

## Big data environnemental (notice de dictionnaire)

Comment citer ce texte :

Pierre Gautreau, 2020. Big data environnemental, in *Dictionnaire Critique de l'Anthropocène (collectif)*, Presses du CNRS, Paris.

- - -

La massification des données environnementales avec l'entrée dans le XXI<sup>e</sup> siècle est une dimension centrale des débats anthropocéniques. Cette massification est indissociable de la promesse qu'il va être prochainement possible de reproduire informatiquement (modéliser) le fonctionnement de l'enveloppe externe de la planète, de le prévoir, et simultanément de résoudre les problèmes qui l'affectent. En somme, de piloter le "vaisseau-terre" (Boulton 2018). Penser le "big data" environnemental est donc, d'abord, une façon de comprendre comment les interactions entre sciences, techniques et géopolitique ont construit depuis les années 1980 l'échelle globale comme "le" niveau pertinent d'analyse des relations des humains à la Terre. D'un point de vue anthropologique ensuite, le big data est indissociable des débats sur la croissante médiatisation technique du lien des humains à la nature, la relation à celle-ci passant de plus en plus souvent par des indicateurs agrégeant de grandes tendances planétaires -déclin de la biodiversité, température globale moyenne- et par des interfaces numériques. Cette médiatisation est-elle un mal nécessaire, marginalisant les savoirs vernaculaires mais permettant d'atteindre une vision commune des "problèmes" de la planète, seule issue face à l'inaction d'acteurs internationaux qui ne veulent pas se coordonner ? Ou, au contraire, est-elle foncièrement dangereuse, antidémocratique et renforçant une relation principalement utilitaire à la nature, perçue comme pourvoyeuse de ressources ou génératrice de risques, ce qui minerait la propension innée des humains à protéger leur environnement ?

Il n'existe pas de consensus fort sur ce que le terme recouvre, et encore moins sur les conditions qui peuvent le rendre "environnemental". Le big data, c'est d'abord un processus d'accumulation de quantités très importantes de données numériques, que peu d'institutions ont actuellement les capacités de traiter à court terme, voir même de stocker (critère de "volume"). Ensuite, c'est un processus rapide, la majorité de ces données étant captées de façon automatisée, et arrivant sur les disques durs en "temps réel", quelques milli-secondes après le processus qu'elles ont enregistré (critère de "vitesse"). Enfin, ce sont des données "non-structurées", ce qui signifie qu'elles doivent subir un traitement plus ou moins long pour être insérées dans des "bases de données", afin d'être interrogées par des humains (critère de "variété"). Le big data a donc des caractéristiques à rebours des données environnementales conventionnelles, fruit de patientes observations humaines consignées après de longues vérifications. La première difficulté pour appréhender ce qu'il recouvre est de trouver des critères permettant de le distinguer de la tendance multi-séculaire de la science européenne, puis internationale, à accumuler des données biophysiques dans de "grosses" bases. Ainsi, on peut dire que les 1.082.865.106 données d'observation de faune et de flore disponibles sur la plateforme du GBIF (Global Biodiversity Information Facility) en février 2019 ne sont pas réellement du big data. Chacune de ces données a été localisée, identifiée, puis classée, des êtres humains intervenant à chacune de ces étapes: finement

manufacturée, cette production de données reste encore très conventionnelle, et ce même si les sciences participatives ont permis d'augmenter sensiblement le nombre d'observateurs grâce à l'enrôlement des naturalistes amateurs. Le terme de "big data" semble bien mieux adapté à l'ensemble des captations massives et automatisées de données environnementales pour l'étude du climat (température, précipitations...), des flux hydrologiques ou des états de surface du sol (imagerie satellitaire): ainsi, des millions de registres météorologiques sont stockés heure par heure, tandis que la Terre est photographiée plusieurs fois par jour dans son intégralité par des constellations de satellites. Historiquement, ce big data est essentiellement "actif", produit par des capteurs paramétrés par des scientifiques et des institutions publiques, avec un rôle prépondérant des militaires. A partir des années 2000, s'intègrent à cet univers de données les captations "passives", principalement celles issues de notre activité sur internet (comportements d'achats, réseaux sociaux) et de notre usage d'appareils connectés (smartphones). Avec l'importance croissante accordée aux risques, toute donnée personnelle captée de cette façon devient potentiellement "environnementale", lorsque par exemple elle permet à un acteur public d'évaluer en temps réel la population exposée à une catastrophe.

Si la massivité de ces données subjugué nombre d'observateurs, il ne faut donc pas oublier qu'elles ne recouvrent que certaines dimensions du réel ou portions de l'espace terrestre, et sont fondamentalement segmentaires. La biodiversité n'est que très mal incorporée au big data environnemental, à la différence des états météorologiques, qui peuvent être plus facilement résumés à des variables capturables de façon massive (température...). Les milieux urbains et leurs populations sont largement plus couverts par le big data passif que les milieux ruraux où les usagers de smartphones sont plus rares, ce qui implique des inégalités en termes de prévention des risques. Une part croissante de ces données est aux mains d'acteurs privés, qui le transforment activement en produits commercialisables, à destination de l'agriculture ou des assurances, généralement dans un but de prévention des risques météorologiques. Plutôt que d'établir des limites strictes entre ce qui est ou n'est pas du big data environnemental, il est sans doute plus intéressant aujourd'hui de scruter les trajectoires d'incorporation de certaines dimensions du réel au champ du "big". A titre d'exemple, si la biodiversité résiste encore largement à ce qu'on pourrait appeler une "bigdatification", certains obstacles semblent en passe de tomber: en 2019, de nombreux projets scientifiques testent des méthodes afin de coupler imagerie satellitaire et relevés "d'ADN environnemental" pour tenter de cartographier de façon plus fréquente l'état de la biodiversité sur de vastes territoires. Les méthodes d'ADN environnemental doivent permettre à court terme d'estimer la diversité en espèces et en gènes de portions croissantes de l'espace terrestre, grâce au séquençage en série d'échantillons d'eau ou de sol. La réduction exponentielle du coût de ces relevés permet d'envisager à faible échéance une explosion du volume de données de biodiversité.

La période actuelle est dominée par un cadrage des questions environnementales dans lequel la planète est pensée d'abord comme une réalité "globale". Cette pensée est indissociable de l'émergence d'une science qui, au cours de la seconde moitié du XXe siècle, va penser la Terre comme un système, et activement favoriser les instruments qui vont lui fournir cette vision globale, ce dont le big data environnemental est le dernier avatar. Il faut chercher les origines de ce mouvement dans les mega-projets scientifiques qui naissent aux USA durant la seconde guerre mondiale (projet Manhattan), puis se ramifient notamment dans les sciences géophysiques, avec l'International Geosphere-Biosphere

Programme lancé entre 1983 et 1990 (Kwa 2005), puis avec le "Megascience Forum", qui dans les années 1990 développe au sein de l'OCDE l'idée que l'informatisation et la production en masse de données environnementales va permettre des avancées scientifiques décisives. De ce mouvement sont issues le *Groupe d'experts intergouvernemental sur l'évolution du climat* (GIEC), ou la *Plateforme Internationale pour la Biodiversité et les Services Ecosystémiques* (IPBES), deux institutions où l'on promeut une science guidée par les données (*data-driven*) plutôt qu'une science fondée sur l'expérimentation, comme c'était notamment le cas en écologie (Devictor et Bensaude-Vincent 2016). Les années 1980-1990 sont donc celles d'une co-production intriquée entre science "globale" et problèmes environnementaux, la première justifiant le cadrage des seconds comme une question globale, tandis que l'internationalisation des seconds légitimait le développement de bases de données de taille planétaire (Turnhout, Dewulf, et Hulme 2016).

C'est d'abord ce que fait le big data à la science elle-même, son impact épistémologique, qui crée la controverse. Certains déplorent qu'en écologie, la priorité accordée aux efforts de constitution de grosses bases de données (en termes de budgets et de priorités de recherche) fasse perdre de vue l'importance de l'étude des interactions, notamment locales, et qu'elle ait transformé la discipline en "une technoscience orientée vers la gestion de la biodiversité [...] bloquée dans le processus sans fin d'accumulation de données massives" (Devictor et Bensaude-Vincent 2016). Car au centre de la constitution du big data environnemental, il y a la nécessité, technique, de simplifier au maximum les paramètres enregistrés, de façon à rendre toute donnée collectée comparable avec toute autre collectée ailleurs sur la planète. Ce processus entraîne une "décontextualisation" de ces données massives. Outre en tirer des bilans généraux, elles perdent l'essentiel de leur pertinence écologique, de *compréhension* des processus (Devictor et Bensaude-Vincent 2016). En matière climatique, un processus similaire s'observe, les études sur la différenciation géographique des climats du globe tendant à être reléguées au second plan derrière celles d'un climat globalisé (Miller 2004). Les conséquences politiques immédiates de ce vaste processus sont liées au fait qu'il diffuse l'idée fautive que le big data est un champ neutre, qui produit un savoir objectif (Boyd 2010), disponible pour l'action mais sans être normatif. Ce faisant, il légitime insensiblement la nécessité d'une gouvernance globale de l'environnement, et les institutions construites dans cet objectif (Turnhout, Dewulf, et Hulme 2016).

En termes politiques, les répercussions de cette tendance à la bigdatafication du monde sont donc multiples. Dans le champ de la biodiversité, la prééminence croissante des approches quantifiées du changement environnemental a favorisé des acteurs en capacité de mettre en scène leur maniement des grandes bases de données. C'est notamment le cas du secteur international de la conservation où, dans une optique néolibérale, de grands ONG revendiquent un rôle privilégié dans la gouvernance de la biodiversité mondiale. Elles démontrent pour ce faire leur capacité à orienter les financements pro-conservation sur les territoires où ceux-ci seront le plus "optimisés" en termes de rapport dollar investi/biodiversité préservée (Milian et Rodary 2010; Dumoulin Kervran et Rodary 2005). Plus généralement, cette centralité du big data dans la gouvernance mondiale de l'environnement peut être considérée comme un pas supplémentaire dans l'occultation des savoirs vernaculaires ou "locaux". Malgré des tentatives menées par exemple par l'IPBES pour leur donner une place dans l'espace numérique des bases mondiales, celles-ci semblent illusoire et en contradiction avec la nature même

de ces données vernaculaires, sensibles, hétérogènes et contextualisées. Les implications en termes de justice sont donc profondes, le big data tendant dans le même mouvement à marginaliser les savoirs "autres" des bases elles-mêmes... et à leur ôter progressivement toute légitimité politique, du fait même de leur absence de ces bases. Alors même que le rôle du big data pour la mobilisation mondiale contre les changements environnementaux semble nécessaire, le coût à payer en termes de perte de vue de la complexité des processus écologiques (Devictor et Bensaude-Vincent 2016), et en termes démocratiques, semble déjà bien lourd.

Pierre Gautreau

BOULTON Geoffrey. 2018. « The Challenges of a Big Data Earth ». *Big Earth Data* 2 (1): 1–7.

BOYD William, « Ways of Seeing in Environmental Law: How Deforestation Became an Object of Climate Governance », *Ecology Law Quarterly*, 37, 2010, p. 843–916.

DEVICTOR Vincent et Bernadette BENSAUDE-VINCENT, « From Ecological Records to Big Data: The Invention of Global Biodiversity », *History and Philosophy of the Life Sciences*, 38 (4), 2016, p. 23.

DUMOULIN KERVRAN David, et Estienne RODARY, « Les ONG, au centre du secteur mondial de la conservation de la biodiversité », In *Représenter la nature? ONG et biodiversité*, édité par Catherine Aubertin, IRD Editions, 2005, p. 59-97.

KWA Chunglin, « Local Ecologies and Global Science: Discourses and Strategies of the International Geosphere-Biosphere Programme », *Social Studies of Science*, 35 (6), 2005, p. 923–50.

MILIAN JOHAN, et Estienne RODARY, « La conservation de la biodiversité par les outils de priorisation: Entre souci d'efficacité écologique et marchandisation », *Revue Tiers Monde*, 202 (2), 2010, p. 33.

MILLER Clarck, " Climate Science and the Making of a Global Political Order », In *States of Knowledge: The Co-Production of Science and the Social Order*, édité par Sheila Jasanoff, Routledge, 2004, 46–66.

TURNHOUT Esther, Art DEWULF, et Mike HULME, « What Does Policy-Relevant Global Environmental Knowledge Do? The Cases of Climate and Biodiversity », *Current Opinion in Environmental Sustainability*, 18 (février) 2016, p. 65–72.