HAL
open science

Privacy by Design et Big Data

Philippe Pucheral, Alain Rallet, Célia Zolynski

To cite this version:

Philippe Pucheral, Alain Rallet, Célia Zolynski. Privacy by Design et Big Data. Les big data à découvert, 2016. hal-01429075

HAL Id: hal-01429075

https://hal.science/hal-01429075

Submitted on 6 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Privacy by Design et Big Data

Philippe Pucheral. Informaticien, Professeur à l'Université de Versailles Saint-Quentin & Inria. *Philippe.Pucheral@inria.fr*
Alain Rallet. Economiste, Professeur à l'Université Paris-Sud. *Alain.Rallet@u-psud.fr*
Célia Zolynski. Juriste, Professeur à l'Université de Versailles Saint-Quentin. *Celia.Zolynski@uvsq.fr*

Le développement fulgurant du Big Data dans tous les pans d'activité, sur tous les types de données, y compris données à caractère personnel (DP), introduit un challenge majeur dans la recherche d'un équilibre entre recherche et innovation et protection de la vie privée. Les sources de données concernées et les possibilités de les croiser sont multiples, qu'il s'agisse de données captées par les majors du Web, de bases de données publiques ouvertes au titre de l'*Open Data* ou de données produites par les individus eux-mêmes, directement (données personnelles répliquées sur le Cloud) ou indirectement (données captées par des appareils domotiques ou équipements de mesures de soi enregistrant la diversité de leurs activités, leur localisation et leur temporalité).

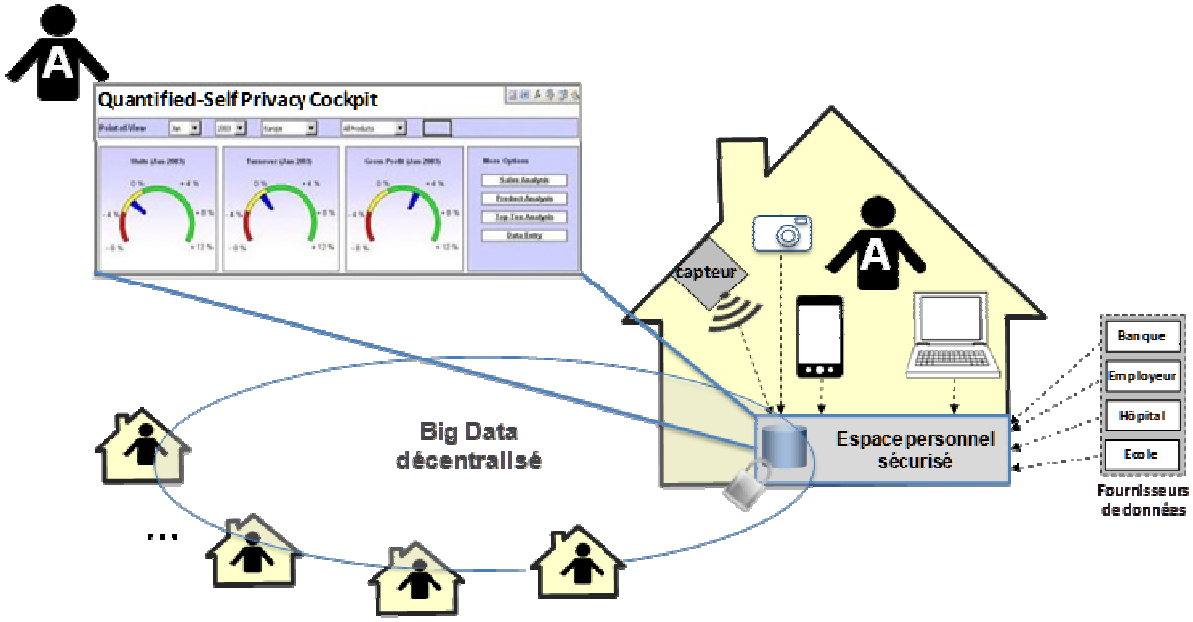
Définition et consécration juridique de la Privacy by Design. Traduction du principe de prévention, la *Privacy by design* (PbD) est un mode de régulation intégrant la protection des DP dès la conception des outils de collecte et de traitement. 7 principes fondateurs[1] la structurent : (1) *Proactivité*, prévenir les risques plutôt que d'essayer d'en corriger les conséquences ; (2) *Protection par défaut*, protéger l'individu même sans action préalable de sa part ; (3) *Protection par construction,* dès la conception du système ; (4) *Somme positive*, garantir un service non altéré par la protection de la vie privée ; (5) *Protection de bout en bout*, pendant toute la durée de vie d'une information ; (6) *Visibilité et transparence*, permettre 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 le concernant. La PbD semble l'instrument idoine pour opérer la balance des intérêts entre protection et innovation[2] et est à ce titre plébiscitée par les autorités de régulation. Le règlement général pour la protection des données dans l'Union européenne lui consacre d'ailleurs une disposition[3]. La charge de la PbD pèse alors principalement sur les exploitants de données.

Limites et paradoxe de la Privacy by Design. Les principes de PbD sont simples à énoncer mais introduisent des verrous technologiques. Ils 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* (en interdisant toute forme de ré-identification), *l'effacement de données* (qui doit être irréversible et effectif quel que soit le nombre de copies de cette donnée), *l'auditabilité* (qui doit présenter des garanties d'infalsifiabilité). De tels mécanismes sont confrontés à (1) des contextes de plus en plus ouverts multipliant les copies de données et diluant les responsabilités (*Open data*, objets connectés,

[1] A. Cavoukian, « Privacy by design - The 7 foundational Principles » https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf
[2] 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.
[3] Article 23, Règlement sur la protection et le libre échange des données personnelles dans l'Union européenne.

hébergement sur le Cloud), (2) des traitements *Big Data* impliquant une collecte massive de données et aux objectifs parfois définis ex post par le traitement et (3) des outils techniques souvent imparfaits (compromis entre protection et performance, entre effectivité de l'anonymisation et conservation des usages, etc). L'incertitude frappe également l'évaluation économique des préjudices imputables à une fuite d'information pour l'individu, conduisant les opérateurs à surévaluer ou sous-évaluer ces risques. La PbD crée ainsi une situation paradoxale : focalisée sur l'incorporation en amont d'une prévention des risques dans des dispositifs techniques, elle implique de nombreux ajustements aval (veille technologique, analyse périodique des risques, réexamen des arbitrages protection/innovation).

De la Privacy by Design à la Privacy by Using. Ces ajustements aval remettent en jeu le rôle des individus dans la protection de leurs données. Ainsi, une protection décentralisée des données et de leurs modes de traitement au niveau des individus[4] parait souhaitable alors qu'un des principes fondateurs 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, un nouveau concept de « *Privacy by using* »[5]. Il s'agirait de développer les instruments technologiques, juridiques ou informationnels permettant de développer une capacité d'apprentissage de l'individu. Celui-ci serait mieux informé de la nature des données collectées, de leur utilisation et des conséquences indirectes et/ou différées de leur divulgation, par exemple au moyen d'un tableau de bord fournissant un diagnostic personnalisé, sorte de « quantified self » de la *privacy*. L'utilisateur serait alors en capacité de construire, par apprentissage, un comportement éclairé de protection (on parle aujourd'hui d'*empowerment*). Ceci impose également de reconsidérer l'algorithmique sous-jacente au Big Data afin de ne pas recentraliser a posteriori l'ensemble de ces données personnelles dans l'unique but d'en faciliter la manipulation.



Instance d'architecture Privacy-by-Using

Ainsi, aucun type de régulation, technique ou juridique, ne peut à lui seul prétendre résoudre les problèmes de protection de la vie privée. Il est aujourd'hui nécessaire de dépasser une approche instrument par instrument pour penser une vision architecturale et holistique de la régulation.

---

[4] N. Anciaux et al. 'Trusted Cells: A Sea Change for Personal Data Services'. *Proc. of the 6th Int. Conf. on Innovative Data Systems Research (CIDR)*, USA, 2013.

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