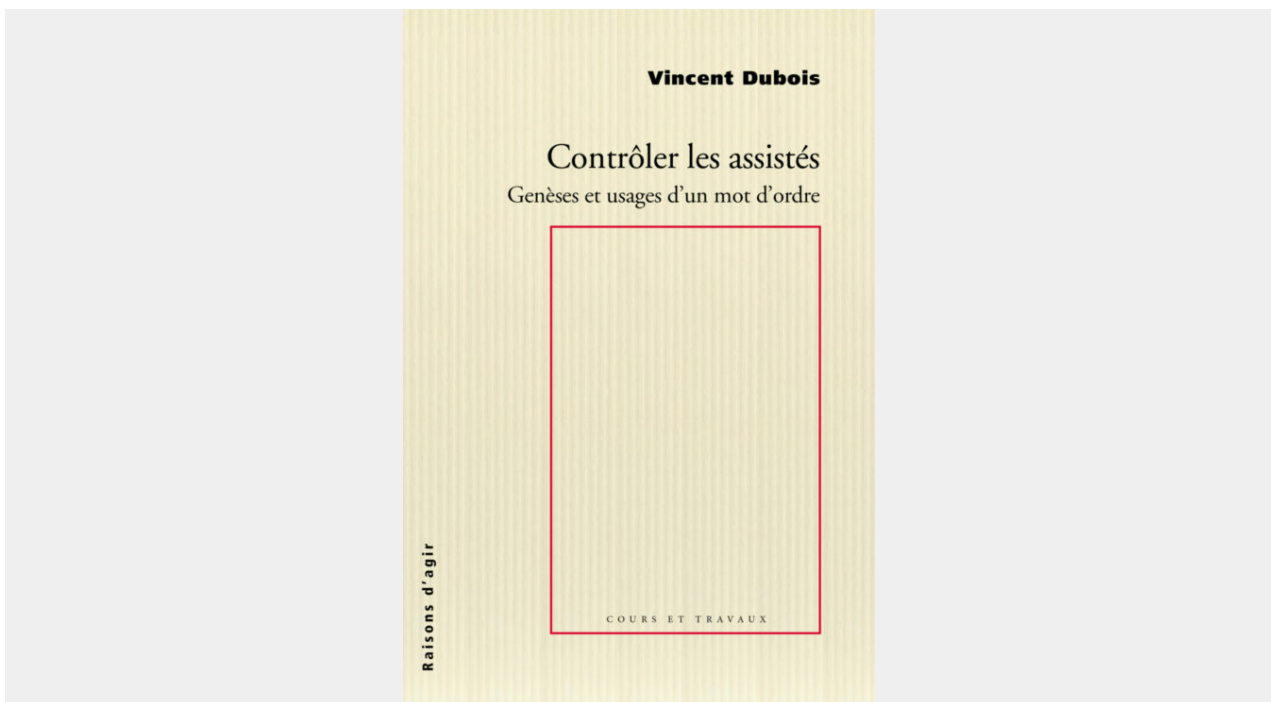


Renforcer le contrôle par les technologies de l'information et les statistiques



On sait que le montant cumulé des aides indues par des fraudeurs reste très en-deçà du non-recours aux prestations sociales par des allocataires qui y auraient droit. Et pourtant les contrôles et la surveillance n'ont eu de cesse de se renforcer. Si ces pratiques ne sont pas nouvelles, elles se singularisent aujourd'hui par l'ampleur que leur confèrent les technologies numériques et les possibilités nouvelles de détection d'anomalies dans les dossiers des allocataires. Dans cet extrait, Vincent Dubois examine ainsi les usages du data mining au sein des administrations publiques françaises désormais rompues au modèle managérial de la gestion des risques : les analyses statistiques de l'État néolibéral délaissent le cadre collectif de la population pour se focaliser sur l'individu et anticiper ses potentielles déviances.

Cet extrait est tiré du chapitre 6 de l'ouvrage de Vincent Dubois, [Contrôler les assistés: genèses et usages d'un mot d'ordre](#), Raisons d'agir, 2021. Silo remercie l'éditeur et l'auteur de nous avoir autorisés à le reproduire ici.

Virginia Eubanks voit dans les technologies numériques de traitement des pauvres aux États-Unis développées à partir du dernier quart du XX^e siècle le prolongement technique d'une tradition séculaire de limitation de l'accès aux aides publiques, de surveillance, et de sanction de ceux qui en bénéficient. Un système sophistiqué d'informatisation des données personnelles, de traitements automatisés et de statistiques prédictives constituerait ainsi la version numérique des institutions disciplinaires et punitives « de brique et de mortier » (*poorhouses*) qui, selon elle, posent au début du XIX^e siècle les orientations principales du *welfare* étasunien : une *digital poorhouse*[1]. Cette analyse présente l'avantage d'éviter la fréquente fascination (positive ou négative) pour la « nouveauté » qu'apporteraient les technologies de l'information. À l'inverse, elle ne doit pas masquer les changements associés à leur usage désormais massif.

Les technologies au service d'une nouvelle politique de contrôle

En France au moins, l'essor sans précédent des technologies d'information et de traitement des données personnelles coïncide comme on l'a vu avec l'avènement d'une nouvelle politique de contrôle. Plus que la continuation d'une tradition de surveillance des pauvres par de nouveaux moyens techniques, cet essor va de pair avec la structuration et la rationalisation de cette politique. Au-delà des changements bien réels de support (du papier au numérique) et d'échelle (l'augmentation du volume de données traitées), l'usage des technologies numériques ouvre des possibilités inédites d'accès, de traitement et de recoupement d'informations personnelles. Il n'améliore pas seulement les dispositifs existants ; il met en place un nouveau système de surveillance et de potentielles sanctions. S'y ajoute un changement dans les usages des statistiques comme « outil de gouvernement » des politiques sociales[2]. Ce n'est plus la population comme ensemble qui est l'objet de connaissance, comme dans la statistique descriptive agrégeant des collectifs, mais des individus et leurs comportements, dans une statistique cette fois prédictive, en l'occurrence le *data mining* utilisé pour la détection des risques d'anomalies dans les dossiers des allocataires. Ce changement dans l'objet et la méthode du gouvernement statistique va de pair avec un changement dans sa finalité, de l'identification de risques sociaux collectifs à socialiser et à assurer à celle de risques essentiellement financiers que font courir à l'institution des déviances individuelles à repérer et corriger.

Outiller pour rationaliser : technologies et systèmes d'information

La conjonction dans les années 1990 d'une informatisation devenue centrale dans le fonctionnement des administrations publiques, d'un assouplissement des règles régissant l'usage des données, des progrès techniques dans la fiabilité de leur traitement, et de la montée en puissance de préoccupations sécuritaires a conduit au développement sans précédent de différentes formes de surveillance électronique par les administrations. Techniques d'identification biométriques, usage par l'ensemble des administrations d'un même numéro individuel à des fins d'échanges de données, constitution de fichiers à large échelle, facilitation de l'accès à ces fichiers d'une administration à l'autre ou multiplication des croisements automatisés des données : ces diverses innovations ont marqué l'ensemble des administrations et secteurs d'action publique[3]. C'est évidemment le cas dans le domaine du renseignement et de la sécurité, où l'usage des bases de données, s'il n'a pas révolutionné les pratiques et varie selon la socialisation professionnelle des agents et leur position dans l'espace transnational du contre-terrorisme, n'en demeure pas moins massif[4]. La politique

de l'Union européenne en matière d'immigration a quant à elle donné lieu au développement d'une « infrastructure digitale » constituée de plusieurs bases de données dont l'usage permet la surveillance des immigrés en situation dite irrégulière[5]. Le « gouvernement des chômeurs » au plan national fait un usage massif des bases de données à des fins de profilage, de ciblage et de surveillance, non sans effets sur le traitement dont ils font l'objet[6].

Certes, l'usage des données personnelles pour le gouvernement des populations n'est pas nouveau. La constitution de fichiers est consubstantielle à la bureaucratie, et leur informatisation est engagée dès les années 1960. Des systèmes d'interconnexion des fichiers proches de ceux mis en place aux États-Unis ont été généralisés aux Pays-Bas au début des années 1990[7]. En France, la CNAF organise dès les années 1980 des « liaisons automatisées et transmission de données » avec les partenaires déjà réalisées avec le fisc pour comparer les ressources déclarées à la CAF et les déclarations de revenus, envisagées avec les ASSEDIC pour les changements de situation professionnelle et les indemnités de chômage, les offices HLM pour le montant des loyers, les hôpitaux pour l'hospitalisation des personnes handicapées[8].

Le traitement informatisé des données personnelles par l'administration s'inscrit dans cette histoire, mais son usage ne se réduit pas à un simple progrès technique. L'échelle inédite des bases de données qui concernent désormais des pans entiers de la population, l'abondance et la précision des informations permises notamment par la multiplication et le recoupement des sources qui constitue un « double informatique » des individus[9], les décisions les concernant qui sont prises sans même qu'ils en aient conscience aboutissent en effet « à une véritable révolution du contrôle » qui en change la nature même et pas seulement les modalités techniques[10] [...]

La statistique prédictive dite de data mining au service de la « maîtrise des risques »

[...]

Quand bien même elle reste à un niveau général et n'aborde pas en eux-mêmes les « outils de rationalisation », la manière dont Foucault a mis en évidence l'usage des statistiques dans le gouvernement de la société libérale offre un premier point d'appui pour prendre la mesure des changements auxquels le *data mining* est associé[11]. Le savoir statistique dont l'usage est central dans le gouvernement libéral des risques vise à repérer des régularités, à établir des normes, et se fonde sur la notion de population considérée dans son ensemble ; il ne connaît ni d'individus, ni de différences individuelles. La connaissance sur laquelle se fonde le gouvernement néolibéral est tout autre. Ce n'est plus la population qui constitue l'unité d'analyse statistique mais cette fois l'individu, dont on connaît précisément les caractéristiques et les actes. Ce n'est plus la norme et les écarts à la norme qu'il s'agit de calculer, mais les comportements individuels qu'il s'agit de prédire[12], marquant « un nouveau rapport aux données et à la quantification », qui vise une connaissance des populations à gouverner non plus en fonction de catégories collectives prédéfinies mais en fonction de pratiques individuelles observées et anticipées sur cette base[13]. L'usage du *data mining* dans le contrôle contemporain des assistés sociaux est ainsi exemplaire de la « politique des grands nombres » de l'État néolibéral qui, entre autres, inclut « des profilages “d'individus à risque” sélectionnés sur des fréquences observées (*profiling*) »[14].

Expérimentées à partir du milieu des années 2000, les statistiques prédictives dites de *data mining* sont devenues au début de la décennie suivante la technique privilégiée pour le déclenchement des contrôles des allocataires. C'est que son efficacité s'avère bien supérieure à celle des modalités antérieures de sélection des cas à contrôler. Son caractère prédictif identifiant de façon systématisée les anomalies potentielles en fait une technique en parfaite affinité avec le modèle de pensée et d'organisation de la « maîtrise des risques ». Plus largement, elle compte au nombre des dispositifs permettant de cibler et suivre de façon individualisée des

sous-populations, de calculer leur niveau de risque et de réguler leur comportement, fidèle en cela à la « philosophie sociale » du néolibéralisme basée sur une explication individuelle des phénomènes sociaux^[15]. Les contrôles « sur pièces », par vérification approfondie des éléments du dossier et demande de justificatifs complémentaires ou « sur place », par enquête à domicile, peuvent être demandés suite au signalement d'un tiers (dénonciation d'un particulier ou indication par un organisme), à la demande du technicien chargé du dossier, de façon aléatoire ou ciblée. Avant la généralisation de l'usage du *data mining*, les cibles formaient l'armature des plans annuels de contrôle lancés à partir de 1996 et mis en œuvre par les caisses locales, tant pour des croisements de fichiers que pour le déclenchement de contrôles manuels ou sur place. Ces cibles étaient définies nationalement, au sein du groupe de travail sur la politique de contrôle et en collaboration avec les informaticiens qui en ont assuré la réalisation technique.

[...]

Les contrôles « portent sur l'ensemble de la situation de l'allocataire et sont ciblés sur des prestations ou des situations à risques »^[16]. Quarante-cinq cibles de contrôle renvoyant à des cas identifiés comme tels ont été établies, dont douze pour le seul RMI, les autres concernant surtout la monoparentalité, « la cohérence entre les charges de logement et les ressources », les changements de droits au 1er juillet, la « cohérence entre l'activité et les ressources déclarées ». Le ciblage des prestations et situations a suscité un débat sur la possible stigmatisation de groupes et individus « à risques », et la focalisation sur les allocataires les plus précaires. « Certains directeurs considèrent que cette politique de contrôle vise aujourd'hui essentiellement les populations en difficulté. Une réflexion sur les cibles du contrôle de l'institution leur paraît nécessaire », lit-on ainsi dans le compte-rendu d'une réunion^[17]. Les modalités de définition des cibles rendent ces questions d'autant plus délicates que les cibles apparaissent comme le résultat de choix délibérés posés *a priori* et devant être assumés comme tels. En outre, si leurs résultats sont meilleurs que ceux des contrôles aléatoires, la quête permanente d'une efficacité et d'une « rentabilité » plus grandes conduit les promoteurs de la politique de contrôle à rechercher des outils toujours plus performants.

Le *data mining* est venu offrir la solution technique à ces difficultés^[18]. Le terme générique *data mining* englobe des méthodes d'exploration ou fouille de données à large échelle, dont l'usage a notamment été favorisé par l'essor d'internet, via le traitement par des algorithmes à des fins notamment commerciales des multiples « traces » laissées par les internautes. Plus précisément, pour ce qui nous concerne ici, le *data mining* désigne des méthodes de gestion « actuarielle » qui combinent quatre caractéristiques : l'usage de méthodes statistiques, la mobilisation de vastes bases de données personnelles, la détermination de différents niveaux de risques corrélés à des caractéristiques construites comme facteurs de risque, des prédictions établies sur la base de ces corrélations qui déterminent le traitement différencié des individus^[19]. Cette évaluation statistique des risques procède par induction et systématisation progressive, de l'observation exploratoire des données à la construction de modèles prédictifs à partir des corrélations les plus fortes repérées entre variables, puis à l'usage de ces modèles pour traiter des données à grande échelle, et à leur révision en fonction des résultats obtenus ou de nouvelles explorations.

[...]

En France, dans les administrations publiques, le *data mining* est en particulier promu en tant qu'instrument permettant d'améliorer le ciblage pour la lutte contre la fraude, par les douanes et Pôle emploi à partir de 2013, par l'administration fiscale progressivement à partir de 2014, de façon plus générale à partir de 2018^[20], cette orientation étant renforcée par la loi contre la fraude fiscale promulguée cette même année, et promue dans le rapport de la Cour des comptes sur la fraude aux prélèvements obligatoires l'année suivante^[21].

Le contrôle des allocataires de la CAF par le

Data mining

La Caisse nationale des allocations familiales est dans ce domaine précurseur, puisque le *data mining* a été utilisé de manière expérimentale dès 2004 et généralisé en 2011 pour identifier les anomalies dans les dossiers. L'usage précoce de cette technique a sans doute été facilité par la disponibilité d'une très importante masse de données personnelles, en raison du nombre et de la diversité des prestations versées par les CAF : chacune d'elle repose sur toute une batterie de critères, couvrant non seulement les ressources et données d'état-civil, mais aussi l'historique des situations familiales et professionnelles, l'âge et la scolarisation des enfants, le logement, sans parler bien sûr des allocations perçues, et ce pour près de treize millions d'allocataires soit, avec leurs conjoints, enfants et personnes à charge, près de trente millions de personnes concernées. Mais cette condition technique n'est pas suffisante. Si le *data mining* occupe une place devenue centrale dans le contrôle des allocataires, c'est qu'il offre un outil répondant à la demande d'un renforcement de ce contrôle et de son efficacité, et que son mode de fonctionnement concorde très directement avec le modèle managérial de la maîtrise des risques.

La réussite d'une technique statistique implique « *qu'elle corresponde à de nouvelles façons de penser et d'organiser les rapports sociaux* » et « *qu'un réseau d'innovateurs trouve suffisamment d'alliés pour la diffuser* »^[22]. En l'occurrence, ces calculs mathématiques ont « *pour effet d'individualiser le risque [en le rattachant] à des caractéristiques personnelles* », occultant que « *la plupart des cas de non-conformité résultent de la complexité des politiques et des procédures gouvernementales* »^[23]. La promotion de cette manière de voir et du *data mining* qui l'incarne s'articule ainsi avec les représentations culpabilisantes des assistés sociaux. L'usage de cette technique constitue, avec les échanges automatisés de données, l'une des principales réponses techniques à l'injonction politique au renforcement du contrôle. Quant à ses promoteurs, ils se recrutent au gouvernement, à la Cour des comptes ou la Délégation nationale à la lutte contre la fraude, qui formulent parfois des préconisations très directes quant à l'usage de nouvelles techniques de contrôle. [...]

Cet instrument ne détermine donc pas en lui-même ses usages possibles, qui peuvent être variés. Les « cibles » des modèles de *data mining* peuvent être des dossiers « à risque » financier, mais également des publics éligibles aux prestations. Il est ainsi possible de mobiliser ces techniques pour faciliter l'accès aux droits sociaux et lutter contre le non-recours, en ciblant les populations à partir des facteurs prédictifs de la non-demande^[24], comme en Grande-Bretagne ou au Pays-Bas^[25], ce qui est notamment évoqué par le statisticien qui en a la charge à la CNAF^[26], et proposé dans un rapport parlementaire pour l'heure peu suivi d'effet sur ce point^[27]. Quand bien même ils ont été conçus initialement pour détecter les indus et la fraude, l'application des modèles de *data mining* conduit du reste bien à identifier des prestations non perçues. Mais, tout comme la politique de « maîtrise des risques » dont il constitue un instrument privilégié, le *data mining* se confond essentiellement avec le contrôle des allocataires^[28].

Notes de bas de page :

[1] Virginia Eubanks, *Automating inequality: how high-tech tools profile, police, and punish the poor*, First Edition., New York, NY, St. Martin's Press, 2017 ; Cette expression est d'abord utilisée dans J. Gilliom, *Overseers of the poor*, *op. cit.*

[2] Alain Desrosières, *L'argument statistique 2. Gouverner par les nombres*, Paris, Presses de l'école des mines, 2008.

[3] Sans parler des usages par des entreprises privées *via* notamment le traitement algorithmique à des fins commerciales des données collectées sur internet.

[4] Didier Bigo et Laurent Bonelli, « *"Nous ne sommes pas un Big Brother !" . Autorité et stratégies de légitimation des services de renseignement dans la captation et l'usage des données numériques* », *Cultures & Conflits*, 2019, vol. 2, n° 114-115, pp. 199-226.

- [5] Dennis Broeders, « The New Digital Borders of Europe: EU Databases and the Surveillance of Irregular Migrants », *International Sociology*, 2007, vol. 22, n° 1, pp. 71-92.
- [6] Paul Henman, « Targeted! Population Segmentation, Electronic Surveillance and Governing the Unemployed in Australia », *International Sociology*, 2004, vol.19, n° 2, pp. 173-191 ; Karolina Sztandar-Sztanderska et Marianna Zielenska, « [Changing social citizenship through information technology](#) », *Social Work & Society*, 16-2, 2018, p. online.
- [7] Loïc Wacquant, *Les prisons de la misère*, Raisons d'agir, 2015, pp.121-122.
- [8] Philippe Steck, *La politique du contrôle dans la branche prestations familiales*, CNAF, 15 octobre 1985.
- [9] Armand Mattelart et André Vitalis, *Le profilage des populations. Du livret ouvrier au cybercontrôle*, Paris, La Découverte, 2014, p.112.
- [10] *Ibid.*
- [11] Michel Foucault, *Sécurité, territoire, population : cours au Collège de France, 1977-1978*, Paris, Seuil : Gallimard, 2004.
- [12] Fabrice Bardet, *La contre-révolution comptable. Ces chiffres qui nous gouvernent*, Paris, Les Belles Lettres, 2014 ; Laurence Barry, « The rationality of the digital governmentality », *Journal for Cultural Research*, 2019, vol. 23, n° 4, pp. 365-380.
- [13] À propos des politiques développées en Grande-Bretagne par le gouvernement du New Labour, Evelyn Ruppert, « The Governmental Topologies of Database Devices », *Theory, culture and society*, 2012, vol. 29, n° 4-5, pp. 116-136.
- [14] Alain Desrosières, « [La statistique, outil de gouvernement et outil de preuve](#) » dans Alain Desrosières, *Pour une sociologie historique de la quantification. L'argument statistique 1*, Paris, Presses des Mines, pp. 7-20.
- [15] Paul Henman et Greg Marston, « The Social Division of Welfare Surveillance », *Journal of social policy*, 2018, vol. 37, n° 2 p.194.
- [16] Convention d'objectifs et de gestion CNAF-État de 1997.
- [17] Comité de programme et de concertation des politiques de la CNAF du 19 juin 1998, Synthèse des travaux, CNAF.
- [18] Ce passage reprend des éléments présentés dans Vincent Dubois, Morgane Paris et Pierre-Édouard Weill, « [Des chiffres et des droits. Le data mining ou la statistique au service du contrôle des allocataires](#) », *Revue des politiques sociales et familiales*, 2018, n° 126, pp. 49-60 ; Vincent Dubois, Morgane Paris et Pierre-Édouard Weill, « [Targeting by Numbers. The Uses of Statistics for Monitoring French Welfare Benefit Recipients](#) » dans Lorenzo Barrault-Stella et Pierre-Edouard Weill (dir.), *Creating Target Publics for Welfare Policies: A Comparative and Multi-level Approach*, Cham, Springer International Publishing, 2018, pp. 93-109.
- [19] Bernard E. Harcourt, *Against prediction: profiling, policing, and punishing in an actuarial age*, Chicago, University of Chicago Press, 2007, p.16.
- [20] Rapports de la Délégation Nationale à la Lutte contre la Fraude, 2014 et 2018.
- [21] Cour des Comptes, *La fraude aux prélèvements obligatoires*, Paris, Cour des comptes, 2019.
- [22] Alain Desrosières, *Prouver et gouverner. Une analyse politique des statistiques publiques*, Paris, La Découverte, 2014, p.81.
- [23] Paul Henman, « Targeted! », *International sociology*, 2004, vol. 19, n° 2, p.179.
- [24] Philippe Warin, *Le non-recours aux politiques sociales*, Grenoble, PUG, 2017.
- [25] Arre Zuurmond, « Bureaucratic Bias and Access to Public Services: The Fight against Non-Take-Up » dans Jorrit de Jong et Gowher Rizvi (dir.), *The State of Access. Success and Failure of Democracies to Create Equal Opportunities*, Washington, DC, Brookings Institution Press, 2008, pp. 148-166.
- [26] Pierre Collinet, « [Le data mining dans les Caf : une réalité, des perspectives](#) », *Informations sociales*, 2013, vol. 4, n° 178, p. 129-132, ici p.132.

[27] Gisèle Biémouret et Jean-Louis Costes, *L'évaluation des politiques publiques en faveur de l'accès aux droits sociaux*, Paris, Comité d'évaluation et de contrôle des politiques publiques, Assemblée nationale, 2016.

[28] Un usage comparable de techniques similaires se développe en Australie dès le milieu des années 1980. Paul Henman, « Targeted! », art. cité, pp.177-180.

Pour citer cet article :

Vincent Dubois, « Renforcer le contrôle par les technologies de l'information et les statistiques », *Silomag* 15, juillet 2022. URL :

<https://silogora.org/renforcer-le-controle-par-les-technologies-de-linformation-et-les-statistiques/>