

Chapitre 8

Chantiers et pistes d'action pour réduire les asymétries d'information

Jusque là, nous avons décrit un bilan critique de l'impact de l'ouverture des données sur la transparence, l'économie et la transformation des administrations. Des données sensibles manquantes ou insuffisantes, des données pas assez fiables pour créer des services pérennes, une transformation bien réelle des pratiques administratives, des bricolages et des arrangements instables pour obtenir les données manquantes : difficile de se satisfaire pleinement de la situation actuelle. Que faire alors ? Ce chapitre distingue les chantiers prioritaires à investir et désigne des pistes d'action à conduire pour que les données ouvertes réduisent effectivement les asymétries d'information. Il découle à la fois d'une réflexion collective avec des acteurs proches dans une séance de travail¹, de la reprise de propositions déjà formulées dans différents rapports et de mes propres propositions originales .

Le premier chantier vise à obtenir les données manquantes. On le voit dans de nombreux domaines, les données ouvertes ne sont pas disponibles ou ne sont pas produites. Ce chantier passera selon moi d'abord par renforcer le droit d'accès, encore trop contourné en France, à produire des données d'intérêt général dans certains secteurs prioritaires lorsque les acteurs publics n'en disposent pas et à développer la concertation entre producteurs et réutilisateurs.

Le second chantier consiste à améliorer la fiabilité des données ouvertes. Les réutilisateurs potentiels sont souvent dissuadés par des données qui demandent un investissement trop conséquent pour devenir exploitables. Pour atteindre l'objectif de rendre les données plus fiables, il me paraît essentiel de pérenniser à la fois les projets d'open data et les données ouvertes. Ce chantier passe aussi par le développement d'incitations pour améliorer la qualité et la fiabilité des données ouvertes.

Enfin, pour que l'ouverture des données réduise les asymétries d'information et agisse comme un contre-pouvoir, il me paraît essentiel de développer la culture des données. Si les compétences liées aux données ne sont pas mieux réparties dans la

¹. Je remercie chaleureusement Marin Dacos, Charles Nepote et mes collègues marseillais de Dataactivist (Joël Gombin, Allyson Pallisser, Mathieu Morey) pour leur contribution lors d'une réunion de travail lors de laquelle ils ont imaginé certaines des pistes d'action de ce chapitre. J'espère qu'ils se reconnaîtront dans les propositions qui suivent.

société et dans les organisations, l'ouverture des données risque de donner plus de pouvoir à celles et ceux qui en ont déjà. J'aborderai ici certaines des propositions du mouvement de la littératie de données (*data literacy*) qui défend l'idée que les données ne doivent pas être réservées à une élite de spécialistes. A cette fin, les pistes d'action évoquées ici consistent à placer les données au cœur des politiques publiques, développer la culture critique des données et à former massivement aux données.

Obtenir les données qui manquent

Refonder le droit d'accès

Comme nous l'avons vu précédemment dans le chapitre 4, le droit d'accès en France comporte de nombreuses faiblesses. Ce contexte entrave considérablement l'ouverture des données. Si le droit d'accès est contourné, comment espérer que la transparence volontaire de l'administration sur des sujets sensibles s'exerce simplement et de manière pérenne ? Rappelons quelques-unes des faiblesses identifiées. Dans les administrations, le traitement des demandes n'est pas entré dans les routines de l'administration. En l'absence d'interlocuteurs précis, de processus bien huilés et de connaissance partagée du droit, l'attentisme est souvent la règle : la CADA statue très souvent pour des documents dont la communicabilité ne fait pas débat. Créée pour améliorer l'efficacité du service public et non pour renforcer la transparence, cette institution apparaît paradoxalement comme opaque pour le demandeur qui n'a pratiquement aucun échange direct avec elle. Relativement méconnue du public², avec une communication confidentielle et technique, elle est décrite par certains comme une « institution à la dérive.³ » Ses délais de traitement des dossiers explosent : 35 jours en moyenne en 2008 contre 94 jours en 2017 pour un nombre d'avis en hausse très relative : 4 548 en 2008, pour 5 315 en 2017⁴. Son budget reste faible (1,5 million d'euros au titre de l'année 2019), contre plus de 15 millions d'euros pour la CNIL, 4,2 millions d'euros pour la

². A titre d'exemple, sur le Wikipédia francophone, la page de la CADA a reçu 9 811 vues en un an contre 28 721 pour la HATVP et 88 775 pour la CNIL.

Source :

<https://tools.wmflabs.org/pageviews/?project=fr.wikipedia.org&platform=all-access&agent=user&redirects=0&start=2019-03-01&end=2020-02-29&pages=>

³. *Enquête sur la CADA, une institution à la dérive.* (s. d.). Consulté 2 avril 2020, à l'adresse <https://www.nextinpact.com/news/107986-enquete-sur-cada-institution-a-derive.htm>

⁴. La CADA pointée du doigt pour des « décisions absurdes anti-transparence ». (s. d.). Consulté 2 avril 2020, à l'adresse

<https://www.nextinpact.com/news/107495-la-cada-pointee-doigt-pour-decisions-absurdes-anti-transparence.htm>

Haute autorité pour la transparence de la vie publique (HATVP). Selon l'enquête pré-citée de Xavier Berne, journaliste à NextInpact et spécialiste du droit d'accès, ses avis témoigneraient d'une « application obsessionnelle de la loi » et d'une grande frilosité pour la transparence. Du fait de ces multiples dysfonctionnements, le droit d'accès reste relativement méconnu et les requérants sont des personnes rompues aux formalités administratives. Enfin, une grande partie des demandes relève de l'accès aux données individuelles contrairement aux Etats-Unis ou au Royaume-Uni où la majeure partie des requêtes sert à réclamer des comptes aux instances dirigeantes {Boustany:2013ge}. Le chantier de la refonte du droit d'accès est donc vaste, voici quelques pistes d'action.

Transférer les compétences en matière de transparence de la CADA vers la Haute Autorité pour la Transparence de la Vie Publique (HATVP) : La fusion de la CADA avec une autre autorité administrative indépendante est régulièrement évoquée. La CNIL a longtemps été pressentie pour cette fusion mais, à mon sens, cela réduirait le rôle de transparence de l'institution et le réduirait aux décisions relatives aux documents individuels. Je suggère plutôt de dissocier l'accès aux documents individuels qui relève selon moi plutôt de la CNIL et des droits relatifs aux données à caractère personnel traitées par les acteurs publics (voir proposition suivante). Ayant la transparence dans son ADN, la HATVP me semble plus adaptée pour cette mission. En charge du contrôle des déclarations de patrimoine et d'intérêts de certains responsables publics, ses missions portent aussi sur la déontologie des responsables et agents publics et le contrôle des activités de lobbying. Contrairement à la CADA dont les rapports annuels et les prises de position sont souvent pusillanimes, la HATVP n'hésite pas à dénoncer, parfois de manière virulente⁵, les faiblesses du cadre légal de la transparence de la vie publique. Sa notoriété est forte suite aux affaires Thévenoud ou Delevoye. Avec 74 transmissions d'affaires à la justice en sept ans, l'institution « s'est imposée comme un acteur important du paysage politique français de la transparence » selon Transparency France⁶. Intégrer le droit d'accès aux documents administratifs dans cette institution forte, bien dotée et connue du public serait un acte fort pour réaffirmer que le droit d'accès constitue la clé de voûte de la transparence. A côté des dispositifs de

⁵. *Frais de mandat, lobbying...* Le rapport au vitriol de la Haute Autorité pour la transparence. (s. d.). Consulté 17 août 2018, à l'adresse <https://www.nextinpact.com/news/106642-frais-mandat-lobbying-le-rapport-au-vitriol-haute-autorite-pour-transparence.htm>

⁶. [Communiqué] Didier Migaud pressenti à la présidence de la HATVP: une occasion de poursuivre la construction d'une culture française de l'éthique et de la transparence | Transparency International France. (s. d.). Consulté 2 avril 2020, à l'adresse <https://transparency-france.org/actu/communiqu%C3%A9-didier-migaud-presents-a-la-pr%C3%A9sidence-de-la-hatvp-une-occasion-de-poursuivre-la-construction-dune-culture-fran%C3%A7aise-de-lethique-et-de-la-transparence/>

transparence active comme les déclarations d'intérêts ou le registre des représentants d'intérêts, le droit d'accès renouvelé constitue un instrument sous-employé pour assurer un contrôle citoyen de la vie publique et favoriser son intégrité. Cet acte fort n'aurait de sens que s'il est couplé par une série de décisions pour renforcer le cadre juridique du droit d'accès, à commencer par l'abrogation du secret des affaires qui a souvent restreint la portée de la transparence administrative.

Transférer à la CNIL l'arbitrage de l'accès aux données individuelles traitées par les acteurs publics : Selon mes estimations, au moins un tiers des avis rendus par la CADA concernent l'accès aux données personnelles des individus⁷. Alors que le Règlement Général de Protection des Données (RGPD) consacre le droit d'obtenir la copie de ses données faisant l'objet d'un traitement auprès des acteurs publics comme privés, la CNIL n'est actuellement pas compétente pour traiter une plainte en cas de réponse négative ou incomplète. Dans ces cas, la CADA est compétente mais ces cas me semblent plus du ressort de la CNIL qui traite ces plaintes pour les acteurs privés. Donner compétence à la CNIL pour les demandes d'accès aux données individuelles permettrait de dégager du temps pour les avis relatifs à la transparence de la vie publique.

Créer un point central pour les demandes d'accès avec des interlocuteurs désignés : beaucoup de demandes n'aboutissent pas faute d'arriver au bon destinataire. La procédure de demande peut aussi paraître complexe pour les primo demandeurs et les sites des administrations ou collectivités ne promeuvent certainement pas ce droit. La HATVP pourrait centraliser les demandes aux différentes administrations concernées et, de ce fait, mieux faire respecter le droit avec des rappels automatiques des délais. Pour le traitement des demandes d'avis, la haute autorité disposerait de tous les éléments et échanges. Cette centralisation obligerait aussi la HATVP à tenir à jour le fichier des PRADA (ce n'est pas toujours le cas) et à responsabiliser ces agents en tant que récipiendaire unique des demandes.

Donner à la HATVP le pouvoir de sanctionner les administrations systématiquement récalcitrantes : la CADA défend, depuis son premier rapport d'activité en 1980, une approche de la pédagogie consistant à ne pas sanctionner les administrations contournant la loi et à espérer que les fonctionnaires et les élus intègrent le droit

⁷. Pour réaliser cette estimation, j'ai catégorisé manuellement 100 avis à partir de leur objet.

d'accès comme une procédure utile et normale de la vie publique {Boustany:2013ge}. 40 ans après, les témoignages d'administrations récalcitrantes voire hostiles au droit d'accès continuent de s'empiler. La pédagogie ne suffit plus dans un contexte de défiance généralisée à l'égard des gouvernants. La HATVP, bien connue des élus pour sa capacité à porter aux juges les déclarations fallacieuses, devrait bénéficier du pouvoir de sanctionner les administrations qui contournent systématiquement le droit d'accès. Ce pouvoir de sanction pourra s'exercer en dernier recours afin de crédibiliser le droit d'accès.

Créer un « référé communication » pour accélérer le traitement des demandes : une des propositions phare du rapport d'information de la sénatrice Corinne Bouchoux de 2014 consiste à créer un "référé communication" devant le juge administratif. Lorsque la question de la communicabilité du document a déjà été tranchée, le juge des référés du tribunal administratif pourrait être saisi ce qui permettrait au demandeur de bénéficier d'une ordonnance sous 48 heures. L'administration serait alors fortement incitée à ne plus tarder à communiquer les documents dont la communicabilité ne soulève pas de difficulté juridique et une grande partie des demandes d'avis qui portent encore sur des questions déjà tranchées au fond pourrait ainsi être prévenues.

Produire des données d'intérêt général

Comme on a pu le voir précédemment, beaucoup de données manquent encore à l'appel dans de nombreux domaines. La loi pour une République numérique qui prévoit l'ouverture des données par défaut n'est pas appliquée systématiquement, y compris dans des administrations bien dotées qui disposent des moyens d'avoir une stratégie d'ouverture ambitieuse. Mais, sur de nombreux sujets, le secteur public ne dispose pas nécessairement de données précises. La mise en données du monde ou « datafication » {Cukier:2013tu} a mis entre les mains des entreprises des volumes de données considérables qui défient souvent le niveau de connaissance des acteurs publics de certains phénomènes. La notion de « données d'intérêt général » étend l'ouverture au-delà des données du secteur public et envisage une ouverture ambitieuse par différents mécanismes.

Renforcer le cadre légal des données d'intérêt général : Les débats autour de la loi pour une République numérique ont donné lieu à l'émergence d'une nouvelle notion, celle de données d'intérêt général. Le rapport du député Cédric Villani sur l'intelligence artificielle le définit comme « une forme d'open data privé ayant vocation à s'appliquer aux données qui revêtent une importance particulière pour le fonctionnement efficace du marché ou des politiques publiques d'intérêt général. » {Villani:2018tf} En juin 2015, la secrétaire d'Etat chargée du numérique Axelle Lemaire a demandé la conduite d'une mission sur les données d'intérêt général pour mieux définir le périmètre de cette notion et sa compatibilité avec le cadre réglementaire {Cyterman:2015tc}. Cette mission a montré que, malgré le fait que la collecte de données relève de la liberté d'entreprendre et présente le caractère d'un droit de propriété, il est possible d'atteindre à ces droits pour un motif d'intérêt général de manière proportionnée et avec un encadrement légal suffisant. Selon le rapport, la jurisprudence indique quatre motifs d'intérêt général autorisant qu'il soit porté atteinte aux droits des personnes privées envers leurs bases de données : une conduite plus efficace de politiques publiques sectorielles, l'information des citoyens notamment en matière environnementale et sanitaire, la recherche scientifique et le développement économique. Afin d'éviter que le Conseil constitutionnel requalifie l'atteinte de privation ce qui nécessiterait une indemnisation coûteuse des acteurs privés, la mission propose d'examiner au cas par cas deux modalités possibles : un accès à ces données pour la seule puissance publique qui pourra mettre à disposition des données agrégées ou une ouverture des données, y compris pour les autres acteurs économiques. Ce dispositif permettrait en quelque sorte de « réquisitionner » les données du secteur privé à des fins d'intérêt général. La notion de données d'intérêt général est entrée dans la

loi pour une République numérique mais avec des ambitions au rabais⁸. L'ouverture des données d'intérêt général pourrait pourtant permettre de répondre aux problèmes posés par les asymétries d'information dans certains secteurs comme le logement ou la santé.

Créer des coopératives de données : inspirée par les communs, des ressources partagées et gérés collectivement par une communauté, la notion de coopératives de données est régulièrement défendue par des organisations comme l'Open Data Institute⁹ ou la Fondation Internet Nouvelle Génération (FING)¹⁰. Une coopérative de données est une organisation gouvernée démocratiquement par ses membres dans laquelle des individus décident volontairement de reverser leurs données, généralement en échanges de services et/ou d'une redistribution de la valeur économique des données collectives. Avec le droit à la portabilité permis par le RGPD, les citoyens peuvent récupérer leurs données à caractère personnel et les migrer dans le service de leur choix. À condition d'être contrôlé strictement par ses membres selon le principe coopératif « une personne = une voix », une coopérative de données constitue un levier pour disposer de données jusque-là inaccessibles. Par exemple, pour mieux comprendre les problèmes de pouvoir d'achat, des contributeurs pourraient reverser les données de leurs cartes de fidélité qui collectent panier après panier les prix des produits achetés. En échange d'une analyse de leurs habitudes de consommation, ces données pourraient servir notamment à identifier les produits dont les prix ont connu une hausse brutale dans les enseignes de la grande distribution.

Développer le *crowdsourcing* de données publiques : l'appel à la foule des contributeurs constitue un levier encore sous-exploité pour créer de grandes bases de données publiques. Pourtant, dans le champ professionnel des données, de nombreuses personnes sont désireuses de contribuer à des projets d'intérêt général. Les données étant ouvertes et librement réutilisables, la motivation des contributeurs peut être guidée par le fait de contribuer à une ressource partagée, idéalement gouvernée comme un commun par sa communauté de producteurs et

⁸. Les dispositions sur les données d'intérêt général permettent uniquement au ministre de l'Économie d'imposer à certaines entreprises la transmission de certaines données afin qu'elles puissent être intégrées dans des enquêtes statistiques de l'INSEE. Cette décision doit être précédée d'une concertation et d'une étude de faisabilité. Une amende de 25 000 euros peut contraindre les entreprises récalcitrantes.

⁹. *Data trusts: Lessons from three pilots (report)* – The ODI. (s. d.). Consulté 3 avril 2020, à l'adresse <https://theodi.org/article/odi-data-trusts-report/>

¹⁰. *Retour sur MyData2018: Quelle(s) approche(s) collective(s) des données personnelles ?* | MesInfos. (s. d.). Consulté 3 avril 2020, à l'adresse <http://mesinfos.fing.org/retour-sur-mydata2018-quelles-approches-collectives-des-donnees-personnelles/>

réutilisateurs. On peut déjà citer plusieurs exemples de projets publics de production de données réalisés par une foule de contributeurs. Etalab a déjà initié le projet OpenSolarMap qui vise à évaluer le potentiel photovoltaïque d'un bâtiment, l'opportunité d'installer des panneaux solaires, en analysant l'orientation de la pente d'un toit à partir d'images satellitaires. La plateforme permettait aux contributeurs d'indiquer l'orientation du toit ; 100 000 analyses ont été réalisées en un mois, un échantillon qui a permis de développer un classificateur automatique traitant les images satellitaires¹¹. Toujours initié par Etalab, on peut aussi citer le projet PIAF (Pour une Intelligence Artificielle Francophone) qui vise à créer un jeu de données permettant de trouver des réponses à des questions précises afin d'entraîner des intelligences artificielles capables de répondre à des questions en langage naturel. Le projet part du constat d'un manque de données d'entraînement pour les intelligences artificielles en langue française. Le jeu de données, qui sera ouvert à la fin du projet, est constitué à partir des contributions d'un millier de contributeurs attendus qui collecteront une centaine de milliers de questions-réponses dans un corpus issu de Wikipedia francophone. Ces projets montrent qu'il est possible pour des acteurs publics de s'appuyer sur les citoyens pour répondre aux besoins de données

¹¹. *Rapport 2016-2017 de l'administrateur général des données : La donnée comme infrastructure essentielle.* (s. d.). Consulté 5 septembre 2018, à l'adresse <https://agcd.data.gouv.fr/2018/04/10/rapport-2016-1017-de-ladministrateur-general-des-donnees-la-donnee-comme-infrastructure-essentielle/>

Développer la concertation entre producteurs et réutilisateurs

Un des grands enjeux de l'ouverture des données à l'heure actuelle consiste dans la rencontre entre l'offre, les données ouvertes, et la demande, les réutilisateurs actuels ou potentiels. Penser l'ouverture des données comme une « politique de l'offre » {Gourgues:2012wd} nous incite à imaginer de nouveaux dispositifs pour favoriser la concertation entre producteurs et réutilisateurs.

Créer un conseil national de la donnée : il existe déjà des institutions en charge de la concertation entre producteurs et réutilisateurs. On peut citer le Conseil National de l'Information Statistique (CNIS) créé en 1984 pour « faire apparaître les nouveaux besoins d'information ainsi que les préoccupations des utilisateurs » et offrir « une tribune pour interpeller le service statistique public sur les lacunes éventuelles dans l'observation des phénomènes économiques et sociaux ou dans la diffusion des résultats.¹² » Le CNIS rassemble une diversité de représentants des réutilisateurs : élus nationaux et locaux, syndicats professionnels, associations, organismes publics, universitaires... Dans *Statactivisme*, Bernard Sujobert a montré comment un groupe de travail au sein du CNIS composé notamment de syndicalistes et de chercheurs est parvenu à provoquer au sein de l'Insee une « révolution mentale » sur la question des inégalités {Sujobert:2014wf}. Alors que les données de l'Insee pointaient une stabilité des inégalités, ce groupe a milité pour obtenir des données au centile (au lieu du décile) qui montrent l'explosion des très hauts revenus en France. Dans un autre domaine, le Conseil National de l'Information Géographique (CNIG) a pour mission d'« éclairer le Gouvernement dans le domaine de l'information géographique » et de « prendre en compte les besoins exprimés et en particulier les besoins des utilisateurs » {FaureMuntian:2018vy}. A mon sens, ces instances sont indispensables pour répondre mieux aux besoins des usagers de données mais elles n'ont pas été conçues pour les « sources administratives », sous-produit des activités de gestion, collectées dans des services dont la production de données n'est pas la finalité {Desrosieres:2005ue}. Surtout, ces instances manquent de visibilité. Un Conseil National de la Donnée rassemblant CNIS et CNIG avec des prérogatives étendues à l'ensemble des données publiques permettrait de mieux organiser la concertation entre producteurs et réutilisateurs. Pour leur donner plus de visibilité, un collège d'experts sur le modèle du Conseil National du Numérique pourrait rendre des avis sur des sujets d'actualité et de prospective.

¹². Cnis—CNIS. (s. d.). Consulté 6 avril 2020, à l'adresse <https://www.cnis.fr/cnis/>

Étendre et renforcer le service public de la donnée : l'article 14 de la loi pour une République numérique a créé le service public de la donnée (SPD) qui vise à faciliter la réutilisation des jeux de données de référence, ceux qui présentent le plus fort impact économique et social¹³. La loi prévoit que ces données qui constituent une référence pour nommer ou identifier des produits, des services, des territoires ou des personnes (données pivots) doivent être mises à disposition avec un niveau élevé de qualité. Elles doivent être documentées en détail, mises à jour fréquemment et disponibles 99% du temps {CluzelMetayer:2018ui}. Une procédure de signalement doit permettre de corriger les erreurs rapidement. Ce dispositif, s'appuyant sur les principes en France du service public (continuité, égalité et adaptabilité) présente des atouts importants pour mieux répondre aux besoins des usagers. Il doit selon moi être renforcé sur deux volets. Premièrement, il était prévu une gouvernance ouverte du service public de la donnée ce qui signifie que les réutilisateurs doivent être consultés et informés lorsque la mise à disposition d'un jeu de données doit évoluer. Or, ce n'est pas toujours le cas¹⁴. Il faut donc faire en sorte que la gouvernance de ces jeux de données soit effectivement partagée, par exemple sous l'égide d'un groupe de travail du conseil national de la donnée. Deuxièmement, le périmètre du service public de la donnée devrait être étendu à une cinquantaine de jeux de données, pour couvrir une plus grande variété de domaines tout en conservant une masse gouvernable. On distinguerait alors les données à plus fort potentiel et on garantirait la fiabilité aux réutilisateurs.

Créer des plans nationaux d'ouverture sectorielle : l'expérience du plan national pour la science ouverte¹⁵ devrait créer des émules dans d'autres domaines. Dans

¹³. A l'issue d'une consultation publique organisée par Etalab en octobre 2016, neuf bases de données ont été sélectionnées au sein du service public de la donnée : la base Sirene des entreprises (Insee), le Répertoire national des associations (Ministère de l'intérieur), le Plan cadastral informatisé (DGFIP), le Registre parcellaire graphique (Agence des services et de paiement), le Référentiel à grande échelle (IGN), la Base adresse nationale (IGN, La Poste, OpenStreetMap-France), le Référentiel de l'organisation administrative de l'État (DILA), le Répertoire opérationnel des métiers et des emplois (Pôle-emploi), le code officiel géographique (Insee).

¹⁴. Lorsque l'Insee a changé le format et le contenus des données de la base Sirene des entreprises (qui fait partie du service public de la donnée) avec notamment des champs manquants, l'information a d'abord circulé au sein du forum TeamOpenData avant que l'Insee n'en informe les usagers sans les consulter. Comme l'a expliqué un des contributeurs du forum, cette décision déroge de l'esprit du SPD : « on pensait par contre qu'il n'y avait pas de problème avec Sirene car elle faisait partie du SPD. J'associe dans ma tête ces jeux de données à une infrastructure de la donnée sur laquelle on peut se reposer. La j'ai l'impression que l'on va rapprocher les rails d'un mètre l'un de l'autre et que mes trains ne pourront plus circuler dessus ... ¹ »

¹ INSEE et SIRENE... Changements à venir—#TeamOpenData. (s. d.). Consulté 3 septembre 2018, à l'adresse <https://teamopendata.org/t/insee-et-sirene-changements-a-venir/729/8>

¹⁵. Le Plan national pour la science ouverte : Les résultats de la recherche scientifique ouverts à tous, sans

son deuxième axe, il instaure à rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics. Cet objectif s'appuie sur une gouvernance ouverte et partagée à travers le comité de la science ouverte qui fédère les principales institutions et rassemble près de 200 experts sur le forum de la science ouverte. Dans de nombreux secteurs, les asymétries d'information restent encore la règle. Comme proposé dans une tribune « Vers une société de l'accès libre au savoir » initiée par Lionel Maurel, Silvère Mercier et Julien Dorra, la même démarche devrait être engagée par l'État dans d'autres champs, avec un Plan National pour la Culture Ouverte, un Plan National pour l'Éducation Ouverte, un Plan National pour la Santé Ouverte, portés par les ministères concernés. La réussite de ces plans repose sur l'inclusion d'une grande diversité d'acteurs dans la définition, la mise en œuvre et l'évaluation des actions.

Améliorer la qualité et la fiabilité des données ouvertes

Comme nous l'avons vu dans le chapitre 6, les problèmes de qualité des données ouvertes remettent en cause leur fiabilité et restreignent leur potentiel d'utilisation. La notion de qualité de données doit toutefois être prise avec des précautions car beaucoup de données sur des portails comme data.gouv.fr n'avaient jamais auparavant circulé hors de l'administration d'origine. Pour ce chantier, je propose deux pistes d'action : mieux pérenniser l'ouverture et créer des incitations à l'amélioration des données.

entrave, sans délai, sans paiement—Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation.
(s. d.). Consulté 6 avril 2020, à l'adresse <https://www.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html>

Pérenniser l'ouverture

Les réutilisateurs pointent souvent le problème de la stabilité de la publication de certaines données, il n'y a pour ainsi dire aucune garantie dans le temps de retrouver un jeu de données à jour à l'endroit où il a été déposé. Christian Quest, contributeur à OpenStreetMap et créateur du projet opendatarchives a listé plusieurs cas de figures de disparition de données :

- seule la dernière version du jeu de données est publiée ;
- les données temps réel sont rarement historisées ;
- les données historiques sont effacées lorsque le portail open data est mis à jour ;
- les données sont volontairement dépubliées comme dans le cas des données de ventes et achats de pesticides, semble-t-il à la demande de certains lobbys agricoles puis republiées...

Il faut aussi citer les cas où l'ouverture était un acte ponctuel et le travail dédié à la mise à jour des données n'est pas entré dans les missions ou la charge de travail des agents. Pour pérenniser l'ouverture, je propose trois pistes d'action : l'archivage des données, la structuration d'une organisation d'intérêt général pour les producteurs et les réutilisateurs et la relance du poste d'administrateur général des données.

Archiver les données publiques : Quelques jours après l'investiture du président Trump, la Maison Blanche a supprimé les pages sur les données ouvertes dans la section open.whitehouse.gov créées lors de l'administration Obama. Suite à la suppression unilatérale de ces données, de nombreux acteurs ont commencé à archiver des données publiques que le gouvernement pourrait supprimer¹⁶, notamment sur des sujets sensibles comme le changement climatique. Ce cas a mis en lumière le risque de la disparition de données pour lesquelles il n'existe généralement pas d'archives. En France, Christian Quest a lancé à l'été 2019 le site opendatarchives.fr qui stocke notamment l'ensemble des données publiées sur data.gouv.fr et leurs mises à jour. L'ensemble regroupe environ 5To de données.

¹⁶. *White House Open Data Disappears, Raising Transparency Questions* – MeriTalk. (s. d.). Consulté 6 avril 2020, à l'adresse <https://www.meritalk.com/articles/white-house-open-data-disappears-transparency-donald-trump-sunlight-foundation/>

Pour l'archivage des données, les initiatives de la société civile sont utiles mais on pourrait envisager de manière plus pérenne une intégration des données publiques à l'archivage des documents des administrations. De prime abord, l'immédiateté de l'ouverture des données et le temps infini de l'archivage semblent opposés mais les complémentarités sont nombreuses {Guyon:2015kn}. Déjà, les données publiques sont par définition des archives publiques auxquelles l'archivage peut apporter une profondeur historique. De nombreux jeux de données sont mis à disposition sur des portails en temps réel mais l'historique des données reste souvent invisible de l'utilisateur. Or, des données comme les horaires de transport en temps réel pourraient trouver de nouveaux usages par l'archivage. Elles ne serviraient pas alors uniquement à fournir un service pratique dans l'immédiateté mais pourraient aider les citoyens à contrôler la qualité des transports ou à développer des modèles prédictifs des retards. Verser les données dans un SAE (systèmes d'archivage électronique) pourrait apporter des garanties quant à la fiabilité et la pérennité sur le temps long des données. Les archivistes pourraient aussi apporter une expertise précieuse sur la documentation des données, l'analyse de la valeur des données et la normalisation des métadonnées. Les archivistes ont une vision transversale et exhaustive de la gestion de l'information dans les administrations mais ils restent encore en marge des projets d'ouverture de données. Pourtant, l'association des archivistes de France s'est saisie de longue date du sujet mais cet intérêt s'est traduit dans de trop rares collaborations. Le service interministériel des Archives de France (SIAF) pourrait se saisir de la question de l'archivage des données publiques.

Fédérer producteurs et réutilisateurs de données autour d'une organisation d'intérêt général : on a pu le voir dans le chapitre 2 sur l'histoire de l'ouverture des données publiques en France, les politiques d'open data sont régulièrement remises en cause lors des phases d'alternance. Certaines données ne sont pas mises à jour, retirées ou « édulcorées » pour perdre de leur sensibilité politique. Face à ces situations, producteurs et réutilisateurs de données arrivent souvent en ordre dispersé pour défendre leurs intérêts. Les producteurs de données, responsables de projets d'open data et réutilisateurs de données ont pourtant des intérêts communs : pérenniser les projets d'ouverture de données, maintenir des moyens à cette politique, renforcer le cadre juridique en faveur de l'ouverture, déverrouiller certaines réticences politiques ou administratives... Des associations existent mais elles ne parviennent pas à fédérer l'ensemble du spectre des parties prenantes de l'open data. Du côté des agents publics, l'association Open Data France fait un travail de veille, d'outillage et de production documentaire indispensable. Mais c'est une association de collectivités locales (les administrations centrales n'y sont pas représentées) et elle est gouvernée par des élus locaux ce qui peut limiter sa

capacité d'action. Sa gouvernance par les élus l'écarte du fonctionnement d'une association d'agents ou de techniciens¹⁷ dont les missions prévoient la promotion d'une profession et la défense de ses intérêts. Les chargés de projet open data, une centaine de personnes en France, ont intérêt à défendre cette profession émergente et à maintenir les moyens associés à ces politiques. Pour les réutilisateurs, le GFII (groupement français des industries de l'information) pourrait jouer un rôle de représentation mais cette association rassemble principalement de gros industriels aux intérêts souvent divergents, notamment sur la question de la tarification des données. Sur les enjeux citoyens, des associations comme Regards Citoyens, SavoirsCom1 ou Open Knowledge France réalisent une activité de plaidoyer indispensable mais intermittente et non professionnalisée. Une association (ou une coalition) fédérant les producteurs et les réutilisateurs dans leur diversité permettrait de mieux défendre l'ouverture des données et de faire avancer efficacement certains sujets qui entravent la fiabilité des données.

Relancer l'administrateur général des données : le rôle de Chief Data Officer est désormais répandu dans la plupart des grandes entreprises. L'État avait créé cette fonction d'administrateur général des données mais au départ d'Henri Verdier, le poste n'a pas été réattribué. L'administrateur général des données joue pourtant un rôle essentiel pour coordonner l'action des administrations centrales en matière de production, de circulation et d'exploitation des données. Sans chef d'orchestre, la politique d'ouverture des données fait face à des contournements dans les ministères et perd en cohérence du fait des silos organisationnels. La mission Etalab reprend en partie cette fonction mais elle agit au service des administrations qui la sollicitent et peut plus difficilement s'immiscer dans les politiques des administrations récalcitrantes. Renommer un(e) administrateur(trice) des données directement sous l'autorité du Premier ministre semble être un moyen simple de faire remonter l'ouverture dans la chaîne de valeur des données et d'en faire un levier pour favoriser une meilleure gouvernance des données de l'État.

¹⁷. On peut citer par exemple l'Association des Archivistes de France (AAF) ou l'Afigéo pour les professionnels de l'information géographique.

Créer des incitations à l'amélioration des données

Au-delà du caractère éphémère de certaines initiatives d'ouverture des données, les réutilisateurs font régulièrement remonter des problèmes de qualité des données. Ces problèmes qui prennent des formes très variées proviennent d'abord du fait que les données mises à disposition n'étaient pas initialement conçues pour circuler hors de l'administration. Ce qui apparaît comme une erreur ou une imprécision pour certains peut n'avoir jamais été un problème pour ses usages initiaux. L'amélioration de la qualité des données ouvertes se fait généralement de manière itérative, dans l'échange avec les producteurs de données. Encore faut-il pour les agents pouvoir intégrer ce travail de correction et d'amélioration des données dans leurs missions quotidiennes. Pour se faire, voici plusieurs pistes d'incitation pour améliorer la qualité des données.

Faire remonter la destination des données : dans sa liste des défis pour l'avenir de l'ouverture des données¹⁸, la Fing avait voulu réinterroger le principe de non discrimination des réutilisateurs : « conditionner les usages à une identification est une des lignes jaunes initiales de l'open data. » Ce principe qui vise initialement à enlever les entraves à la réutilisation place les producteurs de données dans une situation de méconnaissance des usagers. Pour beaucoup d'administrations, les statistiques d'utilisation et éventuellement les commentaires constituent les seuls indices d'un usage des données. Savoir qui a utilisé les données et pour quel usage leur est souvent impossible. Sans ces informations, il est difficile de rendre compte de l'utilité de l'ouverture des données et plus encore d'adapter les jeux de données aux besoins des usagers. Pourrions-nous tracer plus finement les usages sans pour autant les entraver ? L'association Telabotanica parvient à connaître ses usagers sans poser d'entraves à la réutilisation. Sur sa page de téléchargement des données¹⁹, elle propose de renseigner la destination des données en indiquant : « rien ne vous oblige à compléter le formulaire sur la destination des données. Toutefois, si vous exportez ces données pour un usage à titre collectif, celui-ci nous intéresse ! Merci dans ce cas de préciser à quoi vous destinez ces données en complétant le formulaire ci-dessous. » Ce simple formulaire est proposé de manière facultative. A ma connaissance, une telle initiative n'a pas été réalisée par d'autres producteurs de données ouvertes, en particulier dans les administrations alors qu'elle n'entrave aucunement les usages et ne nécessite aucun développement spécifique.

¹⁸. http://fing.org/IMG/pdf/Pre_sentation_des_de_fis_Open_Data_impact.pdf

¹⁹. <https://www.tela-botanica.org/ressources/donnees/telechargements/>

Faire entrer l'ouverture des données dans les fiches de poste des agents : de nombreux agents ouvrent des données hors de leurs missions officielles. La hiérarchie administrative ne prend alors pas en compte l'ouverture des données dans l'organisation du travail. Quand les données sont sensibles politiquement, les agents prennent un risque pour leur carrière car ce n'est pas dans leur missions. Il faudrait inclure l'ouverture des données dans les fiches de poste des agents concernés. Ce document décrit, dans la grande majorité des administrations, les fonctions qui incombent à l'agent et déterminent son évaluation, ses moyens mis à disposition ainsi que ses objectifs. On pourrait s'appuyer aussi le Règlement Général de Protection des Données (RGPD) prévoit la désignation d'un délégué à la protection des données (DPO ou *Data Protection Officer*), « chef d'orchestre » de la conformité en matière de protection des données de l'organisation. Le sujet de l'ouverture des données peut paraître orthogonal à celui de la protection des données à caractère personnel. Mais ils s'insèrent dans une question plus générale de gouvernance des données, c'est-à-dire la stratégie d'accès et de circulation des données au sein des organisations. S'appuyer sur un réseau de référents internes sur les enjeux liés aux données en leur attribuant des moyens constitue à mon sens un levier important pour ouvrir et améliorer les données.

Montrer les résultats : pour de nombreux décideurs, l'ouverture des données peut s'apparenter à une politique abstraite dont les résultats en matière de transparence, d'innovation et de transformation de l'administration peinent à voir le jour. Pourtant, les exemples de réutilisation qui ont produit un impact important sont nombreux et mériteraient d'être mis en avant. Pour les données locales, l'association Open Data France et la Gazette des Communes ont mis en place le prix « open data des territoires » qui récompensent les meilleures initiatives d'ouverture et de réutilisation de données. Au niveau national, Etalab animait jusqu'en 2016 le concours Dataconnexions qui célébrait les meilleures réutilisations de données ouvertes. Ces initiatives sont essentielles pour faire perdurer les projets d'ouverture de données et leur donner les moyens d'accroître leur impact. Un recensement des entreprises réutilisant des données, sans exhaustivité, pourrait aussi aller dans ce sens. Dans six pays dans le monde, GovLab a conduit l'Open Data 500²⁰ pour recenser par un sondage les entreprises réutilisant des données ouvertes et mieux comprendre leur usage. Cette démarche donne de la visibilité aux nombreuses réutilisations invisibles car réalisées en interne ou peu connues.

²⁰. <https://www.opendata500.com/>

Citer les agents : même si une meilleure intégration organisationnelle du travail d'ouverture des données est essentielle, cette incitation ne pourra pas être mobilisée pour l'ensemble des agents impliqués dans des projets d'open data. Si ce travail ne trouve pas de reconnaissance dans les missions et la carrière des agents, il est possible d'envisager d'autres formes de valorisations notamment symboliques. Pour cela, il peut être utile de se tourner de nouveau vers les sciences pour lesquelles le partage des données est une pratique ancienne et où la publication de bases de données, a été ainsi hissée au statut de productions scientifiques à part entière {Bowker:2000ws}. On peut ainsi s'inspirer des *data papers* ou articles de données qui sont des publications décrivant un jeu de données scientifiques brutes à l'aide de métadonnées examinées par les pairs. Les données sont déposées dans un entrepôt de données où elles sont stockées de manière pérenne. Publiés dans des revues spécialisées, les *data papers* informent la communauté scientifique de la disponibilité d'un jeu de données brutes et de son potentiel de réutilisation en montrant son originalité {Laurence:2014we}. La publication de données sous la forme d'un *data papers* présente donc une double incitation. Pour les réutilisateurs, elle indique des données originales pouvant être exploitées sous des formes variées. Pour les agents, cette publication permet de créditer les personnes ayant participé au travail d'ouverture des données et de rendre visible ce travail qui disparaît souvent lors de la publication. La revue par les pairs et le formalisme exigé dans un *data paper* permettraient aussi de garantir un niveau de qualité dans la documentation des données qui fait souvent défaut pour les données publiques. Pour concrétiser cette idée, on pourrait imaginer la création d'une revue française des données publiques qui proposerait une curation de données publiques originales pouvant être exploitées.

Créer un label de qualité des données : comme nous l'avons mentionné à plusieurs reprises, le travail d'amélioration de la qualité des données et de leur documentation est rarement intégré dans les missions de l'ouverture des données. De fait, ce travail est rarement valorisé et repose sur la bonne volonté des agents. Une incitation à l'amélioration de la qualité des données consisterait à créer un label de qualité décerné par les usagers sur data.gouv.fr et d'autres portails pour distinguer les producteurs de données faisant un effort considérable pour aider les réutilisateurs à exploiter leurs données. Cette distinction principalement symbolique serait aussi un moyen de légitimer pour ce travail auprès de la hiérarchie administrative et de créer une forme de compétition entre les administrations pour rendre les données les plus exploitables.

Développer la culture des données

Imaginons que nous ayons résolu les problèmes évoqués précédemment liés au manque de données et à leur fiabilité. Pourrait-on se satisfaire de cette situation ? On ne peut pas réduire les asymétries d'information uniquement en se concentrant sur l'offre de données. Notre attention doit aussi porter sur la demande sauf à vouloir maintenir le *statu quo* dans lequel seule une élite détient le pouvoir d'utiliser les données. Ce risque a été identifié dès 2011 par le chercheur canadien Michael Gurstein qui craignait que les politiques d'open data ne donnent que plus de pouvoir à celles et ceux qui en ont déjà (« *empowering the empowered* ») {Gurstein:2011we}. Gurstein plaide plutôt en faveur d'une approche fondée sur l'« utilité réelle » (« *effective use* ») des données qui garantit que les compétences liées aux données soient disponibles au plus grand nombre possible d'utilisateurs. Il invite les acteurs publics à ne pas se contenter de mettre en ligne les données et les rendre librement réutilisables mais à s'assurer que les citoyens soient capables de comprendre, d'interpréter et d'utiliser les données. Si ce n'est pas le cas, il invite les acteurs publics à rendre disponible des formations à l'utilisation et à la visualisation des données. Cette question de la disponibilité des compétences est souvent un impensé dans les politiques d'ouverture de données. Pour Gurstein, cela s'explique par le fait que la plupart des personnes qui plaident en faveur de l'ouverture des données sont elles-mêmes expertes du sujet. Ne pas mettre en place d'actions proactives pour favoriser une appropriation plus large des données risque de renforcer le pouvoir d'une élite qui dispose de ces compétences. Une élite souvent issue des couches les plus favorisées de la société.

Le développement de la culture des données conditionne aussi la pérennité des données ouvertes. Ouvrir des données demande un travail considérable qui est justifié par les agents publics par l'impact de la réutilisation. Pour certains auteurs, les données ouvertes n'ont pas de valeur en tant que telles, elles ne prennent de la valeur que si elles sont utilisées {Janssen:2012kk}. Si personne ne se saisit des données, ou tout du moins qu'on ne puisse pas prouver d'usage réel, il est en tout cas fort possible que les agents ne disposent plus des crédits pour les ouvrir ou qu'on leur retire cette mission de leurs responsabilités.

Pour répondre à ces enjeux d'inclusion essentiels pour l'ouverture des données, le

mouvement de la littératie de données (*data literacy*) émerge. Porté notamment en France par la FING, ce mouvement part du principe que « les données sont désormais une affaire trop importante pour être laissées entre les mains des spécialistes » et considère les données comme une compétence qui, au même titre que la lecture et l'écriture, peut être accessible à tous.²¹ Rahul Bhargava et Catherine D'Ignazio, respectivement chercheurs au MIT et à Emerson College, ont défini la *data literacy* comme la capacité de « lire, de travailler avec, d'analyser et d'argumenter avec des données. » En 2014, l'Organisation des Nations Unies (ONU) à travers d'un groupe de travail sur la « révolution des données » a publié le rapport « *A world that counts* » qui plaide en faveur d'un investissement mondial pour la *data literacy*. Une des recommandations appelle l'ONU à travailler avec d'autres organisations pour développer un programme mondial d'éducation aux données pour améliorer la compréhension et l'usage des données par le plus grand nombre en particulier par les infomédiaires (journalistes, militants, chercheurs...) mais aussi les personnes en situation de pauvreté. Selon ce rapport, le développement de la *data literacy* est une condition pour que les objectifs de développement durable de l'ONU soient atteints. Les actions conduites actuellement ne semblent pas à la hauteur de ces ambitions fortes et globales {Montes:2019ui}.

Pour développer de manière beaucoup plus ambitieuse la culture des données, ici à l'échelle de la France, je propose trois pistes d'action : placer les données au cœur des politiques publiques, développer une culture critique des données et former massivement aux données.

²¹. « Data Literacy Conference », <http://dataliteracyconference.net/>, consulté le 12 juillet 2016.

Placer les données au cœur des politiques publiques

Dans cette section, je m'inspire d'un des quatre futurs souhaitables proposés par les participants d'Open Data Impact, un des programmes de la FING qui vise à créer la nouvelle feuille de route de l'open data en France, dix ans après son arrivée. Ce futur souhaitable s'intitule « 2025, la data est devenue le meilleur allié des acteurs publics » et prend la forme d'un scénario fictif dans lequel les acteurs publics se sont pleinement emparés du potentiel des données et chaque politique publique bénéficie de l'apport des données, de la conception à la mise en œuvre et l'évaluation²². Je reprends ici quelques pistes proposées dans le cadre d'Open Data Impact.

Généraliser les stratégies data: 'Open Data Impact envisage que l'ensemble des territoires aient l'obligation d'ici 2025 de produire et diffuser leur stratégie en matière de données, à l'image des plans locaux d'urbanisme (PLU) ou des schémas de cohérence territoriale (SCOT). Cette stratégie doit faire l'objet d'une co-construction avec l'ensemble des parties prenantes des territoires. A mon sens, cette idée pourrait d'abord s'appliquer aux métropoles françaises qui ont les moyens de ces ambitions et concentre près de 18 millions de Français. Il faudrait aussi que les stratégies data s'appliquent aussi aux ministères et leurs entités qui disposent de bases de données essentielles à la vie publique.

Co-construire des politiques publiques "nativement" avec les données : chaque territoire devrait expérimenter l'apport des données sur une politique publique de son choix dans laquelle les données sont mobilisées à chaque étape (étude et conception, mise en œuvre, suivi et amélioration permanente, évaluation). Il peut s'agir par exemple de positionner un nouvel équipement public, de faire évoluer l'approvisionnement des cantines scolaires, d'agir sur le non-recours aux aides sociales au niveau départemental, ou encore de réguler par la donnée les opérateurs de micro-mobilité comme les trottinettes électriques. Cet exercice se fait en concertation avec l'ensemble des acteurs concernés dont les habitants. Rajoutons, par rapport au futur souhaitable d'Open Data Impact, que là aussi cette action devrait être conduite aussi au niveau gouvernemental.

²². [Open Data Impact #4] 2025, la data est devenue la meilleure alliée des acteurs publics. Comment ? - Open Data Impact—#TeamOpenData. (s. d.). Consulté 29 juillet 2020, à l'adresse <https://teamopendata.org/t/open-data-impact-4-2025-la-data-est-devenue-la-meilleure-alliee-des-acteurs-publi-cs-comment/1483>

Co-construire les indicateurs des politiques publiques : dans cette proposition, les acteurs publics doivent co-construire avec les acteurs concernés des indicateurs de suivi des principales politiques publiques. Cette proposition pourrait s'appliquer dans un premier temps dans un périmètre donné, par exemple sur les questions sanitaires ou environnementales. Une autre proposition portait sur le fait que chaque territoire doive publier, de manière homogène, des tableaux de bord sur les services publics disponibles au niveau local. Les habitants seraient alors en mesure de participer aux orientations des politiques publiques locales sur une base factuelle et mise à jour. Notons que ces deux propositions n'ont pas fait l'objet de consensus au sein des participants d'Open Data Impact, certains soulignant à juste titre que "tout ne peut pas se mesurer via des données" et que les indicateurs qualitatifs sont trop souvent absents ou réducteurs. Scott {Scott:1998uj} avait montré que les décideurs publics rendent la société lisible en procédant à des simplifications. Les complexités et les particularités locales sont souvent effacées par des indicateurs ou des tableaux de bord qui donnent aux acteurs publics une vision schématique de la société et facilitent le pilotage des politiques publiques. Les ouvrir permettrait aux citoyens d'accéder aux fondements de nombreuses politiques publiques et de contester certains indicateurs dès lors qu'ils sont jugés trop réducteurs. Cela permettra aussi d'identifier des données essentielles au pilotage des politiques publiques qu'il serait bon d'ouvrir afin de pouvoir proposer des manières alternatives de représenter la complexité.

Développer le sourcing des réutilisateurs : cette proposition sort des idées proposées lors d'Open Data Impact. Elle s'inspire des pratiques de l'achat public qui a recours au dispositif du sourcing. Il consiste à solliciter des entreprises sur la faisabilité des prestations demandées et leur adéquation avec les pratiques du secteur. Il sert aussi à inciter un plus grand nombre d'entreprises à candidater. La commande publique connaît souvent un problème d'inadéquation entre offre et demande : trop de marchés n'obtiennent pas assez d'offres. Il en va de même dans l'ouverture des données publiques : les acteurs publics proposent une offre de données sans consulter préalablement des réutilisateurs potentiels sur leurs besoins. Il en résulte que bon nombre de données ne sont jamais réutilisées. Les acteurs publics pourraient, avant d'ouvrir leurs données, avoir recours à une pratique de sourcing en consultant des réutilisateurs potentiels sur leurs besoins et sur les données que pourrait proposer la collectivité. Cela permettrait de concentrer l'effort d'ouverture sur des données demandées par les réutilisateurs et d'adapter le contenu ou la documentation des données aux besoins exprimés par les usagers. Le « sourcing » des réutilisateurs peut aussi prendre la forme d'un signalement aux réutilisateurs potentiels lors de l'ouverture d'un nouveau jeu de données.

Développer une culture critique des données

Le développement de la culture des données ne doit pas échapper à une réflexion critique sur leur production et leur usage. Dans la lignée de *Statactivism* {Bruno:2014tm}, je plaide en faveur d'une stratégie inspirée par le judo, « prolonger le mouvement de l'adversaire afin de détourner sa force. » En utilisant des données et en formant à leur exploitation, je suis convaincu qu'on développe une culture critique des données bien plus forte par la pratique que par l'observation détachée des usages.

Systématiser l'ouverture et l'usage des données dans les concertations : les concertations et les débats publics mettent en scène une confrontation entre un maître d'ouvrage et des citoyens. Ils y confrontent leur opinion et interrogent un projet. Les données sont mobilisées en masse dans les rapports d'expertise qui fondent le projet mis en débat. Or, d'une part, les données brutes qui ont fondé les analyses préalables au projet sont souvent inaccessibles et, d'autre part, les citoyens ne sont ni formés ni encouragés à exploiter les données. Leurs avis sont généralement fondés sur leur expérience vécue ou narrative alors que les experts ont recours à une forme d'objectivité fondée sur les données, généralement reconnue plus légitime et valide {Birchall:2014eo}. Dans ce contexte, la parole des citoyens pourrait sortir renforcée si chacun avait la capacité de saisir des données du débat. Dans un article publié avec David Prothais d'Eclectic Experience {Goeta:2017tn}, nous constatons que la rencontre entre participation et ouverture des données était un « mariage non consommé », une métaphore pour désigner une union qui ne s'est pas produite bien que souhaitée par bon nombre d'acteurs. Nous y avons documenté une expérience qui montre l'intérêt de mobiliser le public d'un débat dans l'ouverture et la réutilisation de données. À l'automne 2016, la commission du débat public sur le projet de troisième ligne de métro de Toulouse a souhaité organiser des ateliers de cartographie afin d'identifier et de localiser les enjeux du projet. Plusieurs ateliers ont été organisés rassemblant jusqu'à 70 personnes. La réalisation de ces cartes a exigé au préalable d'identifier les données disponibles, et de solliciter différents fournisseurs de données afin d'en ouvrir de nouvelles. Cette expérience montre trois apports de la mobilisation des données ouvertes dans les concertations. Premièrement, cela permet d'ouvrir des données en partant des besoins des citoyens dans leur diversité. Deuxièmement, ces ateliers ont été réalisés dans une démarche d'inclusion : les participants ces ateliers n'étaient pas nécessairement des professionnels de la donnée mais plutôt des habitants impliqués dans la vie de leur quartier. Troisièmement, les ateliers ont conduit à adopter un regard critique sur les données et ont remis en cause

l'objectivité voire l'autorité associée aux données. La Commission Nationale du Débat Public (CNDP) devrait étendre ces dispositifs d'ouverture et d'analyse de données dans le cadre des débats qu'elle conduit.

Développer les programmes qui apportent un regard critique sur les données : plusieurs programmes existants peuvent servir d'inspiration pour encourager le public à développer une meilleure compréhension des données et être en capacité de les critiquer. *#DataGueule*, diffusé en ligne depuis 2014 par France 4, condense en quelques minutes des chiffres pour décrypter différents sujets. L'émission, dans sa description, reprend le principe dataactiviste de prolonger le mouvement de l'affaire afin de détourner sa force : « chaque jour, nous sommes bombardés par des milliers de molécules d'information. Des faits, des noms, des chiffres qui s'empilent et se percutent sans que, pourtant, jamais rien ne se crée. Alors pour une fois, plutôt que de rester passifs face à cet assaut, jouons avec.²³ » On peut aussi citer la websérie *Dopamine* diffusée par Arte et coproduite avec le réseau Canopé qui explique de manière ludique les mécanismes de captation de l'attention inscrits sur les principaux réseaux sociaux. Chacun des épisodes prête une attention fine au fonctionnement des algorithmes qui déterminent les interactions sur les réseaux sociaux. Ils mettent en lumière la collecte massive de données indispensables à leur fonctionnement. Ces deux exemples montrent qu'il est possible de traiter de manière à la fois ludique et sérieuse des sujets liés aux données pour s'adresser à une audience large dans une optique de médiation. Il appartient aux groupes audiovisuels, en particulier publics, de donner une plus grande audience à ce type de programmes, au-delà des plateformes en ligne où elles sont généralement diffusées.

Développer la factualité : *Factfulness*, le livre posthume du médecin suédois Hans Rosling nous invite à lutter contre notre vision biaisée du monde par une meilleure compréhension des données. Selon lui, les stéréotypes sur le développement humain viennent en grande partie de biais cognitifs qui perturbent notre compréhension du monde, et donc notre lecture des données qui le décrivent. Il a identifié dans son livre {Rosling:2018ub} dix instincts dramatiques, dix réflexes de pensée qui font barrage à une vision réaliste du monde²⁴ :

- l'instinct du fossé : tracer une ligne droite entre deux parties distinctes de la

²³. <https://www.youtube.com/user/dataqueule>

²⁴. Je m'appuie ici en grande partie sur la présentation de mon collègue Sylvain Lapoix à propos de cet ouvrage : https://dataactivist.coop/infolab_poitiers/facto_sandwich/#1

population aux modes de vie très différents sans prendre en compte tout ce qui se situe entre ces deux points ;

- l'instinct de négativité : montrer que « tout se dégrade » par une glorification ou une mystification du passé ignorant souvent les faits historiques ;
- l'instinct de la ligne droite : considérer qu'une tendance historique comme l'évolution de la population va nécessairement continuer dans la même direction ;
- l'instinct de peur : entretenir l'image d'un monde dangereux en ignorant que nos conditions de sécurité, globalement, n'ont jamais été si bonnes et l'espérance de vie jamais été aussi élevée ;
- l'instinct de taille : mettre en avant le gigantisme d'un nombre sans échelles de comparaison ;
- l'instinct de généralisation : considérer que des conditions valables dans un contexte vont nécessairement se dérouler de la même manière ailleurs ;
- l'instinct de la destinée : assigner à une catégorie de population ou à un pays un destin tracé ;
- l'instinct de perspective unique : user de formules d'autorité pour défendre la légitimité d'un point de vue ;
- l'instinct du blâme : chercher nécessairement un coupable.

En réponse à ces dix instincts, Hans Rosling propose dix règles générales pour développer ce qu'il appelle la « factuelité », une conception du monde basée sur les faits plutôt que les instincts. Cette approche nous aide à mieux appréhender les chiffres, graphiques et visualisations qui pullulent dans le débat public et sont parfois utilisées à des fins de manipulation en mobilisant un des instincts dramatiques. Hans Rosling appelle à la fin de l'ouvrage à enseigner massivement les faits les plus à jour en aidant les élèves à comparer leur situation socio-démographique par rapport au reste du monde ou en montrant l'évolution du développement dans le monde.

Former massivement aux données

Le développement de la culture des données doit aussi passer par des actions massives de sensibilisation et de formation. Voici trois pistes d'action pour mobiliser une large part de la population autour des données.

Développer les Infolabs : En France, la Fondation Internet Nouvelle Génération (FING) a porté le concept d'Infolabs pour créer de nouveaux lieux de médiation aux données inspirés par l'essor des fablabs dans le domaine de la fabrication. Dans ces dispositifs d'innovation ouverte, le public peut prototyper des usages et trouve à sa disposition des ressources, des compétences, de la veille, des rencontres avec des acteurs, des événements de médiation²⁵. Pour l'heure, les Infolabs sont en France des structures temporaires qui organisent des événements réguliers mais ne sont pas engagés dans des démarches de formation massive. Le développement et la pérennisation de ces lieux faciliterait pourtant l'appropriation des données ouvertes par le plus grand nombre. Au lieu de considérer la réutilisation comme une conséquence automatique de l'ouverture des données, ces lieux pourraient accompagner le développement des usages des données en prenant en compte les problématiques des acteurs d'un territoire et en proposant des actions de formation ou de sensibilisation adaptées.

Inviter les données ouvertes dans les médiathèques : les médiathèques proposent une grande multitude de ressources mais les données n'en font pas partie. Pourtant, il n'est pas rare qu'un public large soit à la recherche de données. Par exemple, une personne faisant l'acquisition d'un logement pourrait vouloir explorer les données des demandes de valeurs foncières qui comprennent le prix, la superficie et la localisation des ventes immobilières dans toute la France. Un médiateur pourrait aider la personne à connaître ces données, lui montrer des services existants et, si ces services ne répondent pas à son besoin, lui indiquer comment exploiter ces données. On pourrait imaginer des exemples de médiation par exemple sur l'usage ou la collecte de données sur la qualité de l'air. Les médiathèques en tant que lieu de médiation pourraient aider les projets d'ouverture de données à toucher des publics bien plus larges et divers.

²⁵. *Avons-nous besoin d'Info Labs ?* | *InternetActu.net*. (s. d.). Consulté 29 juillet 2020, à l'adresse <http://www.internetactu.net/2012/05/15/avons-nous-besoin-dinfo-labs/>

Intégrer les données dans les programmes scolaires : le code informatique, la cuisine, la botanique, l'astronomie, la plomberie... Les demandes d'intégrer telle ou telle pratique ou discipline dans les programmes scolaires sont adressées très régulièrement à l'Éducation nationale. Si on écoutait tous les entrepreneurs de cause qui plaident pour compléter les programmes, les élèves pourraient ne jamais quitter l'école. Alors, plutôt qu'une autre pratique, pourquoi intégrer les données ouvertes dans les programmes scolaires ? En plus des enjeux propres aux données, je vois deux atouts à l'usage de données en classe : le caractère interdisciplinaire de cette pratique et la granularité des données qui donne l'occasion aux élèves de travailler sur leur territoire vécu. L'exemple des data sprints pédagogiques développé par le réseau Canopé l'illustre bien. Le data sprint Traces de Soldats²⁶ aide les élèves à partir des traces inscrites sur les monuments commémoratifs et mises en données pour comprendre les parcours des Poilus de la Première Guerre mondiale. Cet exercice fait appel aux savoir-faire de toutes les disciplines : histoire, géographie, mathématiques, technologie, arts plastiques, lettres... Les élèves sont invités à représenter les ravages de la guerre sur la forme de visualisation de données qui sont ensuite publiées sur [data.gouv.fr](https://www.data.gouv.fr)²⁷. Plus de soixante infographies y sont publiées et les résultats sont dans l'ensemble remarquables. L'expérience du data sprint a été renouvelée sur la thématique de l'accidentologie routière à vélo avec Traces 2 roues. En mobilisant des données, les élèves ont pu développer des compétences au-delà des disciplines et ont analysé des phénomènes au plus proche de leur lieu de vie, réinscrivant l'enseignement dans leur territoire vécu et décroissant les savoirs.

Conclusion

Tout au long de ce chapitre, j'ai cherché à identifier les chantiers et explorer les pistes d'action afin que l'ouverture des données réduise les asymétries d'information. J'ai identifié trois chantiers prioritaires à cette fin : obtenir les données qui manquent, améliorer la qualité et la fiabilité des données ouvertes, développer la culture des données. Pour chacun de ces chantiers, j'ai exploré des pistes d'action qui pourraient permettre de résoudre les problèmes identifiés.

²⁶. *Traces de Soldats: Le datasprint historique*—Réseau Canopé. (s. d.). Consulté 30 juillet 2020, à l'adresse <https://www.reseau-canope.fr/service/traces-de-soldats-le-datasprint-historique.html>

²⁷. *Première Guerre mondiale—Les Poilus morts pour la France (à compléter)*—Data.gouv.fr. (s. d.). Consulté 30 juillet 2020, à l'adresse <https://www.data.gouv.fr/fr/datasets/premiere-guerre-mondiale-les-poilus-morts-pour-la-france-a-compléter/>

On pourrait m'accuser de naïveté ou d'excès d'optimisme si je prétendais que ces quelques actions suffiront à elles seules à réduire les asymétries d'information et à créer un monde plus égalitaire. Je reste conscient que trop souvent les données servent à contrôler, réprimer voire à manipuler. Les scandales comme l'affaire Cambridge Analytica reviennent à intervalles réguliers et érodent petit à petit la confiance du public dans le traitement des données. Scandale après scandale, on pourrait rejeter en bloc les données, les considérer comme un instrument d'aliénation. Cette critique s'est exprimée de manière singulière dans l'ouvrage *Contre l'Alternumerisme* {Lainae:2020uk} qui dénonce la naïveté des partisans de l'ouverture des données quant au caractère oppresseur des données :

« ce que nous critiquons, ce n'est pas le caractère public ou privé d'une donnée, c'est la pertinence de la « prise de données ». Résistant au puçage des troupeaux, aux technologies RFID, à la numérisation de l'école et du travail, à la mesure et au stockage de toute chose, nous défendons la thèse que la prise de données procède d'une réduction du réel et d'une dépossession. »

Cette position s'inscrit dans le registre de la critique radicale qui considère « la quasi-impossibilité de modifier certains éléments – au moins des éléments importants – sans changer, d'un coup, le tout » {Boltanski:2014tw}. Cette conception du monde se distingue de la position réformiste qui, toujours selon Boltanski, « se fonde sur la croyance selon laquelle les éléments qui composent la réalité sont suffisamment indépendants les uns des autres pour que certains d'entre eux puissent être progressivement améliorés sans que les contours de la réalité ne soient radicalement transformés, au moins d'un seul coup et en bloc. » Les convictions personnelles de chacun orienteront vers une position ou une autre. Je considère que la critique réformiste a son utilité pour remettre en cause les excès de la collecte généralisée de données, ce que Kenneth Cukier {Cukier:2013tu} a appelé la « mise en données du monde » ou *datafication* pour désigner « la capacité à paramétrer des aspects du monde et de la vie humaine qui n'avaient encore jamais été quantifiés » et à « numériser non plus des documents, mais tous les aspects de la vie. » A l'heure où l'Internet des objets quantifie les aspects les plus intimes de nos vies et où chaque instant de nos vies laissent des traces sous forme de données, je pense que ce registre de critiques peut susciter une prise de conscience sur l'utilité relative et les excès causés par ces dispositifs de collecte de données permanents et ubiquitaires.

Mais ce renoncement risque de renforcer la situation que ces mêmes acteurs critiquent. La situation présente une grande proximité avec celle décrite par les

auteurs de *Statactivism* {Bruno:2014tm} confrontés aux accusations de connivence des statistiques avec le pouvoir et la sanction. Comme pour les statistiques, les usages émancipateurs des données sont nombreux et organisent des mobilisations d'ampleur. *Statactivism* montre que rejeter en bloc la statistique ne peut conduire qu'à « laisser le monopole de ces instruments aux puissants. » Pour paraphraser l'ouvrage, il n'y a pas de raison que les données se trouvent toujours du côté de l'Etat ou du capital. Pour cela, le dernier chantier qui consiste à développer massivement la culture des données me paraît essentiel. Penser l'impact de l'ouverture des données sans développer les compétences pour les traiter crée un pouvoir d'agir limité à une élite.