

Données numériques de masse, « données citoyennes » et confiance dans la statistique publique

Citizen Data and Trust in Official Statistics

Evelyn Ruppert*, Francisca Grommé*, Funda Ustek-Spilda**
et Baki Cakici***

Résumé – Des smartphones, compteurs, réfrigérateurs, des automobiles aux plateformes Internet, les données issues des technologies numériques et les citoyens sont indissociables. Au-delà des questions politiques et éthiques sur le respect de la vie privée, la confidentialité et la protection des données, cela implique de repenser les relations avec le public dans la production de données statistiques si l'on veut que les citoyens leur fassent confiance. Nous esquissons une approche qui implique la co-production de données, avec des citoyens comme partenaires de la production statistique, de la conception de plateformes de production de données à leur interprétation et leur analyse. Si des questions sur la qualité et la fiabilité des données méritent d'être posées, nous estimons que la co-production a le potentiel d'atténuer les problèmes associés à la ré-utilisation de données massives à des fins statistiques. Nous avançons que, dans une période en proie aux « faits alternatifs », ce qui fonde la légitimité d'un savoir et d'une expertise fait l'objet de controverses et de confrontations politiques non négligeables, imposant de dépasser la défense des pratiques existantes pour en inventer de nouvelles. Dans ce contexte, nous estimons que l'avenir des statistiques publiques repose non seulement sur des données et des méthodes inédites, mais aussi sur la mobilisation des possibilités offertes par les technologies numériques pour établir de nouvelles relations avec le public.

Abstract – *From smartphones, meters, fridges and cars to internet platforms, the data of digital technologies are the data of citizens. In addition to raising political and ethical issues of privacy, confidentiality and data protection, this calls for rethinking relations to citizens in the production of data for statistics if they are to be trusted by citizens. We outline an approach that involves co-producing data, with citizens as partners of statistical production, from the design of a data production platform to the interpretation and analysis of data. While raising issues such as data quality and reliability, we argue co-production can potentially mitigate problems associated with the re-purposing of Big Data. We argue that in a time of “alternative facts”, what constitutes legitimate knowledge and expertise are major political sites of contention and struggle and require going beyond defending existing practices towards inventing new ones. In this context, we argue that the future of official statistics not only depends on inventing new data sources and methods but also mobilising the possibilities of digital technologies to establish new relations with citizens.*

Codes JEL / JEL Classification : A14, C93, O35, O38

Mots-clés : sciences participatives, co-production, expérimentalisme, protection de la vie privée dès la conception, statistiques intelligentes

Keywords: *citizen science, co-production, experimentalism, privacy-by-design, smart statistics*

Rappel :

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

* Department of Sociology, Goldsmiths University of London (E.Ruppert@gold.ac.uk; F.Gromme@gold.ac.uk)

** Department of Media and Communications, London School of Economics (f.ustek-spilda@lse.ac.uk)

*** Technologies in Practice, IT University of Copenhagen (bakc@itu.dk)

Nous remercions deux rapporteurs anonymes pour leurs commentaires et suggestions. Nous tenons aussi à remercier, pour leur soutien et leur participation, les nombreux statisticiens des instituts de statistique du Royaume-Uni (ONS), des Pays-Bas (CBS), d'Estonie, de Turquie, de Finlande, ainsi que ceux d'Eurostat et de l'UNECE, qui ont rendu cette recherche possible. Cette recherche a été soutenue par un financement du Conseil européen de la recherche (ERC, convention n° 615588, sous la direction d'Evelyn Ruppert, Goldsmith – Université de Londres).

Reçu le 18 août 2017, accepté après révisions le 6 juin 2018

L'article en français est une traduction de la version originale en anglais

Pour citer cet article: Ruppert, E., Grommé, F., Ustek-Spilda, F. & Cakici, B. (2018). Citizen Data and Trust in Official Statistics. *Economie et Statistique / Economics and Statistics*, 505-506, 179-193. <https://doi.org/10.24187/ecostat.2018.505d.1971>

Depuis maintenant environ cinq ans, les instituts nationaux de statistique (INS) mènent des expériences sur le potentiel d'exploitation dans le cadre de l'élaboration de la statistique publique des données massives générées par diverses technologies numériques. Ces expériences ont permis d'identifier plusieurs sujets de préoccupation, tels que l'accès aux données, la propriété des données, le respect de la vie privée et de l'éthique, la représentativité des données, la qualité des données, et ainsi de suite. Comme l'indique un rapport du groupe de travail de l'UNECE sur la confidentialité des Big Data (UNECE, 2014), les préoccupations traduisent notamment des risques d'atteinte à la réputation et à l'image publique des INS qui ont recours aux sources de données massives. Le rapport en question fait la synthèse d'un certain nombre de stratégies visant à atténuer ces risques, parmi lesquelles : appliquer des principes d'éthique au moyen de dispositifs de redevabilité et de consentement éclairé ; mettre en place des contrôles de conformité rigoureux ; élaborer des systèmes de veille pour le suivi des menaces réputationnelles ; assurer la transparence et la compréhension, en communiquant clairement avec les parties prenantes sur l'usage des données et en organisant des discussions avec le public ; et concevoir un plan de communication de crise. Ce même rapport fait aussi valoir, à l'instar de rapports d'autres organismes internationaux, tels que celui du groupe de travail d'Eurostat sur le Big Data, que réutiliser les données de masse, au-delà des défis techniques, peut également miner la confiance des citoyens dans la manière dont les INS génèrent les données et produisent la statistique publique. Des défis semblables émergent lorsque les INS entreprennent de réutiliser des données administratives, ce qui présente là aussi des enjeux techniques, mais suscite également, du moins pour certains instituts nationaux, des inquiétudes au niveau du partage, de l'appariement et de l'utilisation de données pour une finalité autre que celle prévue initialement.

Bien entendu, la question de la confiance du grand public dans la statistique publique ne date pas d'hier. Si elle préoccupe aussi d'autres parties prenantes, notamment les ministères, les organismes gouvernementaux, les médias, les universités et d'autres organismes de recherche publics ou privés qui s'appuient sur la statistique publique, c'est la confiance des citoyens qui nous intéresse ici. L'histoire des méthodes établies de production de statistique sociale et de population, telles que les questionnaires de recensement ou d'enquêtes, les carnets d'emploi

du temps, montre que s'attacher la confiance des citoyens dans la manière dont les données sont générées et utilisées dans la statistique publique a nécessité des pratiques élaborées. Avec des pratiques telles que la mobilisation de groupes de réflexion, les questionnaires pilotes et la consultation d'organismes de la société civile sur les problématiques de consentement, de protection des données, de confidentialité, d'impartialité et de normes professionnelles, les INS se sont attachés à gagner la confiance des citoyens (Struijs *et al.*, 2014). Vue sous cet angle, la confiance est le produit non pas d'une, mais de maintes pratiques qui concourent à la fiabilité de la statistique publique.

Les Big Data n'étant pas produites par les administrations publiques mais par des entreprises privées, comme les propriétaires de plateformes, leur exploitation éventuelle à des fins statistiques est susceptible de mettre à mal ces pratiques et la confiance qu'elles ont su relativement bien établir. Comme certains statisticiens l'ont observé, « les implications de l'utilisation des Big Data sur la perception d'un INS par le grand public sont de la plus haute importance, du fait de leur impact direct sur la confiance dans la statistique publique » (Struijs *et al.*, 2014). Alors que Struijs *et al.* affirment qu'il est possible d'atténuer ces risques en adoptant d'autres pratiques, comme « faire preuve de transparence quant à la nature et au mode d'utilisation des sources des Big Data », nous envisageons les insuffisances d'une telle approche, sans pour autant en contester la nécessité, à la lumière d'un autre enjeu de taille : l'idée que la réutilisation des données massives à des fins de statistique publique marque une rupture et une distanciation dans la relation entre les INS et les citoyens. Bien qu'imparfaites, les méthodes traditionnelles de collecte impliquent peu ou prou un rapport direct entre les INS et les citoyens, qui ancre les données dans la réalisation collective et le bien commun. Ce rapport permet au public une relative identification à ces données, par exemple, lorsque les enquêtes transmettent leurs connaissances et leurs expériences en répondant à des questions. Il contribue par la suite, selon notre hypothèse, à asseoir la confiance dans la statistique publique et la légitimité de celle-ci.

Cette proposition a été initialement formulée dans le cadre du projet « *Socialising Big Data* » (socialisation des données massives), qui comprenait des ateliers collaboratifs avec des statisticiens nationaux et internationaux et a débouché sur un projet de cadre social pour les Big Data (Ruppert *et al.*, 2015). Ce cadre

proposait des modèles d'appropriation sociale privilégiant les possibilités de partage, de collaboration et de coopération, et appréhendant les Big Data plutôt comme une ressource sociale et collective que privée. L'approche que nous exposons dans le présent article repose sur l'ambition de développer le concept de « données citoyennes » dans une logique de rattachement et d'appropriation sociale fondant de nouvelles relations avec des citoyens co-producteurs de données pour la statistique publique, et non pas sujets de plus en plus éloignés dont il faut gérer les perceptions et la confiance.

Comprendre ces nouvelles relations nous semble crucial à double titre. Tout d'abord, aux antipodes de certains usages qui définissent les données citoyennes comme des données concernant les citoyens, notre conception admet que les Big Data et les citoyens ne font qu'un : les données issues des technologies numériques sont les données des citoyens. Ensuite, des relations impliquant une participation plus directe du public sont nécessaires pour remédier à une autre conséquence de la distanciation, née de la réutilisation de données que génèrent par exemple les médias sociaux, les téléphones mobiles et les navigateurs Internet : le risque d'un écart grandissant entre les actions, les repères et les expériences des citoyens et les modes de catégorisation, d'inclusion et d'exclusion correspondants en termes statistiques, l'interprétation de ces données et l'identification des citoyens avec les statistiques obtenues¹. Nous qualifions de « grandissant » ce risque d'écart, car cet écart n'est pas entièrement nouveau, ni l'apanage du Big Data². L'ancien Directeur général d'Eurostat, Walter Radermacher, a évoqué plus généralement un écart entre les expériences des citoyens et la statistique publique, appelant une « statistique subjective »³. Ce faisant, il avait souligné la nécessité d'un débat plus démocratique entre les citoyens et les producteurs et propriétaires de données pour aboutir à une « compréhension différenciée et plus subjective de notre monde », plutôt que de voir « des technocrates et des politiciens se réunir à huis clos et faire face aux citoyens a posteriori »⁴. S'agissant de notre concept de « données citoyennes », cela exige des processus de co-production qui impliquent de produire les données nécessaires à l'élaboration de la statistique publique moyennant des relations directes avec les citoyens.

Les arguments aboutissant à notre concept de données citoyennes sont le fruit d'un travail ethnographique de terrain de plusieurs années mené avec des INS et deux organismes statistiques

internationaux (voir encadré et Grommé *et al.*, 2017). Cette recherche avait conduit à identifier quatre principes de « données citoyennes », sur la base de sujets de préoccupations exprimés par les statisticiens dans le cadre de notre travail de terrain. Nous considérons ces principes comme des sujets de préoccupation pour deux raisons. Tout d'abord, pour leur reconnaître la qualité de normativités qui influencent et orientent l'action des statisticiens et le développement de solutions pratiques (Boltanski & Chiapello, 2007). Ensuite, pour entreprendre une critique qui n'écarte pas les concepts des statisticiens, mais s'intéresse en premier lieu à la manière dont ces derniers conçoivent et définissent les concepts, avant d'envisager leur redéfinition (Latour, 2004). En d'autres termes, prendre en compte les préoccupations, et les suppositions, exprimées par les statisticiens ne signifie pas les partager, mais repenser ces préoccupations. Les quatre sujets de préoccupation qui nous sont apparus particulièrement importants par rapport à notre concept de données citoyennes sont l'expérimentalisme, les sciences participatives, les statistiques intelligentes et la protection de la vie privée dès la conception. Dans la section suivante de cet article, nous présentons chacune de ces préoccupations, en nous appuyant sur un ensemble de publications dans le domaine des sciences sociales pour les redéfinir, puis nous les présentons sous forme de principes des données citoyennes. Au cœur de notre redéfinition, l'idée que l'avenir de la statistique publique ne repose pas seulement sur l'exploitation des nouvelles technologies numériques, sources de données et invention de méthodes, mais également sur l'instauration d'une nouvelle relation au citoyen (Ruppert, 2018).

Les réflexions développées ici sur le concept de données citoyennes s'adressent principalement aux statisticiens, mais également aux chercheurs en sciences sociales, pour trois raisons principales. La première, c'est que nous

1. Par exemple, les expériences s'appuyant sur les données de téléphonie mobile pour modéliser la mobilité se heurtent à des obstacles dans leurs efforts pour interpréter la signification des schémas de déplacement.

2. Nous sommes conscients que les questions de représentation affectent également les méthodes statistiques établies. Le produit intérieur brut, par exemple, constitue l'une de ces statistiques publiques faisant vivement débat. L'économiste Joseph Stiglitz, de l'Université de Columbia, dénonce la « fétichisation » avérée du PIB en tant qu'indicateur « roi » de la performance d'une économie nationale, en dépit de ses insuffisances (Stiglitz *et al.*, 2009). Par conséquent, Fleurbaey (2009) suggère d'aller « au-delà du PIB » et préconise d'autres approches, y compris les évolutions récentes en matière d'analyse de la durabilité, le bonheur et la théorie du choix social et du partage équitable, ainsi que les études sur le bien-être social. Les indicateurs sur l'emploi se sont vu opposer des arguments similaires, s'agissant notamment des personnes occupant des formes d'emploi irrégulier (cf. Hussmanns, 2004).

3. Notes de terrain, Eurostat Agility Conference, novembre 2016.

4. Notes de terrain, Eurostat Agility Conference, novembre 2016.

ENCADRÉ – Le projet de recherche

Notre concept de « données citoyennes » est le fruit de plusieurs années d'un travail ethnographique mené sur le terrain auprès de cinq INS et de deux organisations statistiques internationales, qui a consisté à savoir observer des conférences et des réunions, faire le suivi et l'analyse de publications, mener des interviews et échanger avec des statisticiens. Plus précisément, cet article s'appuie sur un document de travail ARITHMUS de Grommé *et al.* (2017) et en résume les points clés. ARITHMUS (*Peopling Europe : How data make a people*), un projet financé par l'ERC (European Research Council), a débuté en 2014 avec une équipe de 6 chercheurs : Evelyn Ruppert (directrice du projet), Baki Cakici, Francisca Grommé, Stephan Scheel et Funda Ustek-Spilda (chercheurs post-doctorants), et Ville Takala (doctorante).

Nous avons suivi les pratiques de travail de cinq INS (Office for National Statistics du Royaume-Uni, Statistics Netherlands, Statistics Estonia, l'institut statistique de Turquie et Statistics Finland) et de deux organisations internationales (Eurostat et UNECE). Entre autres, nous avons

notamment suivi les débats des statisticiens à propos des expérimentations menées sur les technologies numériques et les Big Data, et leur incidence sur la statistique publique. Sur la base de ce travail de terrain, nous avons organisé deux ateliers avec un groupe consultatif de statisticiens pour échanger sur certaines de nos analyses, comme l'évolution du rapport entre les INS et les citoyens avec ces nouvelles technologies et les sources de données massives. Cela a débouché sur un document de travail synthétisant certains des arguments développés dans cet article, et introduit le concept de données citoyennes, qui a été examiné dans le cadre du groupe consultatif (cf. Grommé *et al.*, 2017). Cet examen a abouti à un atelier collaboratif avec ce groupe et un groupe élargi de statisticiens, de chercheurs universitaires et de concepteurs d'information, portant sur l'élaboration de principes de conception pour la co-production d'une application de données citoyennes. Plutôt que de synthétiser le matériau empirique issu de notre ethnographie et des ateliers, notre ambition est ici d'explicitier la notion de données citoyennes que nous avons développée dans le prolongement de cette recherche.

mobilisons des concepts et des perspectives des sciences sociales pour aborder les sujets de préoccupation exprimés par les statisticiens. Nous contribuons ainsi, de manière plus globale, aux méthodes de recherche en sciences sociales. La deuxième raison, c'est que le concept de données citoyennes et les principes qui s'y rattachent valent également pour les débats sur les méthodes de recherche s'appuyant sur les technologies numériques et les sources de Big Data qui animent le champ des sciences sociales. Autrement dit, si les problématiques et les objectifs de la recherche en sciences sociales diffèrent, la relation au citoyen dans le cadre de la production des savoirs relève d'une préoccupation partagée. La troisième, reflétée dans notre démarche de recherche qui comportait des ateliers avec des statisticiens, est qu'un concept de données citoyennes appelle des collaborations expérimentales non seulement avec les citoyens, mais également entre spécialistes des sciences sociales et statisticiens.

L'expérimentalisme

L'expérimentalisme est le premier sujet de préoccupation rencontré dans le cadre de notre travail de terrain. Pour les organismes gouvernementaux et les entreprises, l'expérimentation fait partie intégrante de l'innovation. La statistique publique en est une bonne illustration, comme l'atteste le développement de laboratoires d'innovation, de *sandboxes*, hackathons et de projets de

recherche exploratoire⁵. Pour les statisticiens, les expérimentations faisant appel aux nouvelles technologies numériques et aux Big Data sont un moyen de développer des modes de pensée, des techniques et des compétences inédits en matière de production de statistique publique. En outre, diverses composantes du champ des sciences sociales se tournent vers l'expérimentalisme. Toutefois, le choix de l'expérimentation comme méthode pour ouvrir l'expertise scientifique et technologique à différents acteurs dans l'optique de générer de nouveaux modes de pensée est, lui, relativement récent. Dans des domaines aussi différents que la conception de fauteuils roulants, les Big Data ou la biologie synthétique, les spécialistes en sciences sociales ont adopté l'expérimentalisme pour créer de nouveaux espaces de formulation des problèmes, collaborer avec différents acteurs et étudier différents scénarios⁶. En d'autres termes, un postulat majeur est que les modes expérimentaux de collaboration sont de nature à engendrer de nouveaux modes de pensée.

Globalement, on distingue deux modèles permettant aux expériences collaboratives d'y prétendre. Le premier consiste en diverses formes de participation visant à atteindre un certain degré de démocratisation, en ouvrant au public l'accès aux débats scientifiques et techniques et aux processus (Marres, 2012). Le second repose

5. Voir, par exemple, les statistiques expérimentales produites par Eurostat : <http://ec.europa.eu/eurostat/web/experimental-statistics/>.

6. Pour ces trois exemples, voir : <https://entornoalasila.wordpress.com/english/> ; Ruppert *et al.* (2015) ; et <http://www.anthropos-lab.net/about>.

sur l'expérimentation collaborative dans l'optique de développer et d'explorer de nouvelles formulations de problèmes, de transcender des modes de raisonnement solidement ancrés, de bousculer les hiérarchies existantes et d'examiner de façon critique le processus de création des savoirs (Rabinow & Bennett, 2012). C'est le modèle du « collaboratoire » (ou co-laboratoire), selon lequel les participants explorent un thème en commun. Le projet « *Socialising Big Data* » précédemment mentionné a fait usage de ce modèle en animant des ateliers et des discussions avec des statisticiens, des chercheurs en génomique et des ingénieurs en gestion des déchets de divers pays pour définir et élaborer des concepts partagés permettant d'appréhender les Big Data (Ruppert *et al.*, 2015). Autre forme de collaboration : co-produire une « chose » – soit un produit fini tangible –, au travers de laquelle les collaborateurs explorent et développent en pratique des concepts et des problématiques qu'ils partagent. Travailler à un produit commun permet « une disponibilité expérimentale des problématiques dans une mesure telle que le « possible » devient tangible, façonnable et accessible » (Binder *et al.*, 2015, p. 12). Du point de vue méthodologique, cela oblige les participants à préciser les futurs modes de travail (Muniesa & Linhardt, 2011). En général, l'étude sociale des sciences nous enseigne que les expériences collaboratives de ce type exigent aussi de redessiner les relations entre participants, technologies et savoirs. Il s'agit là, par ailleurs, d'un des principes de ce qu'il est convenu d'appeler dans le champ des sciences sociales et humaines « la recherche fondée sur la pratique », qui implique, en dehors du texte et de la parole, un rapprochement entre les participants et les compétences, les matériaux, les tâches mineures et le travail de tous les jours inhérents à la fabrication des choses (Jungnickel, 2017). Fabriquer des choses, plutôt que les décomposer ou les défaire, nécessite de se mêler étroitement aux différents participants et peut enrichir la compréhension des compétences, des relations et des infrastructures qui composent un produit fini.

L'expérimentalisme est tout particulièrement perçu comme une approche fondamentale face à l'incertitude et au changement. À titre d'exemple, dans un article portant sur la collaboration entre universitaires, agriculteurs et écologistes, Waterton et Tsouvalis (2015, p. 477) se posent la question de savoir comment « la politique de la nature peut s'envisager dans une ère consciente de la complexité, de la contingence et de la relationnalité du monde ». Ils analysent une collaboration entre eux-mêmes, en leur qualité de spécialistes en sciences sociales,

et des experts environnementaux, ainsi que des agriculteurs. Dans le cadre de leur expérience, la recherche partagée a amené des questions sur la manière d'appréhender la pollution de l'eau : causes isolées ou bien rapports ou antécédents socio-techniques au sens large ? Ils ont ainsi adopté un programme d'expérimentation qui conçoit la production des savoirs comme nécessitant des « forums hybrides » (Callon *et al.*, 2011) ou des « nouveaux collectifs » (Latour, 2006), au sein desquels les participants s'engagent de manière réflexive dans la reconstruction des relations, des histoires et des parties prenantes impliquées dans une problématique. L'incertitude n'est pas quelque chose qui appelle une résolution, mais plutôt une reconnaissance et un apprivoisement dans le cadre d'un processus collectif continu de production des savoirs. Dans la pratique, il s'agit d'une approche bienveillante et prudente dite « *care-full* » (Grommé, 2015) qui implique : l'exercice de responsabilités pour le suivi et la documentation de ce qui est (inévitablement) inclus et exclu ; d'éviter les ambiguïtés quant aux modalités d'évaluation en précisant de façon explicite comment les résultats sont évalués ; d'admettre qu'un échec s'explique vraisemblablement par une myriade de facteurs ; et de comprendre que les valeurs sont indissociables des faits. « *Care-full* » ne renvoie donc pas uniquement à une approche prudente, mais aussi à la reconnaissance du fait que les expérimentations remodelent continuellement les relations et en redistribuent les effets de manière parfois inattendue.

En tant que principe des données citoyennes, l'expérimentalisme fait donc non seulement appel à l'expérimentation, mais également à la collaboration pour ouvrir les modes de pensée et la production des savoirs à l'influence et à la réflexion d'autrui et, ce faisant, imaginer et réfléchir à des alternatives et des possibilités (Stengers, 2010). Il impose de rendre compte et de répondre des procédures et des pratiques relatives aux expériences. Enfin, il signifie être réceptif à une organisation potentiellement différente des relations entre les divers participants associés à la construction des savoirs. Dans la perspective de nouvelles relations entre les citoyens et les INS qui est la nôtre, l'expérimentalisme implique donc des formes actives et ouvertes de participation et d'influence. Nous approfondissons cette idée avec un deuxième principe, celui de la « science citoyenne », pour explorer comment les relations entre les INS et les citoyens dans le cadre l'élaboration des données et des statistiques officielles pourraient être redéfinies plus avant.

La science citoyenne

En matière de production de données, certains organismes statistiques ont commencé à expérimenter des modèles de participation citoyenne. Ces modèles s'inspirent souvent de conceptions des sciences participatives, dont nous ferons ici un bref examen avant d'envisager leur possible redéfinition. À l'inverse des approches traditionnelles qui considèrent généralement les citoyens comme de simples répondants, divers modèles de sciences participatives voient en eux non seulement des sujets d'étude, mais aussi des acteurs de la production de données. S'agissant de l'élaboration de données, les définitions et les interprétations relatives aux sciences participatives et aux modalités de participation citoyenne ne manquent pas. La Commission européenne, par exemple, les définit comme la « production de connaissances hors du cadre scientifique professionnel, souvent désignée par connaissances générales, locales ou traditionnelles » (Commission européenne, 2013, p. 5). Goodchild (2007) utilise cette dénomination pour décrire les communautés ou les réseaux citoyens qui interviennent en tant qu'observateurs dans un domaine scientifique quelconque. C'est la définition la plus communément acceptée, comme l'atteste en particulier l'importante dynamique enregistrée par les sciences participatives dans le domaine des sciences naturelles ces dernières années (Kullenberg & Kasperowski, 2016, p. 2). Cela dit, la pratique consistant à associer des membres du public à la collecte et à la transmission de données à des fins scientifiques date au moins des années 1960, bien qu'il ait fallu attendre les années 1990 pour que le terme en lui-même entre dans l'usage (*ibid.*)⁷.

Une variante de cette approche fait intervenir les citoyens non seulement en qualité d'observateurs, mais également en tant que co-producteurs ou producteurs de données et d'études scientifiques traduisant leurs propres préoccupations, besoins et interrogations. Cette variante comprend des approches locales et militantes que l'on désigne par « audit communautaire », « science civique », « surveillance environnementale communautaire », « sciences de la rue », « épidémiologie populaire », « sciences citoyennes » et « science *do-it-yourself* » (Kullenberg & Kasperowski, 2016, p. 2). Ces approches accueillent aussi bien les citoyens en quête d'alliances étroites avec les établissements scientifiques et les centres de connaissance que les citoyens participant à la production de savoirs indépendants aux côtés de scientifiques.

La participation des citoyens à la production de données scientifiques répond à des objectifs multiples, allant de la documentation des préoccupations sur les questions environnementales à la création de cartes d'archives de sites historiques locaux, ou encore à la transcription de textes des contemporains de Shakespeare⁸. Goodchild (2007) laisse entendre que les personnes qui participent et partagent des informations sur Internet en général sont plus disposées à communiquer volontairement des renseignements géographiques et à contribuer aux initiatives de collecte de données type « OpenStreetMap ». Sur cette base, il fait valoir que deux types d'individus sont susceptibles de participer : ceux qui cherchent à faire leur auto-promotion et communiquent sciemment des renseignements personnels sur Internet afin de les rendre « accessibles à des amis et des relations, indépendamment du fait qu'ils deviennent alors accessibles à tous ; et ceux qui recherchent la satisfaction personnelle que procure la transmission d'informations anonymes et l'apparition de celles-ci au sein d'une mosaïque de contributions collectives en devenir » (Goodchild, 2007, p. 219).

Sheila Jasanoff (2003, pp. 235–236) remarque que les modèles des sciences participatives peuvent faciliter une interaction fructueuse entre les décideurs, les experts scientifiques, les entreprises et l'opinion publique. Elle soutient que la pression pour la responsabilisation des experts dans la prise de décisions est manifeste dans la revendication de transparence accrue et de participation plus large. Toutefois, les initiatives participatives ne peuvent prétendre assurer seules la gouvernance démocratique et représentative des sciences et de la technologie. Jasanoff souligne que les États modernes ont focalisé leur attention sur l'amélioration des « technologies de l'hubris », conçues pour faciliter la gestion et le contrôle en faisant peu de cas de l'incertitude, des objections politiques et des complexités imprévues de la vie quotidienne (*ibid.*, p. 238). Ce n'est pas seulement la connaissance qui fait défaut, mais les moyens de faire intervenir des méthodes, et des processus incertains et inconnus, dans la dynamique du débat démocratique (*ibid.*, pp. 239–240). C'est pourquoi Jasanoff met en avant les sciences participatives comme modèle possible de concertation démocratique entre les différentes parties prenantes de la production scientifique. Ainsi, on peut considérer

7. Certains chercheurs incluent dans cette définition le recensement annuel des oiseaux de Noël de la National Audubon Society du début des années 1900, à l'occasion duquel des citoyens ont participé à l'observation et au comptage des espèces aviaires.

8. Certains de ces exemples sont documentés sur www.zooniverse.org.

les sciences participatives comme des « technologies de l'humilité », autrement dit, des technologies *sociales* (en italique dans le texte original) qui impliquent des relations entre les administrations, les décideurs, les experts et les citoyens au niveau de la gestion des technologies pour « évaluer l'inconnu et l'incertain, des “évaluations modestes” » qui mobilisent les citoyens en tant qu'agents actifs de la connaissance, des idées et de la mémoire (*ibid.*, p. 243).

L'une des préoccupations sur le rôle des non-scientifiques dans la production savante concerne les implications vis-à-vis de principes scientifiques établis⁹. Cependant, comme le démontre Goodchild (2007), bien que les sciences participatives ne répondent pas toujours, *stricto sensu*, aux critères scientifiques en soi, elles peuvent ouvrir la voie à de nouveaux modes de pensée et d'appréhension des données. C'est particulièrement pertinent pour les pratiques de démocratisation qui appellent différentes formes de raisonnement, comme l'illustre dans le domaine de la prise de décisions la notion de « *satisficing* » d'Herbert Simon (1947), par opposition à la « maximisation » ou à l'« optimisation ». À contre-pied d'abstractions comme la théorie de l'utilité, il prône une interprétation basée sur le mode de raisonnement pratique des individus. Le raisonnement pratique, fait-il valoir, consiste à jongler avec une foule de critères et à parvenir à une solution « suffisamment bonne », plutôt que de se lancer dans une quête sans fin de toutes les options possibles, de les évaluer, puis d'aboutir à la meilleure. Gabrys et Pritchard (2015) empruntent une approche similaire pour suggérer que l'adéquation d'une réponse dépend de la dimension pratique des questions. En lieu et place, elles définissent le concept de « données tout juste suffisantes » pour contrecarrer la toute-puissance de la précision de mesure en tant qu'unique objectif et critère pour évaluer les données environnementales recueillies par le biais de pratiques de télé-détection citoyenne. La mesure des phénomènes environnementaux s'inscrit dans des objectifs ou des questions de natures différentes, souvent méconnus au départ. Par exemple, il est possible qu'une mesure « approximative » visant à identifier un épisode de pollution, en cours ou terminé, soit suffisante et « suffisamment bonne ». Ce que Gabrys et Pritchard mettent en évidence, c'est que, la plupart du temps, les utilisations potentielles ou la valeur des données ne sont pas connues à l'avance, et qu'il est utile d'instituer la production et l'interprétation des données comme pratiques de recherche de potentiel plutôt que de répéter et de reproduire

des objectifs ou des questions déjà connus avec des méthodes précédemment établies.

Parmi les expériences reposant sur des modèles de participation citoyenne menées récemment par les organismes statistiques figure un projet pilote de l'INS Canadien qui utilise OpenStreetMap (OSM) dans le cadre d'une opération participative visant à combler les lacunes des données de géolocalisation (Statistics Canada, 2016)¹⁰. OSM est une initiative collaborative visant à créer une carte du monde gratuite et modifiable. L'application OSM de Statistics Canada permet aux utilisateurs de sélectionner une géolocalisation et de modifier le nom d'une rue, par exemple. Un autre exemple est le Centre commun de recherche de la Commission européenne sur les sciences et données ouvertes citoyennes, qui a étudié des modèles possibles de participation citoyenne pour le suivi de la propagation de espèces exotiques envahissantes (EEE) (Cardoso *et al.*, 2017). Le rapport a estimé que la mise en œuvre du règlement EEE pouvait bénéficier de contributions citoyennes pour la transmission « d'informations précises, détaillées et opportunes relatives à l'apparition et à la répartition d'EEE afin de permettre une prévention efficace, une détection précoce, une intervention rapide, ainsi qu'une évaluation des mesures de gestion » (p. 5). De plus, cette forme de participation citoyenne pourrait sensibiliser et renforcer le soutien de l'opinion publique en faveur de la réglementation, tout en aidant les citoyens à acquérir des compétences et une meilleure compréhension des travaux scientifiques (Socientize Consortium, 2014). Les Nations Unies ont également jugé, pour les questions environnementales, que la production de données issues des sciences participatives était nécessaire à la mesure et au suivi des objectifs de développement durable (ODD) (ONU, 2016). Les modes de participation citoyenne sont perçus comme essentiels pour que l'Agenda pour le développement durable 2030 soit maîtrisé par le pays concerné, adapté au contexte et assorti d'objectifs en lien avec les valeurs et les priorités nationales. Bien que ces initiatives envisagent la participation citoyenne sous des angles différents, elles limitent généralement celle-ci à des tâches telles que la production, la vérification et/ou la classification

9. Voir aussi Gabrys et al. (2016) pour des réflexions concernant la qualité et la crédibilité des données.

10. Le projet pilote a été organisé par Statistique Canada en collaboration avec OpenNorth, MapBox, la Ville d'Ottawa et OSM Canada. OpenNorth est un organisme à but non lucratif qui élabore des outils numériques au service de la participation civique.

de données. Ces formes de sciences participatives ont, par conséquent, fait l'objet de critiques dénonçant une exploitation des citoyens comme de la main d'œuvre publique gratuite (DataShift, s. d. ; Piovesan, 2017 ; Paul, 2018). Ces critiques font observer que les tâches de nettoyage, de codage ou d'analyse des données, tout comme celles de conception, d'architecture ou d'interprétation, sont réservées aux experts, tandis que les citoyens sont cantonnés au rôle de sujets d'étude ou d'assistants de recherche.

Notre redéfinition des sciences participatives est plus proche de ce que Jasanoff décrit comme la production inclusive des savoirs. Mais, dans la lignée de notre argumentaire sur la distanciation, nous suggérons que l'inclusivité comprend le droit d'émettre des revendications et d'exprimer des préoccupations sur la façon dont il convient de catégoriser et de qualifier les questions environnementales, économiques et sociales¹¹. C'est sans doute là la revendication précise qu'émettent les « scientifiques citoyens » lorsqu'ils collaborent à la production indépendante de données ou enrichissent les connaissances scientifiques et officielles. Néanmoins, notre concept de données citoyennes n'envisage pas les citoyens comme autonomes, mais comme co-producteurs. À ce titre, nous concevons les données citoyennes comme reposant sur des nouvelles relations entre les citoyens et les INS mariant sciences statistiques et sciences participatives. Une telle approche pourrait supposer la participation des citoyens à tous les stades de la production, pour aboutir à une statistique plus représentative et inclusive de leurs préoccupations, de leurs besoins et de leurs expériences, sans oublier leurs propres repères. Cela demanderait une approche souple et expérimentale en termes de critères (Paul, 2018) pour pouvoir s'adapter à l'évolution des besoins et des exigences des citoyens, certes, mais également de ce qui leur tient à cœur. Comme nous le proposons ci-dessous, cela implique une conception élargie de l'éthique, qui va au-delà du consentement, de l'équité et de la protection des données, pour s'orienter vers ce qui est sans doute le moteur de l'ascension des sciences participatives : des citoyens contribuant activement à l'élaboration et au façonnage des données servant à produire la statistique et la connaissance publiques. Dans la section suivante, nous examinons les implications potentielles d'une telle conception de l'éthique dans une autre perspective : celle de propositions pour des « statistiques intelligentes ».

Les statistiques intelligentes

Les propositions d'Eurostat pour l'élaboration de « statistiques intelligentes » s'appuient sur l'idée des « villes intelligentes », qui regroupe en général l'utilisation des Big Data, des capteurs urbains, de l'internet des objets (IoT, *Internet of Things*) et d'autres formes de production et d'intégration de données visant à simplifier la gouvernance municipale et les infrastructures de transport, à régénérer l'économie locale, à transformer l'environnement urbain pour le rendre plus durable, vivable et socialement inclusif (voir par exemple Henriquez, 2016). Si les villes intelligentes font l'objet de définitions en tous genres, le concept renvoie généralement, d'une part, « à l'informatique envahissante et omniprésente qui caractérise de plus en plus la composition et la surveillance des villes, dont, d'autre part, l'économie et la gouvernance sont portées par l'innovation, la créativité et l'esprit d'entreprise décrétés par des individus intelligents » (Kitchin, 2014). Vues sous cet angle, les Big Data permettent d'analyser la vie urbaine en temps réel, de faire émerger des modes de gouvernance urbaine nouveaux, et d'imaginer et de bâtir des villes plus efficaces, plus durables, plus compétitives, plus productives, plus ouvertes et plus transparentes.

L'un des objectifs des propositions en faveur de « statistiques intelligentes » avancées par le groupe de travail sur le Big Data d'Eurostat consiste à tirer parti des « systèmes intelligents » tels que l'énergie intelligente, les compteurs intelligents, les transports intelligents et ainsi de suite. Ces propositions visent à exploiter le potentiel associé à la prolifération des périphériques numériques et des capteurs connectés à Internet, et la manière d'intégrer les données qu'ils génèrent aux systèmes de production statistique pour pouvoir produire des statistiques en « temps réel » et « automatiquement »¹². Dans cette perspective, il est envisagé d'intégrer la collecte, l'analyse et le traitement des données dans des activités qui génèrent et analysent simultanément les données. L'adoption d'une telle approche pourrait transformer radicalement le mode de production de la statistique publique et invite à repenser les processus et les architectures opérationnels, les lois et les règlements, l'éthique, les méthodologies, etc.

11. Il s'agit d'une conception avancée dans le champ d'étude de la citoyenneté critique, dont Isin & Ruppert (2015) et Isin & Saward (2013) font la synthèse. Être citoyen s'entend comme une subjectivité politique, qui suppose non seulement la possession de droits, mais également le droit de revendiquer des droits, comme le droit des citoyens de façonner les données produites les concernant, ainsi que les populations auxquelles on les rattache (Ruppert, 2018).

12. « Statistiques intelligentes », groupe de travail d'Eurostat sur le Big Data. Projet de document, octobre 2016.

Pour générer des statistiques intelligentes, deux approches ont été proposées dans cette logique : utiliser des systèmes tiers à vocation autre que statistique, mais depuis lesquels il est possible d'extraire de l'information statistique (les téléphones mobiles, par exemple) ; ou mettre au point des pratiques de production de données entièrement nouvelles, comme des capteurs et des périphériques numériques exclusivement destinés à la production d'informations statistiques¹³. L'approche privilégiant le recours à un tiers est à l'origine de nombre des préoccupations précédemment identifiées, telles que l'accès aux données et leur propriété, le respect de la vie privée et de l'éthique, la représentativité et la qualité des données, ainsi que la confiance et une distanciation accrue entre les citoyens et les INS. En revanche, la seconde approche consistant à concevoir de nouveaux dispositifs de production de données offre la perspective d'atténuer ces problèmes. Autrement dit, notre redéfinition des statistiques intelligentes exige non seulement de repenser les aspects techniques et organisationnels des systèmes de production statistique, mais aussi leur relation au citoyen. Comme indiqué dans nos observations sur les sciences participatives, cela pourrait impliquer des modèles de production qui favorisent la participation des citoyens à tous les stades de la production de statistiques intelligentes.

Toutefois, cela signifierait adopter l'approche « *care-full* » décrite plus haut, y compris une conception élargie de l'éthique tout au long du processus de production de la statistique publique. L'éthique figure évidemment depuis longtemps au cœur des principes fondamentaux de la statistique publique, qui intègrent des valeurs telles que l'utilité, les normes professionnelles et règles déontologiques, les principes scientifiques, la transparence, la qualité, l'actualité, les coûts, la charge pour les enquêtés et la confidentialité (ONU, 2014)¹⁴. Ces principes constituent ce que l'on pourrait qualifier d'« éthique du *care* appliquée aux données », à savoir le souci de la qualité, de l'accessibilité et de la clarté des données, mais également celui de la relation au citoyen et des responsabilités à son égard, à travers des pratiques telles que la protection des données, la confidentialité, le consentement et la confiance. Bien que ces principes trouvent leurs racines dans un imbroglio de logiques et d'exigences juridiques, étatiques, politiques et professionnelles, ils ont tendance à s'inscrire dans les valeurs et les engagements professionnels de tous les jours, ce qui ressort des déclarations des statisticiens à propos de l'utilisation des sources de Big Data : « pouvoir faire n'est pas devoir faire. »

Les principes fondamentaux de la statistique publique manifestent ainsi une conception élargie de l'éthique qui intègre la relation au citoyen, ou « éthique procédurale », selon le terme consacré par la recherche en sciences sociales (Guillemin & Gillam, 2004), qui renvoie à l'appréciation de problèmes éthiques pouvant survenir dans le cadre de la recherche ou de la production de données. Cependant, Guillemin et Gillam identifient une seconde dimension en matière d'éthique dans la recherche, qu'ils nomment « l'éthique en pratique » (*ibid.*, p. 261). Celle-ci a trait aux moments éthiques récurrents, itératifs et incertains qui émaillent la recherche et peuvent aller à l'encontre des résultats d'une évaluation d'éthique procédurale. Cette dernière est pertinente dans le cadre de pratiques de co-production de statistiques intelligentes, qui, par définition, supposent incertitude, adaptation et réactivité face aux interactions, aux intérêts et aux demandes des diverses parties prenantes. En cela, la co-production exige une éthique de la responsabilité qui admette et prenne en compte la dépendance vis-à-vis de la relation au citoyen et ses apports en matière « de création, de détention et de pérennité des données » (Puig de la Bellacasa, 2012, p. 198).

Le concept de données citoyennes que nous proposons redéfinit donc les statistiques intelligentes comme impliquant de nouvelles relations avec des citoyens co-producteurs de plateformes de production de données. Il nécessite une approche « *care-full* » qui offre une conception élargie de l'éthique associant les demandes, les intérêts et les contributions des citoyens à tous les stades du développement des nouveaux dispositifs de production de données, plutôt qu'*a posteriori* en aval ou en mode correctif. À ce titre, c'est un modèle qui s'appuie sur la protection de la vie privée dès la conception, autre sujet de préoccupation qui aborde les questions du respect de la vie privée et du consentement dans la conception logicielle en amont, que nous traitons maintenant.

13. *Ibid.* Un exemple possible est la collecte par l'INS des Pays-Bas (ECB) de données à vocation statistique relatives à l'intensité du trafic routier qui sont générées exclusivement au moyen de capteurs routiers. Voir : <https://www.cbs.nl/en-gb/our-services/innovation/nieuwsberichten/recente-berichten/new-steps-in-big-data-for-traffic-and-transport-statistics>.

14. Six principes stipulent que la statistique publique doit : satisfaire à un critère d'utilité pratique ; être déterminée en fonction de considérations purement professionnelles, de principes scientifiques et de règles déontologiques ; fournir, en fonction de normes scientifiques, des informations sur les sources, les méthodes et les procédures qu'elle utilise ; pouvoir être générée à partir de toutes sortes de sources, qu'il s'agisse d'enquêtes statistiques ou de fichiers administratifs, la source étant choisie en tenant compte de la qualité, de l'actualité, des coûts et de la charge des répondants ; être strictement confidentielle et n'être utilisée qu'à des fins statistiques ; et porter à la connaissance du public les textes législatifs et réglementaires, et toutes dispositions la régissant.

La protection de la vie privée dès la conception

Qui dit sources de Big Data et nouvelles sources de données dit nouvelles interrogations en matière de respect de la vie privée, de consentement et de confidentialité, auxquelles les cadres réglementaires existants ne répondent pas toujours en détail. Ainsi, la protection de la vie privée dès la conception est devenue un sujet de préoccupation pour les INS. La protection de la vie privée dès la conception renvoie au fait d'intégrer la protection de la vie privée dans la conception logicielle de plateformes de production de données, de périphériques ou d'applications. Elle implique d'avoir le citoyen à l'esprit dès le départ et de mettre en œuvre les conceptions obtenues de manière transparente. De ce fait, la protection de la vie privée dès la conception constitue une réponse logicielle aux problèmes de respect de la vie privée, de consentement et de confidentialité, pouvant s'utiliser de pair avec d'autres dispositifs tels que les analyses d'impact sur la vie privée. La protection de la vie privée dès la conception permet de traiter les questions liées au respect de la vie privée dès l'entame du processus de conception, contrairement à d'autres approches qui visent, elles, à résoudre de telles questions une fois la phase de développement logiciel terminée, ou bien font de cette problématique l'affaire du cadre juridique ou réglementaire.

Cavoukian *et al.* (2010) définissent la protection de la vie privée dès la conception à l'aide de sept principes fondamentaux : mesures proactives et non réactives, et mesures préventives et non correctives ; protection implicite de la vie privée ; intégration de la protection de la vie privée dans la conception des systèmes et des pratiques ; fonctionnalité débouchant sur un résultat à somme positive et non à somme nulle ; sécurité de bout en bout ; visibilité et transparence ; et respect de la vie privée des utilisateurs. Ces principes imposent aux conceptions de respecter la vie privée dès le départ et de limiter la production de données dans le respect des attentes des citoyens. Ils prévoient aussi que la production d'applications générant les données tienne compte du fait que les données survivent probablement au logiciel. Les auteurs insistent également sur la nécessité de prendre en compte le cycle de vie du logiciel au moment de déterminer le meilleur moyen de protéger la vie privée, et notamment de prendre des dispositions quant à la suppression des données une fois l'application en fin de vie. Enfin, ces mêmes principes contraignent les organisations qui traitent des

données personnelles à faire preuve de transparence sur les finalités, et à demeurer redevables envers les citoyens.

La production et le traitement de données à caractère personnel, en revanche, posent de nombreux autres défis, aussi bien en termes de respect de la vie privée que de protection des données individuelles. Selon Nissenbaum (2004), les normes de protection de la vie privée doivent s'inscrire dans des contextes spécifiques. Elle présente trois principes qui ont dominé les débats sur le respect de la vie privée tout au long du 20^e siècle, à savoir : limiter la surveillance des citoyens par les pouvoirs publics, restreindre l'accès aux informations personnelles et empêcher les intrusions dans les espaces privés. Elle propose un nouveau terme, l'« intégrité contextuelle », pour faire face aux nouveaux défis qu'engendrent les technologies numériques. L'intégrité contextuelle suppose une collecte d'information adaptée au contexte, ainsi que le respect des normes intrinsèques qui en régissent la circulation. L'idée directrice est que les normes de circulation de l'information varient en fonction des cultures, des périodes historiques, des lieux, ainsi que d'autres facteurs. En outre, l'intégrité contextuelle exige de connaître non seulement le lieu précis de production de données, mais également la pertinence des institutions sociales qui y sont associées (Nissenbaum, 2009).

Les approches visant à protéger les données individuelles peuvent produire malgré tout, dans le cadre de la production de données à grande échelle, des résultats indésirables. Lorsque des données ayant fait l'objet d'une anonymisation individuelle sont liées pour créer des profils, les personnes correspondant à un profil donné peuvent en subir les effets, même si elles ne peuvent pas être identifiées personnellement. Graham (2005), par exemple, explique comment on peut utiliser un logiciel pour catégoriser différentes zones urbaines en fonction de la réussite scolaire, du prix de l'immobilier, du taux de criminalité, etc., ce qui peut éventuellement provoquer des inégalités et des discriminations entre les habitants, quand bien même ils ne sont pas personnellement identifiés. De même, Zwitter (2014) a distingué et problématisé le potentiel discriminatoire d'« effets de groupe », comme dans des profilages à partir de données anonymisées.

L'utilisation des Big Data amène également d'autres défis en matière de respect de la vie privée. Barocas et Nissenbaum (2014) sont d'avis que l'anonymat et le consentement sont souvent

compromis dans les applications des Big Data, et que d'autres approches sont nécessaires afin de protéger l'intégrité des données, à la manière des stratégies fondées sur des principes moraux et politiques qui répondent à des objectifs et des valeurs contextuels précis. Plutôt que de se focaliser sur l'anonymat dans les applications des Big Data, ils mettent l'accent sur l'obtention d'un consentement éclairé, non seulement pour donner le choix aux personnes concernées de retirer ou non leur consentement, mais également pour obliger les collecteurs de données à justifier leurs actions à l'aune des règles, des normes et des attentes. Dans une certaine mesure, ce point est couvert par le Règlement général sur la protection des données (RGPD) récemment mis en œuvre dans les États membres de l'Union européenne, qui s'appuie sur une conception élargie des données à caractère personnel et de la vie privée, et mettra fin aux pratiques de consentement général par défaut en ce qui concerne la production de données personnelles¹⁵. Le RGPD oblige l'ensemble des parties prenantes publiques et privées qui sollicitent, détiennent ou archivent les données personnelles à réfléchir à la définition des « données à caractère personnel », ainsi qu'aux pratiques éthiques nécessaires à leur traitement, étant donné la complexité et la connectivité des systèmes de données, et la non-neutralité avérée des algorithmes. En somme, le respect de la vie privée n'est pas une entité isolée, mais dépend de l'environnement de production, de la redevabilité à l'égard des effets de groupe et des mécanismes de consentement éclairé.

Récemment, des chercheurs s'attellent à résoudre les défis techniques du respect de la vie privée avec les Big Data ont proposé une méthode de protection mettant à profit la technologie de la « chaîne de blocs », ou *blockchain* (Montjoye *et al.*, 2014 ; Zyskind *et al.*, 2015). La *blockchain* est une méthode de calcul décentralisé, dans laquelle de nombreux périphériques communiquent entre eux sur un réseau partagé, sans qu'un serveur central soit nécessaire pour autoriser la participation de chacun des membres ou la tenue d'une liste des membres connectés. En appliquant la technologie de la *blockchain* au respect de la vie privée, il devient possible de chiffrer et de diffuser des données privées sur un vaste réseau sans avoir recours à un serveur central sécurisé.

Les méthodes de protection de la vie privée reposant sur la *blockchain* ont vocation à résoudre les défis sous-jacents au respect de la vie privée en s'appuyant sur un cadre technique pendant la

phase de développement logiciel. Néanmoins, comme nous l'avons mentionné plus haut, elles n'incarnent pas à elles seules l'unique solution pour garantir la protection de la vie privée, venant plutôt, *via* la conception logicielle, en complément de considérations telles que l'intégrité contextuelle, les effets de groupe et les modes du consentement. Ainsi, notre redéfinition de la protection de la vie privée dès la conception va-t-elle au-delà de l'aspect logiciel pour intégrer le droit du citoyen au respect de sa vie privée, tant dans la conception logicielle en amont que dans les relations avec un citoyen co-producteur de statistique publique à tous les stades de sa production. Autrement dit, tout comme l'éthique, la question du respect de la vie privée relève du processuel et ne saurait se régler uniquement par l'octroi d'un consentement ponctuel ou la conception logicielle, ni sans tenir compte de contextes précis.

* *
*

Nous avons, en définitive, repris les sujets de préoccupation exprimés par les statisticiens, puis procédé à leur redéfinition sous forme de principes des données citoyennes. À travers l'analyse des quatre principes que constituent l'experimentalisme, la science citoyenne, les statistiques intelligentes et la protection de la vie privée dès la conception, nous avons examiné la manière dont les données citoyennes peuvent renouveler les relations entre les citoyens et les INS, entre les actions, les repères et les expériences des citoyens, et les catégories, inclusions et exclusions correspondants en termes statistiques. Nous soutenons à cet égard que les données citoyennes ont le potentiel d'engendrer de nouvelles variables statistiques souhaitées et définies par les citoyens, de renforcer l'identification de ces derniers avec la statistique publique et aussi, éventuellement, de faire évoluer leur usage de la statistique. Car c'est bien là, de fait, un effet collatéral possible de la co-production de statistiques avec les citoyens, selon des procédés plus adaptés à leurs expériences et à leurs connaissances.

Notre concept de données citoyennes trouve son sens dans la prolifération actuelle des plateformes de production de données, qui

15. Le Règlement général sur la protection des données est entré en vigueur en mai 2018. Voir : <https://www.eugdpr.org/>.

permettent à une pléthore de producteurs (les propriétaires de plateformes, par exemple) et d'analystes (tels que les chercheurs, les pouvoirs publics et les médias) de créer de la statistique et de la connaissance sociétales (Ruppert *et al.*, 2013). En effet, les données massives générées par les navigateurs, les médias sociaux ou les périphériques tels que les téléphones mobiles – que divers acteurs peuvent consulter et analyser – permettent de mesurer de nombreux thèmes d'intérêt pour les INS, comme le niveau des prix, l'économie, la confiance des consommateurs ou le tourisme. D'aucuns y verront la marque d'une « démocratisation » du savoir et du déclin des connaissances et des expertises sociétales homologuées. Cependant, comme l'avance Ruppert *et al.* (2013), étendre de la sorte les champs de diffusion des données et d'analyse signifie que les connaissances sociétales ne forment pas des blocs d'un seul tenant faisant autorité, en tout cas sans commune mesure peut-être avec un passé récent. À l'inverse, ce qui fonde la légitimité d'un savoir et d'une expertise fait à présent l'objet de controverses et de confrontations politiques non négligeables, comme l'illustrent les débats actuels sur les « faits alternatifs ».

Les propositions qui appellent les INS à défendre la qualité et la légitimité de la statistique publique au moyen de pratiques de contrôle telles que faire la preuve de sa fiabilité en assurant la transparence, et donc l'évaluabilité, de leur pratique statistique, vérifier les statistiques contradictoires et « montrer du doigt les mauvais élèves » ont indubitablement un rôle à jouer. En revanche, elles nourrissent probablement le postulat selon lequel il ne s'agit là que de remporter une bataille de « faits ». Elles méconnaissent la nécessité de soumettre ce qui constitue un « fait public » à la contestation et à la délibération démocratiques, car celui-ci suppose inévitablement des jugements normatifs en termes de signification sociale et des choix quant aux réalités expérientielles pertinentes (Jasanoff & Simmit, 2017). Nous suggérons donc que les INS ont aussi vocation à promouvoir la statistique publique comme une réussite sociale et collective, dont la légitimité procède de modalités de co-production qui sollicitent les personnes concernées en qualité de citoyens habilités à participer activement. Une telle approche conçoit les données et la statistique publique comme des technologies sociales qui nécessitent des nouvelles formes de participation et de relations entre les experts, les décideurs et les citoyens en vue de la résolution des problèmes collectifs (Jasanoff, 2003), et

comme des questions relevant de la délibération démocratique, où les citoyens contribuent activement à construire et à façonner les savoirs des sociétés dont ils sont membres.

Nous convenons que le concept de données citoyennes soulève de nombreuses questions pratiques et politiques. En premier lieu, loin de nous l'idée de sous-entendre que les méthodes existantes et leur relation au citoyen seront frappées d'obsolescence. Cela dit, les méthodes comme les enquêtes et les questionnaires seront probablement amenées à se réinventer, face à l'adoption croissante des technologies numériques. Un concept de données citoyennes peut, le cas échéant, ouvrir la voie à de tels changements. Cela signifie, au-delà des sources de Big Data et selon notre modèle de données citoyennes, pouvoir redéfinir la manière dont les INS produisent des données à l'aide de diverses méthodes. Bien qu'en cours d'adoption, les enquêtes et les recensements en ligne ou numériques, par exemple, n'envisagent pas de scénarios de co-production. On pourrait opter pour différents modes de co-production s'appuyant sur les possibilités offertes par les technologies numériques et capables de produire des données reflétant plus fidèlement les expériences et les connaissances des citoyens.

Tout au long de notre réflexion, nous avons défini la co-production comme la participation des citoyens à tous les stades du processus de production. Les implications pratiques qui y sont liées représentent, bien entendu, une inconnue majeure et posent également la question de la représentativité et de l'inclusion au sein du processus. Il s'agit là d'une préoccupation inhérente à toutes les méthodes statistiques, en particulier si l'on considère l'hétérogénéité de la population. S'agissant des méthodes qui mobilisent les technologies numériques, telles que les recensements et les enquêtes en ligne, cette problématique est potentiellement exacerbée par ce qu'il est désormais convenu d'appeler la « fracture numérique ». Ce n'est là qu'un aperçu des questions politiques et pratiques éventuelles que font naître les données citoyennes, interrogations qui ont par ailleurs été abordées dans le cadre de l'atelier collaboratif avec des statisticiens lors de notre recherche. Bien que le présent article ne fasse pas état des résultats de l'atelier concerné, l'un d'entre eux imaginait d'autres « feuilles de route » pour permettre aux citoyens de participer à tous les stades du processus de production des données, allant de la co-conception de prototypes pour plateformes et applis de production de données

à la mise en place de formes coopératives de propriété des données. En d'autres termes, les données citoyennes appellent à repenser les processus de production statistique et certains de leurs principes fondamentaux.

Par exemple, la normalisation et la qualité des données sont deux aspects de la production statistique qu'il serait bon de revoir. Cependant, comme cela a déjà été dit, le principe même de l'expérimentalisme invite à s'ouvrir à ces questions et à ne pas y apporter de réponse préalable, y compris concernant ce qu'il convient ou conviendrait d'entendre par qualité. Paradoxalement, cela vaut également pour les expériences des INS qui s'appuient sur des données massives générées par des systèmes tiers, où des préoccupations relatives à la qualité, et d'autres comme la représentativité des données, sont apparues. L'une des solutions proposée par les statisticiens consiste à exploiter les statistiques qui réutilisent des données massives non pas dans une logique de remplacement, mais en qualité d'accessoire, de complément ou de supplétif aux sources de données existantes.

Si cela équivaut peut-être à reléguer les données concernées à un rang et un rôle autres, cette réponse est en tout cas l'occasion de repenser à ce qui fait les statistiques « publiques ». Elle suggère, par ailleurs, qu'il n'est pas de mode de production ou de série de normes établissant le caractère public des données. Pour nous, cela vaut également pour les méthodes existantes de production de données pour la statistique publique, qui impliquent une myriade de normes et où la qualité n'est pas nécessairement définie ou mesurable. Cela étant, le concept de données citoyennes que nous avons développé présente une différence cruciale, au-delà des questions normatives et qualitatives. Il propose que l'autorité et l'expertise nécessaires pour décider du caractère public de la statistique dépendent non pas d'une seule et même institution, mais de processus de co-production et de relations directes avec les citoyens. À cet égard, les données citoyennes envisagent les revendications de « faits alternatifs » non pas sous l'angle de l'exactitude et de la norme, mais de la relation au citoyen qui permet aux données, et par conséquent à la statistique, de devenir publiques. □

BIBLIOGRAPHIE

Barocas, S. & Nissenbaum, H. (2014). Big Data's End Run Around Anonymity and Consent. In Lane, J., Stodden, V. Bender, S. & Nissenbaum, H. (Eds.), *Privacy, Big Data, and the Public Good*, pp. 44–75. Cambridge, MA: Cambridge University Press.

Binder, T., Brandt, E., Ehn, P. & Halse, J. (2015). Democratic Design Experiments: Between Parliament and Laboratory. *CoDesign*, 11(3-4), 152–165. <https://doi.org/10.1080/15710882.2015.1081248>

Boltanski, L. & Chiapello, E. (2007). *The New Spirit of Capitalism*. London: Verso.

Cardoso, A. C., Tsiamis, K., Gervasini, E. et al. (2017). Citizen Science and Open Data: a model for Invasive Alien Species in Europe. Joint Research Centre (JRC) and the European Cooperation in Science and Technology (COST Association), *Workshop Report*. Brussels, BE. <https://doi.org/10.3897/rio.3.e14811>

Callon, M., Burchell, G., Lascoumes, P. & Barthe, Y. (2011). *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, MA: MIT Press.

Cavoukian, A., Taylor, S. & Abrams, M. E. (2010). *Privacy by Design: Essential for Organizational Accountability and Strong Business Practices*. *Identity in the Information Society*, 3(2), 405–413. <https://doi.org/10.1007/s12394-010-0053-z>

DataShift (n.d.). Global Goals for Local Impact: Using Citizen-Generated Data to Help Achieve Gender Equality. <http://civicus.org/thedatashift/wp-content/uploads/2017/01/LanetUmojaProcessandApproach.pdf> (accessed 22 February 2018)

Commission européenne (2013). Environmental Citizen Science. *Science for Environment Policy InDepth Report N° 9*. Bristol: University of the West of England, Science Communication Unit. http://ec.europa.eu/environment/integration/research/newsalert/pdf/IR9_en.pdf (accessed 22 February 2018)

Fleurbaey, M. (2009). Beyond GDP: The quest for a measure of social welfare. *Journal of Economic literature*, 47(4), 1029–1075. <https://doi.org/10.1257/jel.47.4.1029>

- Gabrys, J., Pritchard, H. & Barratt, B. (2016).** Just Good Enough Data: Figuring Data Citizenships Through Air Pollution Sensing and Data Stories. *Big Data & Society*, 3(2), 1–14.
<https://doi.org/10.1177/2053951716679677>
- Gabrys, J. & Pritchard, H. (2015).** Just Good Enough Data and Environmental Sensing: Moving Beyond Regulatory Benchmarks toward Citizen Action. In *Infrastructures and Platforms for Environmental Crowd Sensing and Big Data*. Barcelona: European Citizen Science Association.
<https://ecsa.citizen-science.net/sites/default/files/envip-2015-draft-binder.pdf> (accessed 22 February 2018)
- Goodchild, M. F. (2007).** Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221.
<https://doi.org/10.1007/s10708-007-9111-y>
- Graham, S. (2005).** Software-Sorted Geographies. *Progress in Human Geography*, 29(5), 562–580.
<https://doi.org/10.1191/0309132505ph568oa>
- Grommé, F., Ustek-Spilda, F., Ruppert, E. & Cakici, B. (2017).** Citizen Data and Official Statistics: Background Document to a Collaborative Workshop. ARITHMUS *Working Paper* N° 2.
http://arithmus.eu/wp-content/uploads/2015/02/ARITHMUS-collaborative-workshop-wp_final-version-060717-1.pdf
- Grommé, F. (2015).** *Governance by Pilot Projects: Experimenting with Surveillance in Dutch Crime Control* (Doctoral thesis). Amsterdam: University of Amsterdam.
<http://hdl.handle.net/11245/1.486712>
- Guillemin, M. & Gillam, L. (2004).** Ethics, Reflexivity, and “Ethically Important Moments” in Research. *Qualitative Inquiry*, 10(2), 261–280.
<https://doi.org/10.1177/1077800403262360>
- Henriquez, L. (2016).** *Amsterdam Smart Citizens Lab: Towards Community Driven Data Collection*. Amsterdam: De Waag Society and AMS Institute.
<https://waag.org/sites/waag/files/media/publicaties/amsterdam-smart-citizen-lab-publicatie.pdf> (accessed 2 April 2017)
- Hussmanns, R. (2004).** Measuring the Informal Economy: From Employment in the Informal Sector to Informal Employment. *Working Paper* N° 53.
http://www.ilo.org/wcmsp5/groups/public/---dgreports/---integration/documents/publication/wcms_079142.pdf (accessed 30 April 2018)
- Isin, E. & Ruppert, E. (2015).** *Being Digital Citizens*. London: Rowman & Littlefield International.
- Isin, E. & Saward, M. (2013).** *Enacting European Citizenship*. Cambridge: Cambridge University Press.
- Jasanoff, S. (2003).** Technologies of Humility: Citizen Participation in Governing Science. *Minerva*, 41(3), 223–244.
<https://doi.org/10.1023/A:1025557512320>
- Jasanoff, S. & Simmet, H. R. (2017).** No Funeral Bells: Public Reason in a “post-Truth” Age. *Social Studies of Science*, 47(5), 751–770.
<https://doi.org/10.1177/0306312717731936>
- Jungnickel, K. (2017).** Making Things to Make Sense of Things: DIY as Research Subject and Practice. In: Sayers, J. (Ed.), *The Routledge Companion to Media Studies and Digital Humanities*. Oxon: Routledge.
- Kitchin, R. (2014).** The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1–14.
<https://doi.org/10.1007/s10708-013-9516-8>
- Kullenberg, C. & Kasperowski, D. (2016).** What Is Citizen Science? – A Scientometric Meta-Analysis. *PLOS ONE*, 11(1), e0147152.
<https://doi.org/10.1371/journal.pone.0147152> (accessed 2 April 2017)
- Latour, B. (2004).** Why has critique run out of steam? From matters of fact to matters of concern. *Critical Inquiry*, 30(2), 225–248.
<https://doi.org/10.1086/421123>
- Latour, B. (2006).** Which Protocol for the New Collective Experiments? *Boletín CF+S*, (32/33).
<http://habitat.aq.upm.es/boletin/n32/ablat.en.html> (accessed 2 April 2017)
- Marres, N. (2012).** *Material Participation: Technology, the Environment and Everyday Publics*. Basingstoke: Palgrave Macmillan.
- Montjoye, Y.-A. (de), Shmueli, E., Wang, S. S. & Pentland, A. S. (2014).** openPDS: Protecting the Privacy of Metadata through SafeAnswers. *PLOS ONE*, 9(7).
<https://doi.org/10.1371/journal.pone.0098790>
- Muniesa, F. & Linhardt, D. (2011).** Trials of explicitness in the implementation of public management reform. *Critical Perspectives on Accounting*, 22(6), 550–566.
<https://doi.org/10.1016/j.cpa.2011.06.003>
- Nissenbaum, H. (2004).** Privacy as Contextual Integrity. *Washington Law Review*, 79(1), 119–158.
<https://nyuscholars.nyu.edu/en/publications/privacy-as-contextual-integrity> (accessed 2 April 2017)
- Nissenbaum, H. (2009).** *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford: Stanford University Press.

- Paul, K. T. (2018).** Collective organization of discourse expertise using information technology – CODE IT! *Information Technology*, 60(1), 21–27. <https://doi.org/10.1515/itit-2017-0022>
- Piovesan, F. (2017).** *Statistical Perspectives on Citizen-Generated Data*. [Online]. http://civicus.org/thedatashift/wp-content/uploads/2015/07/statistical-perspectives-on-cgd_web_single-page.pdf (accessed 22 February 2018)
- Puig de la Bellacasa, M. (2012).** “Nothing Comes Without Its World”: Thinking with Care. *The Sociological Review*, 60(2), 197–216. <https://doi.org/10.1111/j.1467-954X.2012.02070.x>
- Rabinow, P. & Bennett, G. (2012).** *Designing Human Practices: An Experiment with Synthetic Biology*. Chicago: University of Chicago Press.
- Ruppert, E. (2018).** *Sociotechnical Imaginaries of Different Data Futures: An Experiment in Citizen Data*. 3e Van Doornlezing. Rotterdam, NL: Erasmus School of Behavioural and Social Sciences. <https://www.eur.nl/sites/corporate/files/2018-06/3e%20van%20doornlezing%20evelyn%20ruppert.pdf> (accessed 21 Jan 2019)
- Ruppert, E., Law, J. & Savage, M. (2013).** Reassembling Social Science Methods: the Challenge of Digital Devices. *Theory, Culture & Society, Special Issue on “The Social Life of Methods”*, 30(4), 22–46. <https://doi.org/10.1177/0263276413484941>
- Ruppert, E., Harvey, P., Lury, C., Mackenzie, A., McNally, R., Baker, S. A., Kallianos, Y. & Lewis, C. (2015).** A Social Framework for Big Data. CRESC, The University of Manchester and The Open University, *Project Report*. <http://research.gold.ac.uk/13483/> (accessed 2 April 2017)
- Simon, H. (1947).** *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan.
- Socientize Consortium (2014).** Green paper on Citizen Science. Citizen Science for Europe: Towards a better society of empowered citizens and enhanced research. European Commission Digital Science Unit. <https://ec.europa.eu/digital-single-market/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research> (accessed 22 February 2018)
- Statistics Canada (2016).** *Open Building Data: an exploratory initiative*. <http://www.statcan.gc.ca/eng/crowdsourcing> (accessed 18 February 2018)
- Stengers, I. (2010).** *Cosmopolitics*. Vol. 1–2. Minneapolis: University of Minnesota Press.
- Stiglitz, J. E., Sen, A. & Fitoussi, J.-P. (2009).** *Report by the Commission on the Measurement of Economic Performance and Social Progress*. Paris: CMESP. <http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report> (accessed 30 April 2018)
- Struijs, P., Braaksma, B. & Daas, P. J. H. (2014).** Official statistics and Big Data. *Big Data & Society*, 1(1), 1–6. <https://doi.org/10.1177/2053951714538417>
- UNECE (2014).** The Role of Big Data in the Modernisation of Statistical Production Project. Report of the Big Data Privacy Task Team. <http://bit.ly/2eTHDOe> (accessed 2 April 2017)
- United Nations (2014).** “Fundamental Principles of Official Statistics”. *Resolution adopted by the General Assembly on 29 January 2014. A/RES/68/261*. <http://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf> (accessed 21 Jan 2019)
- United Nations (2016).** Make Sustainable Development Goals Relevant to Citizens. New York: Economic and Social Council. <https://www.un.org/press/en/2016/ecosoc6782.doc.htm> (accessed 2 April 2017)
- Waterton, C. & Tsouvalis, J. (2015).** On the Political Nature of Cyanobacteria: Intra-Active Collective Politics in Loweswater, the English Lake District. *Environment and Planning D: Society and Space*, 33(3), 477–493. <https://doi.org/10.1177/0263775815594305>
- Zwitter, A. (2014).** Big Data Ethics. *Big Data & Society*. 1(2) 1–6. <https://doi.org/10.1177/2053951714559253>
- Zyskind, G., Nathan, O. & Pentland, A. (2015).** Decentralizing Privacy: Using Blockchain to Protect Personal Data. *2015 IEEE Security and Privacy Workshops*, pp. 180–184. <https://doi.org/10.1109/SPW.2015.27>

