

La statistique publique à l'ère du numérique : entre déclin, mission impossible et nouveau départ ?¹

Hugues Picard

Membre du groupe Commerce électronique de l'Académie des Sciences Commerciales

Dans un article publié le 17 janvier 2017 dans le journal « The Guardian », le sociologue et économiste britannique William Davies estime que les statistiques, plus particulièrement les statistiques publiques, sont de plus en plus mises en doute par le public alors que les données issues des Big data seraient plus fiables. Cet article a été traduit en Français et publié dans la revue « Statistique et société » d'avril 2017. La revue a également traduit le communiqué de John Pullinger, directeur de l'Autorité Statistique du Royaume Uni, à propos de cet article et a questionné Dominique Bureau, président de l'Autorité de la statistique publique.

Que dit William Davies ? Il se place sur le plan politique et note que, pour la droite populiste d'abord, mais aussi pour de plus en plus de monde, outre les doutes que l'on peut avoir quant à la validité des chiffres, il paraît « insultant ou arrogant de ramener les questions sociales et économiques à des agrégats numériques et des moyennes ».

Sur la validité, ce n'est pas la justesse des indicateurs qui est en cause, mais le fait que ceux-ci sont souvent en contradiction avec le ressenti. Quant à l'utilité des données publiées, elle est mise en cause en arguant que nous sommes dans une société de plus en plus diversifiée où les agrégats et les moyennes cachent une diversité de plus en plus grande des situations individuelles, ce qui fait que, ne se reconnaissant pas dans les catégories pour lesquelles les chiffres sont calculés, on a tendance à les considérer comme inutiles sinon comme faux. D'autant plus que les politiciens sont renommés pour n'utiliser que les données qui confortent leurs idées et, souvent pour ignorer les limites d'utilisation de tel ou tel indicateur.

L'histoire du développement des statistiques permet de saisir le pourquoi des positions actuelles. Conçues, au départ, pour donner une vue générale de la population d'un état ou d'une région « autonome », elles se sont développées vers l'étude plus détaillée de ces populations et de leurs mouvements (naissances, décès, mariages) et également vers des mesures de production agricole, d'exportations, d'importations et de prix, mais de façon assez fruste et peu agrégée. Puis naissent les données avec critères croisés (tableau croisant lignes et colonnes) qui permettent de mieux décrire la complexité des sociétés. Bien entendu, on ne pouvait que présenter certains aspects des populations ou de leur production, en précisant bien à quoi les chiffres se référaient (par exemple, le PIB ne comptabilise que ce qui est produit par un travail rémunéré, à l'exclusion de la production domestique).

La production de données sur la société et l'économie s'est développée avec le temps, pour satisfaire les besoins de plus en plus grands des Etats. Mais, dès la fin du 18ème siècle, d'autres utilisateurs, notamment les universitaires analysant la société, se sont

¹ Le point d'interrogation ne figure pas dans le titre de l'article présenté ici. Il a semblé nécessaire de l'ajouter compte tenu du caractère plus optimiste des deux commentaires qui sont également résumés.

placés comme utilisateurs de statistiques. L' « invention » de la comptabilité nationale, le développement des politiques économiques, surtout après 1945 (la reconstruction) et de la modélisation de l'économie (développement de l'économétrie) et, également, l'étude et l'amélioration des enquêtes par sondage, ont marqué le point d'orgue des statistiques après le milieu du siècle dernier.

« A mesure que les indicateurs de santé, de prospérité, d'égalité, d'opinion et de qualité de la vie sont venus nous dire qui nous sommes collectivement et si les choses s'améliorent ou non, les politiciens se sont lourdement appuyés sur la statistique pour soutenir leur autorité » écrit William Davies.

La réduction de la description d'une population ou de l'économie d'un pays à un nombre limité d'agrégats forcément globaux apparaît de plus en plus difficile à accepter. Les agrégats cachent des situations individuelles contrastées voire très différentes (une production globale en expansion ne rend pas compte des secteurs ou des régions en récession). Par ailleurs, les agrégats ne sont pas forcément adaptés au phénomène que l'on souhaite mesurer : une baisse du nombre de chômeurs ne rend pas compte de la durée individuelle du travail. Certes, on essaie de présenter tous les aspects d'un phénomène, il n'en reste pas moins que, ce faisant, d'une part, on complexifie les indicateurs par l'augmentation de leur nombre et, d'autre part, en affinant les nomenclatures, on arrive à la limite des mesures correctes des phénomènes (diminution drastique du nombre d'individus pour un poste d'une nomenclature, source d'une plus grande variabilité des chiffres correspondants).

L'utilisation des big data est-elle la réponse aux critiques évoquées ci-dessus ? Certes, à partir de données « captées d'abord, avant que ne se posent les questions de recherche, on dispose de données beaucoup plus abondantes que dans les sondages ou les modèles statistiques. Mais, nous dit William Davies, il n'y donc pas de prédétermination de cadre d'analyse (la nation par exemple) ni de catégories (les chômeurs par exemple).

Les champs et les nomenclatures sont établis a posteriori, au vu des données collectées. Ce qui fait que l'on a affaire à un nouveau type d'experts qui ne cherchent pas à remplir un cadre fixé a priori mais sont guidés par la nature des données dont ils disposent. On passe donc d'une logique statistique à une logique de données.

Les nouveaux spécialistes, selon William Davies, travaillent plutôt au « coup par coup », à la commande, cherchant des régularités dans d'énormes banques de données. Mais surtout, ils travaillent pour des commanditaires qui n'ont guère la volonté de diffuser les résultats dans le domaine public. « Mais, même s'il y avait un Bureau pour l'Analyse des Données agissant pour le compte du public et de l'Etat, il n'est pas clair qu'il offrirait la sorte de perspective neutre que les « libéraux » défend aujourd'hui. Le nouveau mode de calcul convient bien pour détecter les tendances, percevoir l'humeur et découvrir des choses comme le gonflement d'une bulle. Il sert très bien les directeurs de campagne et de marketing. Il convient moins bien à la formulation de constats sur la société non ambigus, objectifs, faisant consensus, ce pourquoi les statisticiens et économistes sont payés ».

Comment s'en sortir ? William Davies reste très vague sur le sujet. Il faut tenir compte des réalités nouvelles, se servir de l'analyse pour développer et réorienter les statistiques

classiques. « La bataille à mener n'est pas, à long terme, entre une politique élitiste de faits et une politique populiste de sentiments. Elle est entre ceux qui défendent encore la connaissance publique et le débat public et ceux qui profitent de leur désintégration en cours ».

Le communiqué de John Pullinger est assez bref. Il réfute d'une part l'idée que le public croit de moins en moins aux statistiques publiques, si grande est la demande pour des statistiques plus rapides et plus détaillées aidant à comprendre et traiter les problèmes qui se posent à la société. Il appelle également à la réalisation d'études « dignes de confiance » sur les biais de sources d'information utilisés pour déguiser un avis (et préjugé) en « fait irréfutable ».

Dominique Bureau donne acte de la méfiance de certains vis-à-vis des statistiques publiques et des critiques relatives à leur manque de détail. Néanmoins, il ne faut pas exagérer ces critiques et considérer que les statisticiens publics, en l'occurrence l'Insee et les services statistiques ministériels, ne font rien pour améliorer la situation. Par exemple, on dispose maintenant de résultats infra-communautaires sur le revenu disponible et la pauvreté monétaire.

Certes, l'autorité des chiffres n'échappe pas au constat de défiance. Mais il ne faut pas confondre les interrogations sur la qualité des statistiques disponibles et sur leurs lacunes.

En matière de qualité, un Comité stratégique de la qualité a été créé à l'Insee ; il a établi un plan de gestion de la qualité au sein du processus de production statistique. D'ailleurs, les enquêtes de satisfaction montrent que si le public se reconnaît plus ou moins dans les chiffres publiés, il met plutôt en cause la manière dont ils sont utilisés.

Sur les lacunes, il faut noter que ce n'est pas d'aujourd'hui que les statisticiens publics ont essayé de prévoir les demandes à venir et d'établir, en conséquence, des plans pour développer de nouvelles statistiques et fournir des résultats plus détaillés pour les indicateurs existants. On peut citer, à titre d'exemples, un rapport du CNIS sur les statistiques de l'emploi, du chômage et de la précarité, qui date de 2008 et les travaux sur la recherche et la mise en place d'indicateurs du progrès social plus pertinents que le PIB.

Quant à l'utilisation des macro données, les statisticiens publics, soutenus par l'Etat sont plus optimistes que ne l'est William Davies. Ainsi, un projet de traitement des données de caisse des magasins de grande surface a été lancé, avec objectif de le rendre opérationnel en 2019. Un groupe du CNIS a été créé pour examiner la potentialité des développements statistiques à partir des grands fichiers ; il a déjà proposé de mettre à l'étude l'utilisation de certains d'entre eux pour les statistiques démographiques et de consommation. Le groupe étudie les potentialités d'autres fichiers. De plus, la récente Loi sur la République numérique autorise maintenant la statistique publique à demander la transmission (sous forme sécurisée) d'informations issues de bases de données détenues par des personnes de droit privé.