individu garde le contrôle de ses données au niveau de leur hébergement ou des algorithmes de traitement plutôt que de déléguer cette tâche à des tiers centralisant ces données sur des serveurs et augmentant ainsi leur niveau d'exposition. Redonner le pouvoir aux individus sur leurs données tout en respectant le principe de *Proactivité* (7e principe identifié par Ann Cavoukian) pourrait ainsi se traduire par des solutions de stockage et de gestion de données décentralisées, terrain désormais investi par les éditeurs de Cloud personnel. Un mouvement connexe à l'Open data, qualifié de *Self Data*, est également à l'œuvre et s'inscrit dans cette direction : des initiatives emblématiques de Self Data se font jour comme les projets *Blue Button* (données médicales) et *Green Button* (données énergétiques) aux Etats-Unis, le projet *miData* en Grande-Bretagne et son équivalent *MesInfos* en France. Dans ces initiatives, les individus sont amenés à prendre conscience de ce qu'ils sont source de données valorisables ce qui pourrait les inciter à en plébisciter une meilleure protection. Mais, s'ils y gagnent un meilleur contrôle sur leurs données, ils héritent dans le même temps de la responsabilité de les protéger. La gestion du filtre de protection est ainsi déportée des exploitants vers les individus. La PbD conduirait alors à ce que les individus soient le sujet actif de la gestion de leurs données, ce que précisément la PbD cherchait à éviter (principe n°2 énoncé par Ann Cavoukian). Car le coût cognitif de gérer soi-même la protection de ses données est très élevé et nous fait s'éloigner du principe de *Protection par défaut*. Les éditeurs de Cloud personnel l'ont bien compris en proposant une prise en charge de cette administration. La boucle est alors bouclée : les acteurs ont changé mais la centralisation resurgit. De nouvelles architectures sécurisées, centrées sur l'individu et auto-administrées restent dès lors clairement à inventer. Cela montre que la PbD peut-être utile si on la décharge d'une prétention à fournir une solution technologique simple à un problème complexe. La leçon de la gestion décentralisée illustre combien l'individu ne doit pas se dessaisir de son propre rôle et dans quelle mesure la technologie et le droit doivent lui offrir les outils pour le soutenir dans cette démarche.

La PbD est un instrument utile de régulation de la protection des données personnelles, que ce soit dans le cadre du développement de l'*Open data* ou dans celui des objets

[36] En ce sens, G. Force et G. Pillet, rapport au Sénat préc., p. 65 et la recommandation n°11.

[37] Sur cette diffusion modulée en fonction du risque, v. G. Force et G. Pillet, rapport au Sénat préc., p. 61 et recommandations 7 et 8 – *Adde*, préconisant également un approche au cas par cas, l'avis 6/2013 du Groupe de l'article 29, préc., p. 9.

connectés. Mais elle n'offre pas toutes les garanties de prévention ou de protection que son intervention en amont de la collecte et de l'exploitation des données laisse supposer. En effet, la séduction qu'exerce cet outil est très largement fondée sur l'illusion que les problèmes de protection des données à caractère personnel pourraient être réglés, avant même que ces données soient utilisées, par l'incorporation préventive de la protection dans des dispositifs techniques.

Il ressort pourtant de l'analyse qu'au contraire, la mise en œuvre de la PbD suppose des ajustements aval importants en raison des limites des techniques employées, de la méconnaissance de l'ensemble des risques potentiels et de la nécessité d'arbitrages entre protection et innovation dépendants des contextes d'utilisation.

La nécessité de procéder à d'importants ajustements en aval implique de remettre en jeu le rôle des individus dans la protection de leurs données. Ainsi, il serait souhaitable, de mettre en place une protection décentralisée des données et de leurs modes de traitement au niveau des individus alors qu'un des principes essentiels de la PbD est de leur assurer une protection sans action préalable de leur part. Il conviendrait donc de promouvoir, aux côtés de la PbD, des mécanismes d'apprentissage dynamiques de l'utilisateur que l'on pourrait traduire par un nouveau concept de « *Privacy by using »*[38]. Il s'agirait de développer les instruments technologiques, juridiques ou informationnels permettant de développer cette capacité. L'individu serait alors mieux informé de la nature des données collectées et de leur utilisation (par ex. par un tableau de bord personnalisé) d'une part et, d'autre part, des conséquences indirectes et/ou différées de la divulgation explicite ou implicite de données personnelles, par exemple au moyen d'un instrument de diagnostic lors de l'usage d'un objet connecté, sorte de « quantified self » appliqué à la *privacy*. Cela placerait l'utilisateur en capacité de mieux définir son comportement de protection de sa vie privée par un mécanisme d'apprentissage. Il ne le fera pas nécessairement, mais il aura les éléments pour adopter un comportement éclairé[39]. Par son *empowerment*, le consommateur, mis en conscience – ou en « capacitation » – pourrait ainsi prendre des décisions en pleine connaissance, ce qui pourrait ainsi favoriser l'émergence de nouvelles normes de *privacy* partagées[40].

Cette incomplétude de la PbD illustre un problème plus général de la régulation de la *privacy*. Aucun type de régulation, technique ou juridique, ne peut à lui seul prétendre résoudre les problèmes de protection. L'efficacité de la régulation ne peut s'envisager qu'en articulant divers instruments. La PbD n'a ainsi de sens que prise dans un dispositif plus vaste, intégrant par exemple le rôle nécessaire de la *Privacy by using*. Or, ce qu'il manque aujourd'hui à la régulation, c'est certainement la représentation d'un dispositif d'ensemble où serait spécifié le rôle de telle ou telle composante de ce dispositif. La discussion se fait composante par composante, d'où ses limites. Il est désormais

[38] A., Rallet, F. Rochelandet F. et Zolynski C., « De la *privacy by design* à la *privacy by using*: regards croisés droit / économie », *Réseaux*, 2015/33 (189), p. 15-46.

[39] A. Aquisti rappelle ainsi que, même pleinement informé, un individu est affecté dans sa prise de décision par des distorsions psychologiques (actualisation hyperbolique des coûts et bénéfices futurs, ignorance rationnelle, effet de valence, etc.). Si, ce constat tiré de la psychologie comportementaliste semble valide à court terme, il n'en va pas de même à long terme où le comportement bénéficie des connaissances acquises par apprentissage (application de règles prudentielles, sélection accrue des données divulguées, etc.) : « Privacy in electronic commerce and the economics of immediate gratification », *Actes de l'ACM Electronic Commerce Conference*, ACM 21-29 et revue *Réseaux*, 2011, vol.29, n°167.

[40] A. Rallet, F. Rochelandet et C. Zolynski, art. préc.

nécessaire de dépasser cette approche pour penser une vision architecturale de la régulation.

P. P., A. R., F. R. & C. Z.