

INTRODUCTION

*« Non, l'avenir n'offre que peu d'espoir à ceux qui attendent
de nos esclaves mécaniques l'avènement d'un monde
où l'on pourra se passer de penser. »*

Norbert Wiener, *God & Golem Inc.*, 1964, rééd. 2000, p. 89

Au cours des deux dernières décennies, notre production de données numériques a connu une augmentation sans précédent. Celle-ci ne semble pas devoir se tarir actuellement, car avec l'émergence de l'Internet des objets, les capteurs se sont immiscés dans tous les lieux de notre quotidien. De fait, cette croissance évolue au rythme de trois lois, qui restent encore en toile de fond des analyses : la loi de Moore, qui prévoit un doublement de la densité d'inscription sur puce de silicium tous les 18 mois, la loi de Kryder, qui suppose un doublement de la densité de stockage sur disque magnétique tous les 13 mois ; et la loi de Nielsen, qui annonce le doublement de la capacité des réseaux publics tous les 21 ans. Ces tendances devraient donc inévitablement déboucher sur de nouvelles avancées.

Pourtant, le développement continu de nos capacités de traitement et de stockage est à peine suffisant pour faire

face à l'explosion du nombre de données : en 2015, ce sont à peu près 10^{22} octets nouveaux qui ont été générés et ce nombre devrait être multiplié par 5 d'ici à 2020. Le vocabulaire lui-même doit s'adapter pour décrire ces nouveaux volumes. Nous sommes passés au cours de la dernière décennie du temps des exaoctets (10^{18} octets) à celui des zettaoctets (10^{21} octets), voire à celui des yottaoctets, 10^{24} octets... En 2018, la quantité totale de données sur la planète était estimée à 33 Zo, et ce chiffre devrait grimper à 175 Zo d'ici à 2025. Dans ce contexte, de nouveaux outils ont été créés, défiant les capacités de l'entendement humain pour manipuler et traiter ces quantités disparates de données (Big Data, *machine learning*). Et ces nouvelles technologies ont fait l'objet d'un grand engouement médiatique et industriel car les promesses qui leur sont associées semblent potentiellement révolutionnaires. L'hypothèse formulée serait que le traitement de données massives en temps réel permettrait d'améliorer considérablement nos capacités à connaître notre environnement, à y interagir, ainsi qu'à prédire des événements et des comportements à court et moyen termes. Aussi, l'application de ces outils devrait conduire à d'importants progrès, dans des secteurs aussi divers que la santé, l'agriculture, la recherche sur le génome humain, le renseignement géolocalisé, la gestion des villes, celle des maisons individuelles, etc.

En parallèle de ces nouveaux champs de recherche émerge un ensemble de discours annonçant l'obsolescence programmée des sciences humaines avec l'avènement de cette ère des données. En 2008, le rédacteur en chef de la célèbre revue *Wired*, Chris Anderson, présageait ainsi la « fin de la théorie » et des méthodes scientifiques traditionnelles, qui visent à expliquer des phénomènes sociaux

et à les interpréter (Anderson, 2008). Selon lui, « avec suffisamment de données, les chiffres parlent d'eux-mêmes ». La corrélation de données à grande échelle et en temps réel nous permettrait d'observer, de détecter et d'anticiper des événements et des faits auparavant invisibles, ou à peine perceptibles, sans que le recours aux sciences sociales ne soit nécessaire. L'analyse géographique et géopolitique serait donc en passe de disparaître, au même titre qu'un grand nombre de disciplines, détrônées par les capteurs et les algorithmes.

Il pourrait donc sembler plutôt hasardeux de proposer ici l'esquisse d'une géopolitique des données, qui s'intéresserait autant à leur source et aux pouvoirs qu'elles procurent qu'aux conséquences politiques et spatiales que leur usage généralisé a engendrées. Et cette proposition pourrait paraître encore plus périlleuse qu'il s'agirait d'amorcer, par le biais des bonnes vieilles sciences humaines, une approche critique permettant de cerner les limites de telles technologies, et de montrer en quoi le monde n'est pas aussi déterminé que les données et les statistiques semblent nous le dire. C'est pourtant notre objectif. Pour ce faire, nous nous baserons sur une sélection d'études et de recherches réalisées dernièrement sur cette question, tant dans le champ de la géopolitique, que dans ceux de la géographie et des sciences sociales. Mais autant le dire d'emblée, le texte qui suit n'entend nullement dresser un état de l'art exhaustif sur ce sujet. Il s'agit plutôt ici de fournir un ensemble de pistes sur ce que pourrait être une géopolitique des données aujourd'hui. L'idée est de dépeindre un panorama cohérent de ce vaste domaine, autour de ses dimensions politiques et spatiales, qui puisse être, demain, le socle pour de nouvelles recherches plus approfondies.

Comment penser l'espace numérique comme un espace (géo)politique ?

Concevoir l'espace numérique à travers ce prisme géographique et géopolitique peut, à première vue, sembler délicat. Le numérique ne fait-il pas immédiatement écho dans les représentations populaires à un espace intangible, un lieu hors du monde, impossible à localiser et à cartographier, dont l'une des caractéristiques premières serait sa dimension ubiquitaire, car il paraît accessible partout sur la planète, à partir de l'ensemble des lieux disposant d'un point d'entrée Internet. Mais cela n'a pas empêché les géographes de s'intéresser très tôt à cette question. En France, des chercheurs spécialisés dans la géographie des communications, comme Henry Bakis, dans la géographie des réseaux, comme Jacques Lévy ou Gabriel Dupuy, se sont intéressés à cette question dès les années 1990. À la même époque, les travaux sur le sujet en langue anglaise étaient déjà importants. Ces relations précoces entre géographie et numérique ont fait récemment l'objet d'un article de synthèse publié dans la prestigieuse revue *Progress in Human Geography* par trois spécialistes, James Ash, Rob Kitchin et Agnieszka Leszczynski (Ash, Kitchin et Leszczynski, 2016). Selon eux, la discipline géographique s'est confrontée au numérique sous trois formes distinctes, qui ont chacune leur propre chronologie, qu'ils regroupent dans les catégories : géographies produites *via* le numérique, géographies *du* numérique et géographies produites *par* le numérique.

Historiquement, les premières géographies à être apparues sont les géographies produites *via* le numérique. Ce sont celles qui, dès les années 1960, avec les premiers ordinateurs, vont s'intéresser à la manière dont les outils numériques pourraient permettre un traitement automatisé de

données géolocalisées. Cette pratique va être à l'origine des Systèmes d'information géographique (SIG). Ce champ va peu à peu se spécialiser pour devenir un domaine très technique de la discipline (géomatique), régulièrement utilisé dans diverses administrations ou entreprises, pour fournir des éléments stratégiques d'analyse spatiale d'activités. Un courant critique s'est néanmoins développé en son sein, principalement dans la littérature anglophone, qui se détache de l'instrumentalisation institutionnelle pour s'intéresser à des objets plus éloignés du pouvoir – les minorités, les femmes, le genre, entre autres.

À partir de la fin des années 1990 apparaissent les géographies *du* numérique dont le principal objectif est l'application « d'idées et de méthodologies géographiques préexistantes pour étudier ce qui est considéré comme un nouveau domaine matériel, spatial et technique de communication et d'interaction » (*ibid.*, p.4). Dans cette catégorie émergent alors des géographies d'Internet, des mondes virtuels ou des jeux vidéo, du cyberspace, etc. Ces travaux s'intéressent en particulier à la dimension socio-technique de ces phénomènes, montrant en quoi ce nouvel environnement numérique est une production humaine, dont les enjeux ne sont pas strictement techniques, mais aussi sociaux, politiques, économiques, etc.

Enfin, les géographies produites *par* le numérique, plus récentes, ont pris comme objet d'étude la manière dont les technologies ont modifié la production de nos espaces du quotidien, en les régulant ou en les « augmentant », et comment elles ont transformé, ce faisant, les relations socio-spatiales (*ibid.*, p.7). Il est en effet très intéressant de voir comment les outils numériques se sont, au fil des ans, progressivement infiltrés dans notre environnement familier, et ont fini par apparaître presque banals, alors

même qu'ils transforment en profondeur les lieux dans lesquels ils s'inscrivent. Aujourd'hui, la fonction de certains espaces est ainsi en partie générée par les codes et les algorithmes qui les gèrent. L'aéroport, le supermarché, les routes intra-urbaines sont des exemples parmi d'autres d'espaces en grande partie dépendants de codes informatiques, que ce soit pour gérer les données concernant les flux de voyageurs et de bagages, associer un code-barre à un prix dans les paiements en caisse, ou commander les feux de circulation, entre autres.

Toutefois dans l'ensemble de ces catégories, la question du pouvoir demeure un aspect central de la production scientifique qui résulte de l'articulation entre géographie et numérique. Des analyses géopolitiques se sont ainsi développées de manière transverse dans ces différents champs disciplinaires. En France, Frédéric Douzet a été l'une des premières à proposer une lecture pertinente dans ce domaine en publiant sur ce sujet dès 1996 dans la revue *Netcom*. Plus récemment, elle a présenté, avec l'équipe de l'Institut français de géopolitique (IFG), une synthèse des différents enjeux de l'approche géopolitique du cyberspace dans un numéro de la revue *Hérodote* consacré à cette question (Douzet, 2014). Des individus (respect de la vie privée, commerce des données personnelles, pirates informatiques, cybersécurité), aux États (rivalités de pouvoir, questions de la souveraineté, des cyberattaques et des adaptations doctrinales qu'elles provoquent, ou des nouveaux rapports de force induits par l'architecture physique et logique des réseaux), en passant par la gouvernance mondiale (poids croissant des nouveaux géants économiques du numérique, gouvernance d'Internet, luttes de valeurs entre États « démocratiques » et « autoritaires », implication d'acteurs non étatiques), le prisme d'étude est

gigantesque. Sur chacun de ces points apparaissent de nouveaux thèmes d'affrontements et de tensions entre acteurs locaux, nationaux et internationaux.

Dans le monde anglophone s'est également développée une géographie critique du numérique en partie en lien avec cette approche géopolitique. L'un de ses principaux objectifs est d'analyser la manière dont le développement du numérique a accentué les inégalités sociales entre ceux qui possèdent, stockent, traitent les données et le reste du monde. De fait, les utilisateurs sont souvent candides et impuissants, face à des usages et des stratégies, dont ils n'ont, pour la plupart, pas connaissance. Ces chercheurs se sont aussi penchés sur les nouveaux modes de surveillance rendus possibles par ces technologies, ou sur les nouvelles formes de légitimation technique du pouvoir, qui par le biais de l'analyse algorithmique justifient des politiques publiques au nom d'une expertise statistique.

Une conception opérationnelle : la théorie des couches du cyberspace

Parmi ces différents champs de recherche, on peut néanmoins mentionner une conception plus opérationnelle du cyberspace, qui est largement utilisée aujourd'hui, en particulier dans la réflexion stratégique. Utilisant une métaphore géologique, cette approche considère l'espace numérique comme constitué de plusieurs couches, dont le nombre varie en fonction des auteurs. La proposition la plus simple reste celle qui propose d'en distinguer trois, à savoir :

- une couche matérielle, qui est la couche des infrastructures avec tous les périphériques d'accès, les sites, les machines, les dispositifs tangibles nécessaires au fonctionnement du réseau. C'est le domaine physique d'Internet,

constitué d'ordinateurs, de serveurs, de câbles sous-marins, de *datacenters*, etc.

- une couche logicielle, qui, par l'intermédiaire du code, et des différentes strates de langages machine et de protocoles, permet aux ordinateurs de communiquer les uns avec les autres, et d'échanger entre eux d'importants volumes de données en des temps minimes. Cette couche pourrait être elle-même décomposée en deux sous-couches, avec d'un côté les systèmes d'exploitation (Windows, Linux, macOS, etc.) qui sont le socle de fonctionnement des ordinateurs, et de l'autre, les applications, et les programmes plus accessibles et conviviaux, qui permettent aux consommateurs d'utiliser leur machine, sans avoir besoin de connaissances informatiques spécifiques. Cette couche logique est la cible des attaques les plus fréquentes (virus, hacking, cheval de Troie, etc.).

- enfin, une couche dite sémantique (ou cognitive), à savoir tout le contenu informationnel d'Internet. En clair, il s'agit de l'ensemble des messages qui passent entre utilisateurs par Internet. C'est le lieu des interactions sociales, des échanges d'informations, mais aussi des jeux d'influence et de propagande pour différents acteurs politiques.

Ainsi, cette conceptualisation théorique du cyberspace a le mérite de rendre possible des analyses géopolitiques plus concrètes d'Internet. Apparaît derrière ces couches superposées une matérialisation géographique du pouvoir dans le cyberspace. La domination *de facto* des États-Unis, du fait de leur prédominance économique, politique et technique dans ce secteur est par exemple rendue visible, à tous les niveaux, que ce soit pour ce qui est des infrastructures physiques, des avancées techniques, de la recherche technologique, du poids économique, mais

aussi et surtout de l'influence réglementaire. Néanmoins, les investissements russes et chinois de ces dernières années semblent présager d'un paysage numérique mondial en mutation, avec des pôles régionaux émergents, concurrents du bloc américain.

Problématiser la géopolitique des données

Sur la base de ce premier tour d'horizon, quelques questions se posent. Que penser de cette hypothèse émergente d'un monde déterminé par les données, qui serait alors en partie prévisible ? Notre conviction est qu'il n'existe pas de déterminisme technologique. Les données ne sont pas des éléments naturels que l'on devrait subir (un déluge de données, être noyé dans les données, etc.), mais sont des produits humains, issus de choix (ou d'absence de choix) politiques, économiques, sociaux, idéologiques, philosophiques, éthiques. Les données sont donc le fruit d'une construction sociale, et le sens qu'on leur attribue dépend d'un traitement et d'une interprétation qui reflètent en partie un environnement socio-spatial (l'usage des données répond à des besoins et des stratégies sociales) et géopolitique (la capacité de traitement nécessite l'accès à des infrastructures techniques de plus en plus lourdes et complexes, et des moyens financiers, qui deviennent des enjeux stratégiques et politiques sur la scène internationale).

En ce qui concerne la géopolitique des données, il s'agit donc dans cet essai d'écarter, autant que possible, tous les discours déterministes, qu'ils soient géographiques, technologiques ou idéologiques et de débusquer les choix politiques et stratégiques, ainsi que les représentations, que sous-tendent toutes les constructions humaines, techniques et territoriales. En clair, nous proposons ici d'étudier les données au travers de ce que les acteurs en

font. Pourquoi les produisent-ils ? Quels en sont leurs usages ? Et en quoi ces diverses utilisations génèrent-elles, confortent-elles ou fragilisent-elles des relations de pouvoir et d'influence entre les acteurs ? Nous devons également interroger l'architecture du pouvoir, produite par l'organisation et la hiérarchisation des espaces physiques et numériques, et par les concepts d'analyse utilisés pour le faire. Le terme même de « données », et ce qu'il véhicule, fera l'objet de cet examen réflexif.

Or, penser la donnée comme un produit socio-spatial amène à plusieurs réflexions qui structurent cet essai. Tout d'abord, les données peuvent être de différente nature suivant qu'elles ont été générées par un capteur automatique, par un instrument de navigation, par un individu ou un groupe d'individus, etc. – et chacun de ces différents types de données répond à divers types de besoins, appelle divers usages qui profitent à différents acteurs. De fait, les données ne sont pas neutres et indépendantes des groupes sociaux qui les ont produites – elles répondent à des finalités, qui révèlent des stratégies et des représentations d'acteurs.

Mais, en retour, ces produits numériques transforment la géographie et la géopolitique, leurs pratiques et leurs objets. Le contrôle des données (leur production, leur collecte, leur stockage, leur sécurisation, leurs flux) est devenu un enjeu géopolitique majeur pour les États comme pour les individus (comme le démontrent régulièrement les lanceurs d'alerte). Ce contrôle révèle un paysage géopolitique ancien (domination américaine, volonté de souveraineté étatique sur les données), mais aussi une géopolitique mondiale renouvelée (nouveaux acteurs – GAFAM, acronyme désignant les principales firmes numériques américaines Google, Apple, Facebook, Amazon et

Microsoft, etc.). Cette concurrence autour des données peut entraîner des rivalités, des conflits, voire aboutir à ce que certains appellent une « cyberguerre ».

Et, de fait, l'usage des données produit de nouveaux modes d'organisation, de nouvelles pratiques sociales, une nouvelle économie et modifie les rapports de force et les jeux de pouvoir dans la société. L'usage des données produit de nouveaux espaces et de nouvelles manières d'appréhender les territoires et les populations, pour les scientifiques comme pour les gouvernements. Or, sous l'aspect technique de ces nouveaux modes de gestion automatisée se cache une dimension politique que nous devons apprendre à déceler.

Ces phénomènes que l'on désigne comme « données », ceux-là mêmes qui nous semblent si évidents et si clairs que l'on ne s'interroge que rarement sur ce qu'ils sont, doivent donc faire l'objet d'une réflexion préalable. Et pour comprendre ce qui se joue derrière ce que l'on croit tous connaître, il est nécessaire d'entrer dans le détail de ce qui construit les « données », tant sur le plan technique que social, et de prendre le temps d'étayer, quitte à paraître quelquefois un peu pointu, car cet effort éclaire quelques-uns des principaux enjeux politiques de notre époque.