



Le « Big Data » et l'évaluation des politiques publiques

Vers des méthodes d'évaluation plus inclusives ?

Noé Germakopoulos

Master Mégaonnées et Analyse Sociale

USID0E – Humanités Numériques

Conservatoire National des Arts et Métiers de Paris

Mars 2019



Introduction

Apparu en 1997 pour désigner les « grands ensembles de données »¹, le terme « Big Data » a commencé à susciter l'intérêt du public début 2012² allant jusqu'à être intégré à la langue Française sous le terme de « Mégadonnées » deux ans plus tard³. Ces « *données structurées ou non dont le très grand volume requiert des outils d'analyse adaptés* »⁴ sont dès lors présentées comme révolution dans le traitement, l'accès et l'analyse de l'information annonçant un profond bouleversement des dogmes organisationnels pour l'ensemble des structures économiques⁵.

Même si certains changements structurels furent observés au sein des grandes entreprises privées⁶, la vague révolutionnaire des données massives n'a cependant pas atteint tous les secteurs d'activité une décennie après son essor. Le secteur public, connu pour une « lourdeur » administrative depuis au moins deux siècles⁷, semble avoir acquis un retard sur l'intégration des données massives dans ses processus analytiques, décisionnels et managériaux en dépit d'un intérêt grandissant des décideurs publics pour la matière. Car les données massives apparaissent comme prometteuses, notamment dans l'analyse des actions de l'État à travers ses politiques publiques, qu'elles s'ancrent dans une visée de service⁸ ou de puissance⁹ publique. Définies comme une « *réaction [de l'État] à un problème actuel, perçu ou anticipé* »¹⁰, les politiques publiques ont un besoin crucial « *d'identification et de conceptualisation un problème qui pourrait être à traiter* »¹¹. Appliquer des méthodes issues des données massives à l'évaluation de politiques publiques pourrait dès lors ouvrir la voie à de nouvelles appréciations de l'action publique.

Dès lors, comment les données massives peuvent-elles transformer l'évaluation des politiques publiques et dans quelle mesure remettent-elles en cause les dogmes existants ?

Remettant en question des modèles traditionnels d'évaluation des politiques publiques basés sur des méthodes de catégorisation (I), les données massives collectées dans le cadre de missions de service public s'avèrent être un outil puissant et fiable d'évaluation et de

¹ Gil Press, « A Very Short History Of Big Data », *Forbes*, 9 mai 2013

² Google trends, « Big Data », consulté le 13/03/2021

³ JORF n°0193 du 22 août 2014

⁴ JORF n°0193 du 22 août 2014

⁵ K. Carillo, *Le big data, une révolution qui transforme l'entreprise en profondeur*, Harvard Business Review France, 2017

⁶ Yves Eychenne, Jean-Charles Cointot, *La Révolution Big data: Les données au coeur de la transformation de l'entreprise*, 2014

⁷ M. Pinet, *Histoire de la Fonction Publique en France*, 1993

⁸ Louis Rolland, *Précis de droit administratif*, 1926

⁹ Maurice Hauriou, *Précis de droit administratif et de droit public général : à l'usage des étudiants en licence et en doctorat ès-sciences politiques*, 1901

¹⁰ T. Birkland, *Introduction to the policy process : Theory, concepts and methods for policy making*, 2001

¹¹ L. A. Schintler, R. Kulkarni, *Big Data for Policy Analysis : The Good, The Bad and the Ugly*, Review of Policy Research, 2014

prédiction malgré certains risques (II). La Fonction Publique gagnerait dès lors à généraliser le déploiement des données massives au-delà des administrations spécialistes des études pour tendre vers une autonomie des services en matière d'évaluation des politiques publiques.

I) Les Données Massives, une remise en question des dogmes statistiques dans l'évaluation des politiques publiques

Afin de mesurer, comprendre et prévoir pour gouverner, les méthodes d'évaluation des politiques publiques ont jusqu'à présent été basées des dogmes de catégorisation statistique des populations (A). Cet outil ayant atteint des limites, les données massives pourraient dès lors s'imposer comme une solution à la pertinence décroissante des méthodes traditionnelles d'évaluation par la puissance publique (B).

A) Catégoriser pour gouverner

Si Chaptal avait déjà commencé à s'appuyer sur un des méthodes statistiques lorsqu'il était Ministre de l'Intérieur sous le Premier Empire¹², ce n'est qu'en 1833 que les statistiques prirent de l'importance dans la gouvernance étatique avec la création des Statistiques Générales de France (SGF)¹³. Sous tutelle du Ministère du Commerce, ce service absorbe peu à peu les services statistiques des autres ministères, s'autonomise et diversifie son activité¹⁴. Le développement de ses missions ainsi que son indépendance fut achevé lors de la création en 1946 de l'Institut National des Statistiques et des Études Économiques (INSEE)¹⁵, organisme le remplaçant officiellement. De services spécialisés pour chaque ministère, les services d'études statistiques ont laissé place à une administration centrale, indépendante et polymathique dans ses missions.

Le tournant des années 80 fit basculer les statistique étatiques dans « l'analyse des politiques publiques »¹⁶, pratique jusqu'alors circonscrite au Monde anglo-Saxon¹⁷. Présenté comme une « modernisation » par le gouvernement Rocard¹⁸, cette « *construction d'un jugement sur la valeur de l'intervention publique* »¹⁹ pousse à une mesure de l'action publique et une gouvernance placée sous le dogme des chiffres et de la mesure via de nouveaux

¹² Y. ISSELIN, Répertoire numérique de la sous-série F20, Archives nationales (Paris), 2010

¹³ INSEE.fr, *Données historiques de la Statistique générale de France*, 2010

¹⁴ A. Desrosières, *La politique des grands nombres. Histoire de la raison statistique*, 2010

¹⁵ Articles 32 et 33 de la loi de finances du 27 avril 1946

¹⁶ E. Monnier, *Évaluation de l'action des pouvoirs publics*, 1992

¹⁷ A. Desrosières, op. cit.

¹⁸ E. Monnier, op.cit.

¹⁹ Commission Européenne, " Focus on Results: Strengthening Evaluation of Community Activities", ref SEC(2000)1051

indicateurs fortement influencés par le courant néo-libéral des années 80²⁰. On ne résonne des lors plus que par le prisme de catégories²¹ où chaque individu devrait y être circonscrit.

Si cette catégorisation à outrance fut efficace dans l'observation néolibérale de l'action publique des années 1990, elle a atteint très vite une limite au début des années 2000. Politiques hors sol²², déconnectées de la réalité et exclusion d'une part non-négligeable de la population furent le résultat de l'évaluation par la seule fenêtre d'une imparfaite catégorisation, au risque de rendre l'Etat aveugle sur ses propres actions.

B) Collecter pour mieux gouverner

Vers le milieu des années 2000, l'essor de l'influence de l'économie vient raviver des méthodes d'évaluation des politiques publiques devenues poussiéreuses²³. Un outil particulièrement intéressant va faire son apparition et connaître un succès grandissant dans les méthodes d'évaluation : la régression, ou « détermination de la grandeur approximative d'un phénomène correspondant à la grandeur certaine d'un autre phénomène ²⁴ ». Ainsi, en étudiant la corrélation de deux variables liées, il est désormais possible d'observer des phénomènes en s'affranchissant de toute catégorisation. Le défi n'est plus de trouver les bonnes catégories mais de trouver les variables et les modèle de prédiction et d'observation les plus adéquats²⁵.

La place croissante des méthodes économétriques dans l'évaluation des politiques publiques constitue un terrain propice au développement des données massives de par le besoin massif de variables induit par les méthodes de régression²⁶. De par la massivité et la diversité d'informations que le « Big Data » met à disposition de manière quasi immédiate, un nombre infini de régressions ont de quoi être alimentées. Là où les catégories admettaient pour limite l'exclusion de tout individu en dehors de leurs périmètres, chaque citoyen peut être individuellement et indépendamment être représenté et ce à travers un important éventail d'informations. Les oubliés de la gouvernance catégorielle²⁷ reviendraient ainsi dans le champ de la considération étatique, ce qui permettrait de mettre en lumière des faiblesses dans la conception et la mise en œuvre des politiques publiques et les faire évoluer pour les rendre plus inclusives.

²⁰ Telles les « Professions et Catégories socioprofessionnelles », Emmanuel Pierru et Alexis Spire, « Le crépuscule des catégories socioprofessionnelles », Revue française de science politique, vol. 58, n° 3, 2008.

²¹ D. Cardon, *A quoi rêvent les algorithmes*, 2015

²² D. Cardon, op.cit.

²³ D. Cardon, op.cit.

²⁴ Dictionnaire Larousse

²⁵ G. Casela, R. Berger, *Statistical Inference*, 2008

²⁶ D. Cardon, op. cit.

²⁷ J. Hillje, *Les oubliés – Entretiens où prospère le vote extrême*, Das Progressive Zentrum, 2018

Les données massives, associées aux outils d'évaluation économétriques, permettraient dès lors à la puissance publique de mieux observer son action, ses effets ainsi que de mieux prévoir ses actions futures.

II) Les Données Massives, outil principalement fiable pour le Service Public

Malgré un risque de submersion et d'exclusion bien réel (A), les données massives produites par service public sont suffisamment fiables pour être utilisées judicieusement, ce qui rend possible un bouleversement des pratiques au sein des administrations (B).

A) Le risque de la submersion et de l'exclusion

Les données massives furent définies en 2001 à travers trois grands V²⁸ : volume, vitesse et variété. Souvent le premier de ces V, pourtant le plus important, a tendance à être oublié des architectes de l'information. Qui dit « données massives » ne dit pas que « données ». La particularité d'une dimension volumineuse dans l'information des Big Data induit par conséquent à l'implémentation d'un particularisme indispensable à sa gestion. De plus, Les infinies possibilités de corrélation offertes par les méthodes de régression, ou chaque variable est observable et sujette à prédiction en fonction d'une autre, ajoutent une couche de densité supplémentaire à un volume d'information déjà surabondant. Avec d'un côté des données en nombre considérable et des possibilités de couplage exponentiels de l'autre, on frôle (souvent) l'indigestion informationnelle.

Un tel flux de données peut dès lors mener à une construction faussée d'indicateurs qui serait grandement préjudiciable à l'évaluation des politiques publiques. Outre le risque d'un choix de données à observer peu judicieux, il est primordial de noter qu'une différence majeure existe entre données et données justes. Ayant tendance à être biaisées de nature²⁹, les données peuvent subir des altérations tout au long de leur cycle de vie. De l'erreur de saisie au mauvais nom de colonne dans un programme de traitement informatique, les possibilités de détérioration de l'information sont multiples. Dès lors, faire une confiance aveugle à un jeu de donnée et prendre les informations en résultant pour vérité absolue est extrêmement risqué. Des étapes de vérification et de nettoyages sont nécessaires bien que parfois insuffisantes si les biais de la donnée collectée sont trop importants³⁰.

Enfin, les données peuvent être excluantes par leurs moyens de collecte. Souvent associé au développement du digital, les données sont très souvent rendues disponibles de façon massive par les outils numériques. Via des plateformes en ligne, collecter des

²⁸ Doug Laney, Meta Group, 2001

²⁹ D. Boyd & K. Crawford, *Critical questions for Big Data* in Information, Communication and Society, 2012

³⁰ Comme pour les données des réseaux sociaux qui ont tendance à être biaisées, L. A. Schintler, R. Kulkarni, *op.cit.*

informations en large quantité est extrêmement facile. Cependant, toute la population Française n'est pas connectée, ni connectée de la même manière. Que ce soit par des disparités d'infrastructures³¹ ou d'usages³², des inégalités existent dans l'accès au numérique et par conséquent dans la production de données. Si les données peuvent apparaître comme un moyen de redonner une voix au peuple³³, ce n'est souvent que la voix d'une partie du peuple, la plus connectée, qui est entendu.

Ainsi, les données peuvent présenter par nature un caractère excluant, qu'il soit relatif à leur surnombre, leur véracité ou leur origine. Les risques pour les politiques publiques de retomber dans une construction excluante sont également présents avec l'usage des mégadonnées.

B) Des données collectées qualitatives ouvrant de nouvelles voies d'évaluation

Comme défini plus haut, les politiques publiques se créent en réponse ou en prévention à une problématique. Un moyen d'évaluer leur efficacité serait alors de mesurer l'état de la problématique visée pendant ou à la suite de l'action publique. Ainsi, en sélectionnant les indicateurs de mesure adéquats, le nombre de variables à explorer se réduit et par extension le risque de se retrouver submerger par un trop grand nombre de variables. De plus, avec l'automatisation et la digitalisation croissante des services administratifs, le chemin qu'emprunte une politique publique de son concepteur à l'administré et l'effet que celle-ci a sur ce dernier sont renseignés dans un système d'information. Par conséquent, des données peu biaisées et qualitatives sont rendues disponibles. Ces données sont également inclusives. Le plus souvent le bénéficiaire n'a pas à rapporter les effets la politique qu'il reçoit puisque le traitement informatisé a déjà enregistré ces informations en amont. La mise en œuvre digitalisée d'une politique publique devrait donc fournir un retour d'informations sur la pertinence de cette dernière sur l'ensemble de ses destinataire. Le risque d'exclusion développé dans la partie précédente est minime avec de tels programmes. Les données collectées à la suite de la mise en place d'une politique publique sont dès lors relativement fiables et préservées des risques que l'utilisation des données massives peuvent présenter. Cependant, une telle observation n'est que théorique et pourrait être à nuancer en fonction des choix d'implémentation des politiques publiques et les potentielles erreurs humaines telle les erreurs de saisie.

Une telle qualité théorique des données issues des politiques publiques ouvrent la voie à une nouvelle organisation de l'évaluation au sein de la fonction publique. De par la connaissance métier que possède chaque administration maître d'œuvre de politiques publiques et les données produites par celle-ci, apparaît un terrain favorable à une autonomie de chaque administration dans l'évaluation de ses politiques. Les méthodes de mesure issues

³¹ <https://maconnexioninternet.arcep.fr/>

³² E. J. Schweitzer, *Digital divide*, Encyclopedia Britannica, 2008

³³ E. Helderop, T. H. Grubestic, T. Alizadeh, *Data deluge or data trickle ? Difficulties in acquiring public data fir telecommunications policy analysis*, 2018

du big data, bien qu'elles soient nouvelles, sont pour beaucoup facilitées par le développement d'outils automatisant ces tâches. Les « Entreprise Ressource Planning », par exemple, permettent de calibrer des outils d'analyse en fonction des données produites par une organisation et du métier de celle-ci. Là où la statistique étatique s'était centralisée au fil des décennies et régimes, l'apparition des données massives pourrait induire un effet inverse en rendant chaque administration indépendante en terme d'évaluation de son action.

Devant le risque de dérive endogame d'une auto-évaluation des administrations, des précautions sont évidemment à prendre afin d'éviter toute dérive. Mais s'il est admis que l'administration puisse s'auto-juger³⁴, ne pourrait-elle pas également s'auto-évaluer ?

Conclusion

Influencées par le courant néo-libéral anglo-saxon des années 80, l'évaluation des politiques publiques a opéré sous le dogme des catégories pendant plus de trente ans. Par manque de pertinence dans leur construction et d'évolution, ces catégories utilisées pour décrire, observer et gouverner devinrent obsolètes, menant au développement de politiques chiffrées et hors sol. Afin de palier à ce manque de justesse, des méthodes venues de l'économétrie commencèrent à être implémentées. Les données massives, fournissant des informations en grand nombre, pourraient être couplées à ces nouvelles méthodes en information ouvrant alors la voie à de nouvelles possibilités d'évaluation.

Cependant, l'utilisation des données massives n'est pas exempt de tout risques. De par leur volume et leur moyen de collecte, les informations fournies par les mégadonnées ont un risque non négligeable d'exclure une partie importante de la population. Néanmoins, avec le développement du numérique et la digitalisation croissante des services dans l'administration publique, une politique publique devient traçable. De sa conception à son potentiel bénéfique sur les administrés, il est possible de suivre son cheminement. Dès lors, si les variables à observer sont soigneusement sélectionnées dans cet ensemble de données, les risques de submersion se contractent. De plus, avec un tel système la collecte de données peut s'affranchir d'une collecte directe au près des bénéficiaires de la politique, réduisant ainsi les problématiques d'exclusion. L'implémentation de méthodes issues des mégadonnées seraient dès lors une occasion de rendre l'évaluation des politiques publiques plus inclusives et ouvrirait la voie à une autonomisation des services dans ce domaine.

³⁴ Henrion de Pansey, *De l'autorité judiciaire en France* (1818)

Bibliographie

Gil Press, « A Very Short History Of Big Data », *Forbes*, 9 mai 2013

Google trends, « Big Data », consulté le 13/03/2021

JORF n°0193 du 22 août 2014

K. Carillo, *Le big data, une révolution qui transforme l'entreprise en profondeur*, Harvard Business Review France, 2017

Yves Eychenne, Jean-Charles Cointot, *La Révolution Big data: Les données au coeur de la transformation de l'entreprise*, 2014

M. Pinet, *Histoire de la Fonction Publique en France*, 1993

Louis Rolland, *Précis de droit administratif*, 1926

Maurice Hauriou, *Précis de droit administratif et de droit public général : à l'usage des étudiants en licence et en doctorat ès-sciences politiques*, 1901

T. Birkland, *Introduction to the policy process : Theory, concepts and methods for policy making*, 2001

L. A. Schintler, R. Kulkarni, *Big Data for Policy Analysis : The Good, The Bad and the Ugly*, Review of Policy Research, 2014

Y. ISSELIN, Répertoire numérique de la sous-série F20, Archives nationales (Paris), 2010
INSEE.fr, *Données historiques de la Statistique générale de France*, 2010

A. Desrosières, *La politique des grands nombres. Histoire de la raison statistique*, 2010

Loi de finances du 27 avril 1946

E. Monnier, (1992), *Évaluation de l'action des pouvoirs publics*, 2^e Édition, Economica, Paris
Commission Européenne, " Focus on Results: Strengthening Evaluation of Community Activities", ref SEC(2000)1051

E. Pierru et A. Spire, « Le crépuscule des catégories socioprofessionnelles », *Revue française de science politique*, vol. 58, n° 3, 2008.

D. Cardon, *A quoi rêvent les algorithmes*, 2015

Dictionnaire Larousse

G. Casela, R. Berger, *Statistical Inference*, 2008

J. Hillje, *Les oubliés – Entretiens où prospère le vote extrême*, Das Progressive Zentrum, 2018

Doug Laney, Meta Group, 2001

D. Boyd & K. Crawford, *Critical questions for Big Data* in Information, Communication and Society, 2012

<https://maconnexioninternet.arcep.fr/>

E. J. Schweitzer, *Digital divide*, Encyclopedia Britannica, 2008

E. Helderop, T. H. Grubestic, T. Alizadeh, *Data deluge or data trickle ? Difficulties in acquiring public data fir telecommunications policy analysis*, 2018

Henrion de Pansey, *De l'autorité judiciaire en France* (1818)