

R F A S

Revue française des affaires sociales

Appel à contribution pluridisciplinaire sur :

« Big data et protection sociale »

Pour le numéro 4 - 2017

Cet appel à contribution s'adresse aux chercheurs en mathématiques appliquées, en informatique, en économie de la santé, en économie et économétrie, en droit, en philosophie et sociologie, ainsi qu'aux acteurs du champ de la protection sociale.

Les articles sont attendus avant le **31 mars 2017**.

La société contemporaine connaît de rapides changements dont l'un des traits distinctifs est la numérisation de la plupart des secteurs d'activité. Cette numérisation s'accompagne du développement des activités de réseau (objets connectés, réseaux sociaux, forums, plateformes numériques), qui génèrent des données toujours plus nombreuses, des traitements de plus en plus automatisés, complexes et rapides. La donnée est alors au cœur de la création de valeur dans le numérique et, plus largement, dans l'économie. De nouveaux acteurs (Google, Facebook, Uber, Airbnb, pour citer quelques-uns des plus emblématiques) émergent et remettent en question, souvent très violemment, le modèle économique des acteurs installés parfois depuis des siècles. Certains secteurs qui interagissent avec le champ social connaissent ainsi et connaîtront dans un avenir proche des bouleversements particulièrement profonds, notamment la santé, le logement, la mobilité, l'assurance, la finance ou l'éducation. La généralisation de l'exploitation de cette masse croissante de données, tant par la taille que par la diversité crée de nouvelles connaissances, de nouveaux services et de nouvelles organisations sociales. Elle est souvent dénommée « Big data », ou, en français, traitement de données massives.

La quantité et la variété grandissantes de données stockées, ainsi que les capacités de traitement informatique toujours plus puissantes ont ainsi donné naissance à un nouveau domaine de recherche et d'applications, à la frontière entre la statistique et l'informatique. Le Big data fait tout autant référence à certaines caractéristiques des données, qu'aux méthodes utilisées pour les exploiter, ainsi qu'aux finalités en jeu. Sa définition et ses contours sont flous et évolutifs, mais suffisamment présents dans le débat public (colloques, ouvrages, rapports...) et dans les travaux académiques (au-delà des nombreux articles publiés dans des

revues thématiques, le Big data a même sa revue anglo-saxonne¹) pour justifier que la RFAS y consacre un dossier. Dans notre vie quotidienne en France et à l'étranger, le Big data intervient également via la génération de publicités personnalisées lors de la consultation des boîtes de messagerie, la localisation des taxis Uber, la surveillance des lieux publics, la prédiction du développement du diabète ou l'orientation précoce dans les services d'urgence grâce à la mobilisation de données hospitalières (Israël), la détection de la fraude aux allocations, l'obtention d'un crédit immobilier, etc.

Le concept de Big data a émergé à la fin des années 1990 et a trouvé un écho large à la fin des années 2000, lorsqu'il a trouvé sa définition la plus répandue, celle des « 3V » : volume, variété et vélocité. Il sera ici retenu dans une acception plus large, reposant certes sur le croisement de données aux types et aux finalités originellement différentes (« variété »), souvent volumineuses (pas systématiquement) mais il englobera aussi un ensemble de techniques d'analyse et de traitements relevant de l'apprentissage statistique.

On s'intéressera aux deux grandes catégories d'usages du Big data telles qu'elles émergent des projets en cours : des usages à vocation de connaissance et de recherche, permettant d'éclairer les déterminants des comportements, des inégalités, de l'état de santé, etc. ; et des usages, en fort développement actuellement, se rapportant à des cas individuels, à finalité d'aide à la décision : aide à la recherche d'emploi, au diagnostic et à la stratégie thérapeutique, etc.

Polymorphes, les Big data recouvrent des données « structurées » traditionnelles dans le champ des sciences humaines et sociales, comme le sont les fichiers administratifs (données des hôpitaux et de l'assurance maladie, données de gestion des caisses d'allocation familiale ou de Pôle emploi, etc.) ou les données d'enquêtes. Leur nouveauté tient au fait qu'elles portent également sur des données dites « non structurées » comme le sont les données textuelles, les données d'imagerie, les métadonnées, les données des réseaux sociaux, des forums, etc. Le champ des données sur lequel peut s'appuyer la recherche en santé et en protection sociale est ainsi en forte expansion, et s'enrichit de données produites en dehors du système de protection sociale : on pense évidemment aux données produites par de nouveaux acteurs privés (via les objets connectés en santé notamment), mais aussi à des données produites dans des contextes complètement différents et pour des finalités n'ayant a priori pas de lien avec la santé et la protection sociale (e-mails envoyés par des consommateurs ou des usagers, achats au supermarché, passages à proximité des antennes de téléphonie mobile, etc.). Certaines données utiles au Big data relèvent de l'open data (elles sont ou pourraient être en accès libre car elles ne sont pas protégées par un secret prévu par la loi : secret de la vie privée, secret des affaires...) ; d'autres sont protégées par un secret légal, qui ne peut alors être levé que dans certaines conditions et pour certaines finalités. Ce sujet pourrait particulièrement faire l'objet d'articles juridiques.

Cet appel à contributions balaie de manière large les enjeux liés au Big data dans le champ de la RFAS : nouvelles techniques statistiques, nouveaux usages individuels et collectifs, nouvelles régulations. Les enjeux éthiques et juridiques feront l'objet d'une attention particulière. Des éclairages apportés par des expériences étrangères seront également bienvenus.

¹ <https://journalofbigdata.springeropen.com/>

Axe 1 : Statistique classique et apprentissage statistique

Le Big data recouvre un ensemble de méthodes statistiques depuis longtemps connues mais remises au goût du jour par l'existence de ces données massives et des capacités techniques désormais disponibles pour les traiter. Recherche de signaux faibles, détection de profils, réseaux de neurones... viendraient compléter (voire supplanter ?) la statistique classique. Le Big data se caractérise notamment par sa capacité à mettre au jour des corrélations inconnues et cherche essentiellement à les observer et à prédire des événements (ce qui peut poser des difficultés lorsque les corrélations sont instables dans le temps) et non à les expliquer. Quelles sont donc les spécificités du Big data face aux méthodes « traditionnelles » ? N'existerait-il pas des passerelles entre statistique classique, qui repose sur la définition d'un modèle et la recherche d'une causalité, et l'apprentissage statistique dont l'objectif premier est la prédiction d'une variable d'intérêt ? De nombreuses méthodes semblent en effet leur être communes. La RFAS souhaite s'interroger sur leurs liens. Sur ces sujets techniquement complexes, la Revue attend des articles pédagogiques.

En parallèle, une réflexion sur l'évolution du système statistique public, français, européen, voire international, sous l'impulsion du développement du Big data, serait bienvenue.

Plus spécifiquement dans le champ des études sur le médicament, les études « en vie réelle » permettent le plus souvent d'analyser le mésusage et les effets indésirables. Des débats existent cependant sur la valeur probante de ces études pour démontrer l'efficacité des médicaments, alors qu'il est complexe, en dehors du laboratoire, d'imputer à coup sûr un effet à un facteur en particulier, au contraire des essais randomisés en double aveugle. Face à l'afflux de données de plus en plus nombreuses et de plus en plus riches, la possibilité de recourir aux études en vie réelle, au moins en complément des études cliniques, représente néanmoins un enjeu majeur, en termes de coûts pour les laboratoires pharmaceutiques, mais aussi et surtout en termes de connaissance et donc de prise de décision pour les pouvoirs publics.

Par ailleurs, il est légitime de s'interroger sur le statut de ces nouvelles données, au regard des préoccupations classiques des statisticiens : contexte de leur production, qualité, reproductibilité, représentativité. L'opposition entre représentativité et exhaustivité en particulier constitue l'un des angles morts de bien des discours contemporains sur le Big data. Il est probable que ces préoccupations « traditionnelles », garantes de la qualité scientifique de la statistique, ne disparaîtront pas : la RFAS accueillera volontiers des contributions explorant cette question.

Axe 2 : Quels usages du Big data au sein de notre système de protection sociale ?

Le Big data bouleverse des secteurs entiers de l'organisation économique et sociale (transports, télécommunication, industrie, finance, assurances...). Comme bien d'autres secteurs de l'industrie et des services, la plupart des organismes de protection sociale sont concernés par le Big data et le seront de plus en plus. Ils doivent donc s'interroger sur ses enjeux : nouveaux services, nouvelles solidarités, efficience, mais aussi fragilisation des solidarités existantes, voire remise en cause du modèle de la protection sociale issu du XX^e siècle, assis sur la solidarité professionnelle ou nationale, organisé autour de la redistribution de prestations monétaires couvrant les principaux risques de la vie et de services peu individualisés.

En matière de santé, le ministère de la Santé et des Affaires sociales a lancé en 2015 une réflexion qui a questionné les usages possibles du Big data en santé². Le Big data, allié aux potentialités ouvertes par la génomique et les biotechnologies, fait naître de (trop ?) grandes attentes en matière de prévention, de prise en charge des patients, de médecine de précision, voire de médecine personnalisée. Le cancer, par exemple, est une pathologie dont la prise en charge pourrait connaître des progrès considérables dans l'avenir, avec des traitements « sur-mesure » en fonction des spécificités génétiques et biologiques de la tumeur et en tenant compte de l'environnement du patient et de son mode de vie. La connaissance des liens et causalités entre environnement et santé devrait également bénéficier fortement des apports du Big data.

Au-delà de la santé, le Big data pourrait sans doute faire émerger de nouveaux usages dans les autres domaines de la protection sociale, afin de mieux lutter contre la perte d'emploi, la pauvreté et l'exclusion sociale, l'invalidité... On pense en particulier au domaine de l'emploi et de la formation professionnelle : des expériences assises sur le Big data apparaissent qui visent à mieux accompagner les demandeurs d'emploi et les recruteurs. Des exemples et les enseignements qui peuvent en être tirés seraient bienvenus dans le dossier.

L'ONU, au travers de l'initiative *United Nations Global Pulse*, ainsi que la Banque Mondiale, promeuvent également le rôle essentiel que le Big data peut jouer dans la lutte contre la pauvreté dans les pays en voie de développement. Des expériences en la matière pourraient être rapportées. Dans les pays développés, ne pourrait-il pas venir en aide aux personnes en exclusion sociale, par exemple pour lutter contre l'isolement social via les réseaux sociaux et les objets connectés ?

Tous les champs de la protection sociale qui font intervenir des services pourraient à terme être concernés par le Big data, via la création de plateformes (qu'elles soient collaboratives ou qu'elles permettent le développement de marchés bifaces³) et la mise en œuvre de services personnalisés : l'accompagnement de l'invalidité (handicap, dépendance), la garde d'enfants, etc. Ces nouveaux usages posent cependant la question des normes de qualité et de sécurité, ainsi que celle des modalités de leur encadrement.

Dans chacun de ces champs, on s'interrogera sur le rôle que peuvent jouer les acteurs institutionnels financeurs (créateurs, promoteurs, financeurs de nouveaux services ?) et les organismes (publics et privés) producteurs des services, ainsi que sur le rôle régulateur de l'État.

Dès lors que le Big data est porteur d'usages nouveaux, d'organisations et de collaborations innovantes, on peut s'interroger sur sa capacité à améliorer l'efficacité globale de notre système de protection sociale, soit par l'amélioration de la qualité du service rendu, soit par la maîtrise ou la réduction des dépenses sociales. On pense notamment aux exemples suivants :

- l'amélioration du service rendu à l'utilisateur ou au patient et la création de nouveaux services : information sur la qualité des soins ou des services de protection sociale, médecine

² http://social-sante.gouv.fr/IMG/pdf/colloque_big_data_synthese.pdf.

Voir aussi le rapport de l'Institut Montaigne, <http://francestrategie1727.fr/wp-content/uploads/2016/02/resume-objets-connectes-institut-montaigne.pdf>

³ Ces marchés ont pour point commun de mettre en relation deux catégories distinctes de clients par l'intermédiaire d'un agent ou d'une "plate-forme", Voir par exemple http://www.autoritedelaconurrence.fr/doc/ocde_cp_06_2009.pdf

personnalisée, accès à l'offre de soins, télémédecine, plateformes, dans tous les domaines de la protection sociale, gestion de la qualité des services, création d'entreprise et gestion des prélèvements, etc. ;

- la meilleure gestion des moyens, dans le secteur sanitaire comme dans le secteur médico-social : gestion de l'offre de soins, coordination des acteurs, mise en œuvre des recommandations de bonne pratique et détection des pratiques médicales à risque, gestion des achats, etc. ;

- l'évaluation des politiques publiques, de leur efficacité et de leur impact sur les citoyens ;

- la diminution du non recours aux prestations par une détection précoce des personnes potentiellement bénéficiaires mais non recourantes ;

- la détection de la fraude, que cette dernière émane des allocataires, des professionnels, des patients.

Des articles théoriques et prospectifs, et d'autres plus pratiques décrivant des expériences passées ou en cours, seront bienvenus dans le dossier.

Le fonctionnement des acteurs de la protection sociale pourra également être étudié : ces nouveaux outils et services sont en effet susceptibles d'avoir des conséquences profondes sur les organisations existantes, en matière de systèmes d'information, de ressources humaines, de relation avec les assurés.

Axe 3 : Comment concilier Big data et solidarité ?

Le Big data peut favoriser des services nouveaux et des solidarités nouvelles, mais il peut aussi menacer des solidarités existantes et accroître les inégalités. Ce dossier de la RFAS voudrait explorer cette tension, située au cœur des nouveaux usages envisagés. Le champ de l'interrogation est vaste, prospectif, et peut évidemment donner lieu si l'on n'y prend garde à un certain catastrophisme. L'objectif de ce dossier se veut pragmatique et consiste à se placer au plus près des usages connus ou envisagés du Big data.

La problématique la mieux balayée concerne les évolutions de la prise en charge en santé. Les outils numériques rendent nécessaire une éducation à la donnée ; le bon usage des algorithmes repose sur leur transparence et leur lisibilité. De manière plus générale, on associe souvent santé numérique, Big data et autonomie des patients. Mais celle-ci n'est possible que si le patient est informé, capable et soucieux de l'être. Le Big data peut ainsi faire peser un risque de segmentation croissante entre les patients « autonomes » et les autres. De la même manière, le développement d'une médecine plus prédictive, édictant des diagnostics probabilistes, peut créer des inégalités entre les patients informés et les autres, en renvoyant à la responsabilité individuelle des décisions qui ne peuvent pas toujours l'être. Cela pose en miroir la question du droit à la non information des patients : un patient peut ne pas souhaiter connaître le risque encouru de développer telle ou telle maladie. L'essor du Big data invite ainsi à revisiter les concepts de droits des patients et de démocratie sanitaire.

Un autre usage du Big data en santé qui est souvent discuté renvoie à l'assurance, avec un risque d'individualisation des garanties et des contrats. Si des dispositions régissent strictement les modalités de segmentation des contrats selon des caractéristiques individuelles,

si des mesures existent et ont été renforcées dans la loi de modernisation de notre système de santé pour organiser, en matière de santé, un droit à l'oubli interdisant de stigmatiser les personnes ayant été concernées par des problèmes de santé spécifiques, si l'état de santé est lui-même reconnu comme un critère au vu duquel une situation de discrimination peut être caractérisée (cf. art. 225-1 du Code pénal), le Big Data met néanmoins ces principes en tension et appelle à discuter les cadres éthique et réglementaire. L'exemple le plus parlant est celui de l'assurance en santé. La contradiction entre l'individualisation rendue possible par des données de plus en plus nombreuses et la mutualisation qui est au cœur même du principe de l'assurance, est aujourd'hui un objet de réflexion et de préoccupation pour l'ensemble des acteurs de l'assurance.

Il serait intéressant de rapporter des expériences internationales de modulations tarifaires de la couverture assurancielle privée en santé en fonction des comportements effectivement observés. Cela génère-t-il une prise de conscience des assurés ou cela entraîne-t-il plutôt une aggravation de certaines inégalités, notamment de prise en charge ?

Certes, tous les usages du Big data par des assureurs santé ne concernent pas directement la tarification individuelle ; certains des usages prioritairement envisagés à ce jour portent sur des services d'accompagnement, de prévention et d'information. Mais ces usages ne sont-ils pas également porteurs d'un risque de segmentation ou d'éviction des « mauvais risques » (c'est à dire des personnes dont les prestations sont plus élevées que les cotisations) ?

Au-delà de la santé et de l'assurance, les autres secteurs de la protection sociale pourraient également voir les solidarités actuelles fragilisées par les nouveaux outils et pratiques du Big data. Des questions de nature très différentes peuvent ainsi être posées. Si l'insertion dans la formation professionnelle et dans l'emploi repose de plus en plus sur des outils numériques, quid de ceux qui ne les maîtrisent pas ? Si le Big data, allié aux nouveaux outils de la domotique et de la surveillance à distance, améliore l'accompagnement des personnes âgées à domicile, sera-t-il complémentaire de l'aide formelle et informelle existante ou ne risque-t-il pas de s'y substituer en partie, au détriment du temps relationnel qui est nécessaire à un vieillissement réussi ?

A plus long terme, le Big data pose également la question de son impact sur les inégalités sociales, qu'elles soient de santé, mais aussi d'emploi, de conditions de travail, de revenu, ou d'éducation. Peut-on l'évaluer ? Que peut-on en attendre ? Des algorithmes « impartiaux » (si le concept même existe) permettant une objectivation totale de la prise de décision pour l'attribution des logements, des places en crèches, pour le choix d'une école ? Des personnes averties qui sauront en tirer profit et en user et d'autres moins informées qui pourraient en pâtir ?

Plus globalement, un questionnement peut-être encore très théorique au stade de développement qu'a atteint le Big data, pourrait porter sur la conciliation d'une protection sociale numérique avec l'architecture de la protection sociale issue du XX^e siècle. Les dispositifs actuels sont marqués par une grande complexité, reposent sur de nombreux acteurs et des circuits de financement pas toujours perçus par les citoyens. Comment cette situation peut-elle se concilier avec les principes nouveaux des services numériques, dématérialisés, personnalisés, adaptables et expérimentables ? L'individualisation des services qui est au cœur du Big data est-elle compatible avec la solidarité nationale qui fonde la sécurité sociale ? A quelles conditions ?

Au fond, le Big data ne risque-t-il pas de cristalliser la situation annoncée par Pierre Rosanvallon il y a vingt ans, qui voyait une remise en cause de l'État Providence et de la logique d'assurance sous la pression de la montée de l'individualisme ? D'autant plus que les anciennes solidarités ne s'affaiblissent pas seulement sous les coups de boutoir de l'individualisme : se créent en parallèle de nouvelles communautés numériques, potentiellement riches de nouvelles entraides, qui pourraient à terme se poser en concurrentes des solidarités traditionnelles de l'État-Providence.

Axe 4 : Le Big data, un changement sociétal ?

Le Big data n'est certes pas Big brother, en tout cas pas encore. Mais il en revêt certains aspects, à travers sa prétention à recueillir et à analyser une multitude de données sur l'ensemble de la vie humaine, individuelle et collective. Bien évidemment, cette dimension inhérente au Big data fait l'objet de nombreux débats et mises en garde. Comment s'assurer du respect de la vie privée et de la protection des données personnelles ? Le concept même de vie privée a-t-il encore un sens ou doit-il évoluer ?

Les autorités publiques font un effort pour s'assurer du respect des droits des personnes en termes de conservation et de traitement des données. Mais peut-on en dire autant de l'ensemble des données que nous générons et cédon sans en être forcément conscients ? Données de téléphonie mobile, objets connectés, e-mails, requêtes internet... L'ensemble de ces informations dont nous ignorons quelquefois l'existence nous identifient et sont utilisées à notre insu pour définir de nouveaux algorithmes à but le plus souvent lucratif. La protection de la vie privée, assurée en droit par la loi n°78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés et renforcée au niveau européen par le récent règlement général sur la protection des données⁴, pourrait devenir de fait un enjeu essentiel de la protection sociale, publique ou privée.

La prise en compte des résultats des algorithmes dans les décisions humaines (par exemple, en « automatisant » le management) constitue également un terrain de réflexion et de recherche lourd d'enjeux. Si la rationalité algorithmique peut apporter des progrès, individuels et collectifs (en matière de santé notamment), elle comporte des limites et des risques qui doivent être encadrés et débattus.

L'accompagnement des personnes (y compris des plus vulnérables) pour prendre conscience de leurs droits et les faire valoir, en matière de maîtrise des données personnelles, de consentement et d'enjeux liés aux algorithmes, ne pourrait-il pas à ce titre constituer un nouveau service, voire un nouveau service public ?

⁴ RÈGLEMENT (UE) 2016/679 du 27 avril 2017 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données <http://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A32016R0679>

□ Des informations complémentaires sur le contenu de cet appel à contributions peuvent être obtenues auprès de Magali Beffy ou de Franck von Lennep qui ont en charge la préparation du numéro, aux adresses suivantes :

magali.beffy@sante.gouv.fr

et

rfas-drees@sante.gouv.fr

□ Les auteurs souhaitant proposer à la revue un article sur cette question devront l'adresser accompagné d'un résumé et d'une présentation de l'auteur (cf. les « conseils aux auteurs » de la RFAS [en ligne] <http://drees.social-sante.gouv.fr/etudes-et-statistiques/publications/revue-francaise-des-affaires-sociales/>) à cette adresse:

rfas-drees@sante.gouv.fr

Avant le 31 mars 2017