Exemple de fiche de lecture

CATTARUZZA, Amaël. Géopolitique des données numériques : pouvoir et conflits à l'heure du Big Data. Le Cavalier Bleu, 2019. 174 p. ISBN 979-10-318-0348-7

Date et auteur de la fiche de lecture : février 2020 par la cartothèque de Paris 8
Localisation du document : BU de Paris 8, salle violette, cote 911.332 CAT

L'auteur est géographe, maître de conférence à l'ESM Saint-Cyr, chercheur au centre de recherche GEODE.

Mots-clefs :
Monde, années 2000-2020, données numériques, territoire physique et virtuel, gouvernance, cyberguerre, frontières, surveillance, géopolitique, méthodologie.

Résumé
La production de données numériques a modifié en profondeur le champ d'étude de la géopolitique. Le territoire numérique est devenu un lieu de conflit entre acteurs publics et privés. La science géopolitique doit se doter de nouveaux moyens d'analyse.

***

Introduction

Depuis 2000, explosion de la production de données numériques (DN) et des nouvelles technologies permettant leur traitement. Grand engouement du public : promesses de progrès dans tous les domaines. Les sciences humaines semblent dépassées, leurs méthodes incapables d'analyser et d'interpréter les phénomènes sociaux avec l'efficacité du *Big Data*. L'auteur cherche à prouver que la géographie et la géopolitique sont toujours valides et propose des pistes pour élaborer une géopolitique des DN.

La géopolitique s'applique à un espace, ici l'espace numérique. Ce nouvel espace est abordé par la géopolitique depuis les années 90. La notion de pouvoir reste centrale.
Analyse du cyberespace via trois couches : matérielle (serveurs, câbles, machines, sites...) / logicielle (échanges des données entre machines) / sémantique (contenu informationnel). Cette analyse fait apparaître la position dominante des USA même si mutation avec Chine et Russie.
Les DN ne sont pas naturelles mais issues de choix humains politiques et stratégiques. Les DN sont créées selon des objectifs et en retour, modifient l'espace et le pouvoir.

I. De quoi les « données » sont-elles le nom ?

Donnée définie comme une transcription du réel. Les données ont toujours existé mais croissance à partir du 19e. La donnée devient un objet technique, indépendant de son usage social. Courant positiviste => possible de déterminer les phénomènes sociaux en appliquant des lois scientifiques comme le principe de causalité, sur un grand nombre de données. Depuis les années 70, dev. de la géographie quantitative aboutissant à l'utilisation du Big Data en géographie. Même conception du chiffre comme moyen et fin = fétichisation du chiffre.

Nouveautés apportées par les données numériques
Via les algorithmes, la corrélation entre données remplace le lien de causalité. Il y a mise en relations sur des données massives allant de plusieurs populations à l'individu, de nature variée, identifiant individuellement les produits ou les personnes, avec des évaluations permanentes (avis des internautes), un accroissement constant par les utilisateurs et une évolution technique parallèle.
Nouvelles pratiques => débats entre scientifiques sur la place du déterminisme dans l'analyse des DN. Sur les limites des statistiques. Sur l'obsolescence de l'hypothèse, de l'enquête puisque les DN ne semblent plus soumises à la subjectivité du chercheur ou de l'individu. Mais est-ce vrai ?

« La donnée prend la forme d'un nombre ou d'une caractéristique qui apporte une information sur un objet » (page 39). Différence entre D qualitatives et quantitatives / entre D structurées (tableur), semi-structurées (texte avec un titre), non structurées (photo). Métadonnées = données sur les données : elles favorisent les échanges entre systèmes.

Différence entre les données captées par observateur, expérimentation, instruments ou technologies et les données issues du fonctionnement des machines (ex traces laissées par l'internaute). La donnée brute est traitée pour devenir une donnée dérivée : les algorithmes cherchent des liens entre DN (*pattern*). Hypothèse qu'il existe des liens entre les DN analysées mais corrélations abusives à critiquer = critique de l'algorithme.
La DN n'est pas objective : elle est une traduction partielle du réel, elle est normalisée pour être traitée. Choix humain : quelles données capter ou traiter, comment les traiter etc. Cette démarche intellectuelle existe aussi quand la donnée est transformée en information ou en savoir. Cette transformation s'inscrit dans un contexte social, économique, culturel, politique etc. En analysant les DN captées ou délaissées, on peut définir le groupe social qui les produit, ses relations, son niveau technique, ses valeurs... Ex différence USA / Europe sur la notion de données personnelles.

Pour aider à la décision à partir des DN tout en gardant son esprit critique, l'auteur propose une méthode : suivre le cycle de vie de la D au cours de ses quatre étapes : production et collecte / transport / stockage / traitement pour en tirer de l'information. Et même 5e étape avec l'archivage et la suppression de la D. A chaque étape, il faut analyser les opportunités, les vulnérabilités, les dimensions géopolitiques et stratégiques. Tableau page 62.

II. Vers une territorialisation des données

Paradoxe : l'espace virtuel semble sans frontière. Dans la réalité, territorialisation d'Internet. Ex : cyberattaque sur l'Estonie en 2007. Les attaques (virus, sabotage) contre un pays ou le monde entier ont contraint les États à étendre leur contrôle territorial sur le cyberespace. Même réaction après les révélations de Snowden en 2013 : surveillance et l'espionnage généralisés des USA sur leurs alliés. Idem à la suite des interventions d'États dans les élections présidentielles étrangères.
États et acteurs privés cherchent à s'approprier le cyberespace. Apparaissent des territoires numériques. Qui dit territoire dit pouvoir et jeux d'acteurs à différentes échelles pour protéger des ressources, des valeurs ou tenter de les étendre.

Les *datacenters* (DC)
Centres de stockage et de traitement des DN sur des milliers de serveurs et commutateurs réseau. Extension avec le *cloud* qui permet aux particuliers et aux petites et moyennes entreprises d'accéder à des machines distantes puissantes.
Les DC = centralisation des DN / puissance de traitement entraînant le devpt de technologie comme les objets connectés, le *Big Data*, le *machine learning* / accès via le *data mining* à des données massives et production de nouvelles données.
Contraintes : disponibilité permanente des D donc logistique (câbles, énergie, refroidissement, débit...) / intégrité (surtout pour les secteurs à risque comme administration publique, banques...) / confidentialité = le plus délicat à assurer vis à vis des USA et de leurs lois extraterritoriales. Devpt du chiffrement, de réponses juridiques par les firmes du *cloud* contre les réglementations étatiques.
Pour protéger leurs DN et obtenir de la puissance, les États poussent au dev de *datacenters* et de *cloud* nationaux. Ils barrent la sortie des données sensibles du pays dans un but économique et politique.
Enjeu de souveraineté nationale. La notion de souveraineté numérique (contrôle des D et du cyberespace via ses trois couches) est de plus en plus souvent évoquée par les États soucieux de se protéger des grandes puissances (USA, Chine, Russie).
* Exemple de la France. Échec du *cloud* français. Dev de certifications pour les *datacenters*.
* Exemple de la Russie. Construction à partir de 2015 de DC en Sibérie après une loi qui oblige les sites internet de toute nationalité à stocker les D russes en Russie. Avantage : ni dépendance politique ni gain économique pour les USA. La Russie barre l'influence des GAFAM. Propose hébergement de leurs DN à la Chine et aux républiques d'Asie centrale : extension de l'influence russe.

Le flux des données
Les flux passent par l'interconnexion des machines (ordis, appareils). Le protocole commun est l'architecture TCP/IP. Il interconnecte les 50 000 sous-réseaux d'Internet à travers 500 000 points de transfert. L'étude des flux = champ de recherche large et complexe sur les couches physique, logique et sémantique.

* Exemple des câbles sous-marins : 95% des télécom' et des D Internet passent par eux. Les USA sont le nœud central. 97% des échanges de D entre Europe et USA passent par les USA car transit plus rapide via leur supériorité technologique. Multiplication de routes alternatives avec participation des acteurs publics et

privés. Ex. Google et Facebook : câble entre Los Angeles et Hong Kong. En plus du gain économique et d'une éventuelle captation des données (NSA), la possession d'un câble (par une firme ou un État) peut permettre du marchandage envers des États ou des populations et un affranchissement de toute supervision (d'un État ou d'un autre État). Augmentation de la puissance des Gafam suite à leur insertion dans le marché des câbles.

* Exemple des routeurs : points de passage entre réseaux. Depuis les révélations de Snowden, volonté de protéger les D nationales en empêchant le plus possible leur sortie du territoire. En Allemagne, Deutsche Telekom a créé un e-mail *made in Germany* : transit et stockage en Allemagne. Le routage devient intelligent pour éviter certains trajets. Centralisation et politisation du réseau en cours. Les États renforcent leur contrôle sur les D, les fournisseurs renforcent la régulation des contenus et des clients, les individus perdent le pouvoir.

* Exemple des plates-formes (PF) d'intermédiation (Google, Facebook...)
2 fonctionnalités : services personnalisés via lien direct avec l'utilisateur / intermédiation entre internautes. Plus les PF sont populaires, plus elles engrangent des D primaires et secondaires (clics, requêtes sur les moteurs de recherche) utilisées pour profiler les utilisateurs et prédire leur comportement. Cette prédiction a des limites. Ex de *Google Flu trends* qui devait prédire la contagion grippale mais les requêtes étaient supérieures aux cas réels.
Les PF génèrent des flux gigantesques. En Europe de l'ouest, elles profitent aux USA. En Chine, malgré l'existence de PF nationales (moteur Baidu, réseaux sociaux Weibo, WeChat...) et le blocage des PF et réseaux étrangers, un tiers des D sont captées par les USA (*trackers* dissimulés dans les publicités).

Les frontières existent bien dans Internet et les États sont toujours à l'œuvre selon les principes classiques de la géopolitique.

Le droit permet la territorialisation (data localisation) des DN
Conflit entre conception libérale et économique : les DN doivent circuler librement pour créer de la richesse (à la base des traités comme le TTIP ou le TiSA) / conception politique : la supériorité économique et technologique des grandes firmes américaines menace l'industrie numérique française et européenne.
Leur richesse vient de la commercialisation des D secondaires. Leur popularité renforce leur position de leader. Leur avance technologique et leur puissance financière entraînent des investissements colossaux donc elles sont toujours innovantes. Leur clientèle est mondiale donc moins soumises à des aléas territoriaux.

Cas particulier de l'extraterritorialité du droit américain
La puissance des Gafam permet un contrôle mondial des D par les USA car leurs CGU (conditions d'utilisation) se réfèrent au droit américain. Idem pour les lois antiterrorisme. Actuellement, le gvt américain peut contraindre les fournisseurs de service américains à lui fournir des D conservées dans les serveurs américains aux USA et à l'étranger. L'Europe protège ses D par le RGPD : les entreprises et leurs sous-traitants doivent protéger l'intégrité, la confidentialité et la disponibilité des DN des ressortissants européens sur l'UE et hors UE. Mais cette loi peut-elle s'appliquer contre le *Cloud Act* américain ?

Le droit est l'instrument de la data localisation et de la restriction de circulation des DN. Beaucoup de pays adoptent des législations protectionnistes : pour protéger leur économie, pour assurer la confidentialité des D et protéger des valeurs, pour diffuser leurs valeurs contre les valeurs américaines (ex Russie, Chine).

Débat idéologique autour de la fragmentation d'Internet. Qui doit gouverner Internet ? Position américaine : les États, les acteurs privés, les sociétés civiles. Mais cette position conforte la puissance américaine car les Gafam sont plus puissantes que les États.
Position de la Chine et de la Russie : gouvernance du cyberespace par les États via l'UIT, agence de l'ONU (Union internationale des télécommunications). De fait, la gouvernance actuelle d'Internet est mixte et complexe au niveau physique comme dans celui des contenus. Au niveau national, devpt d'une industrie numérique avec infrastructure et PF (ex Yandex, Baidu). On parle des BATX, l'équivalent Gafam des géants chinois. Alphabets propres.

Cyberguerre : définition floue. Attaques sur toutes les couches d'Internet (tableau p 117) par de nombreux acteurs (hackers, cybercriminels, *dark web*). Beaucoup de questions : certifier la source d'une cyberattaque,

acte de guerre ou non, rôle des acteurs privés. Impasse des négociations internationales sur la cyberconflictualité, chaque État est conscient du risque de déstabilisation mondiale mais devpt des capacités offensives contre les autres. Idem pour les acteurs irréguliers. Nouvelles armes : *fake news*, trolls... disponibles pour tous.

III La géopolitique à l'épreuve des données

L'espace numérique conduit-il à une modification de la géopolitique même si les bases de la discipline sont toujours valides ? La société actuelle est toujours davantage mise en données. Comment observer et analyser la société, le rôle de l'État, les rapports entre espace et pouvoir ?

Adage célèbre du juriste Lessig : *Code is law*. Le code (algos, protocoles, logiciels) fait le cyberespace. Le code n'est pas neutre mais dépend de ce que l'humain met dedans (ex du plein accès aux D personnelles ou non). Par csq, chaque espace physique peut être doublé d'un espace codé. Ex du réseau routier (code gestion feux de circulation). Interaction entre les deux espaces. Ex : code dans le smartphone -> D corporelles qui modifient le comportement des connectés. Le code régule l'espace du quotidien et crée des valeurs. Sur tous les territoires, il existe une couche invisible de D.

Ex du champ de bataille comme espace dédoublé. Les DN collectées et traitées sont depuis longtemps incorporées à la stratégie. Mais elles servent aussi l'adversaire, surtout les acteurs irréguliers (récolte d'informations sur otages, sur mouvements des forces de l'ordre). Techniques de profilage comme dans le domaine commercial pour identifier des ennemis potentiels. But : éliminer les personnes-clefs. Les menacer ou leur entourage. Vie privée et vie professionnelle se confondent.
Les DN sont utilisées sur le champ de bataille : capteurs sur les équipements, les véhicules, les drones. Traitement des D par les algos et renvoi d'informations entre composantes. Ouvre un grand nombre de questions : comment trier les D ? Comment garder la main sur les algos ? Quelles relations entre le commandement et les hommes sur le terrain ? Problème : la mise en réseau de flux de D augmente la possibilité des attaques (ex logiciels malveillants). Autre problème : la prédiction à partir des DN pour anticiper les zones instables conforte les acteurs privés du numérique (marché du maintien de l'ordre, de la surveillance des réseaux, de la gestion des réfugiés etc).
Dans le domaine de la sécurité comme de la défense, traitement des DN pour deviner les intentions de l'adversaire à partir de la théorie des jeux, modélisation. Ces modèles sont-ils efficaces ? Peut-on vraiment prédire la violence ? Nécessité pour les décideurs de connaître les limites de ces expertises statistiques.

Ex des frontières. Avec la mise en réseau des appareils et des capteurs détectant les mouvements, les frontières sont mouvantes et ne concernent plus seulement les limites topographiques d'un territoire. On parle de *smart borders*. Ex de la frontière dans un aéroport. Objectif : gestion des flux. Autorisation / interdiction du passage non plus selon la nationalité (visa) mais selon l'individu via croisement de nombreuses bases de données. Sélection des individus selon des prédictions statistiques même si contournement possible (chirurgie). Cette surveillance généralisée entraîne une adaptation des législations, une coopération internationale et une alimentation continue des bases de données.

Data surveillance. Algos plus D semblent indispensables pour la gestion actuelle et future d'un monde efficace et sécurisé. Extension continue de la data surveillance (puces RFID, GPS, smartphones). Ex de la gestion des *smart cities* : la production énergétique est couplée à leur utilisation par les habitants. Mais méconnaissance des critères de décision du réseau. Prédiction statistique de l'algorithme aboutit à la supériorité du groupe sur l'individu, des statistiques sur le comportement réel de l'individu.
On aboutit à une inversion de la relation entre risque et surveillance. La surveillance généralisée permet d'identifier des risques alors qu'avant, on surveillait les risques potentiels. Csq : croissance de la surveillance pour établir un maximum de liens entre les DN. Ex de la cartographie prédictive de la police. Nombreuses erreurs mais se développe quand même pour rationaliser l'utilisation des effectifs. Même tendance dans la surveillance élargie du social sans que soient spécifiées à l'avance les raisons et les besoins (NSA).
L'architecture spatiale traditionnelle de la surveillance = points, lignes, aires. Aujourd'hui = grille régulière de capteurs en réseau sur toute la Terre. C'est le modèle de la *smart city* : toutes les DN collectées peuvent être connectées et servir à la sécurité (caméras, photos personnelles postées etc.).
Les bases de D permettent de prédire le comportement futur des individus. La collecte s'élargit aux groupes en contact avec l'individu. Problème d'atteinte aux libertés des personnes.

Les DN ont modifié le territoire de la géopolitique. Les échanges de D et la suppression de la distance = perte de la dimension topographique (distance + étendue). *« Le territoire semble s'effacer derrière le réseau, et la topographie derrière la topologie » page 162*. L'espace est redéfini comme l'ensemble des relations via les connexions. Entraîne une transformation des rapports entre pouvoir et espace. L'auteur propose la notion de pouvoir topologique = comment les acteurs influent sur des lieux en dehors de toute distance physique. Il n'y a plus d'intérieur et d'extérieur. Le pouvoir exerce une influence sur un lieu et peut en sortir. L'individu ne le peut pas sauf à échapper au réseau. Reste à trouver des méthodes pour analyser (et cartographier) ces pouvoirs topologiques (ex des influences sur une élection via Internet, ex de la propagande de Daech).

Conclusion

La géopolitique des DN est un champ d'étude prometteur. Les analyses doivent se faire au niveau international (États, acteurs territorialisés) et au niveau social (rapports de force sociétaux et nouvelles formes d'expression du pouvoir via les outils numériques). La gestion des populations change jusqu'à interroger la politique et l'éthique car germes de contestations. Nécessité de critiquer les DN car loin d'être neutres et objectives, elles sont des constructions sociales. Le fort investissement des acteurs étatiques et privés dans les DN (*datacenters*, câbles, industrie numérique) prouve que les D sont des instruments de pouvoir et d'influence (*hard power*, *soft power*). Les acteurs privés ont des capacités supérieurs aux États = modification de leurs relations mutuelles. On a redistribution des pouvoirs et disparition progressive des séparations entre guerre et paix, militaire et civil, lieux de guerre et lieux quotidiens. La guerre n'appartient plus aux seuls États, chacun peut la subir ou y participer via les réseaux sociaux ou les cyberattaques.

Les gouvernements mettent en place des contrôles sur les territoires numériques (populations, marchandises, idées. Les questions posées à la géopolitique par l'usage des D sont immenses : acteurs, stratégies, relations de pouvoir, impact sur les populations, pertinence des analyses du Big Data, légitimité de la technique qui masque les pouvoirs à l'œuvre. Mêmes questions pour les méthodes de cette géopolitique des D. Les sciences sociales doivent étudier les techniques derrière le *cloud*, les réseaux sociaux, l'intelligence artificielle... en travaillant avec des experts techniques.

***

Aucune référence bibliographique relevée car le livre est complet.
Citation retenue : « Cette numérisation sans précédent des données les transforme en des objets essentiellement techniques, qui paraissent de plus en plus déconnectés de leurs usages sociaux réels. » (page 23)

Iconographie
Page 62-63 : Tableau « opportunités, vulnérabilités, dimension stratégique et géopolitique » du cycle de vie de la DN
Page 117 : tableau des attaques et menaces potentielles selon les couches du cyberespace

Appréciation personnelle
Bonne synthèse sur l'importance des DN en géopolitique et les transformations de la discipline.
Remarques : il manque une distinction sur les données non numériques entre celles qui peuvent le devenir (numérisation des données imprimées, numérisation en 3D des bâtiments...) et celles qui ne le peuvent pas (paroles non enregistrées et à jamais perdues etc.). Limite temporelle.
Dans le cycle de vie de la DN, il manque la notion de transformation de la donnée, par ex son anonymisation.
Les données numériques ne se confondent pas avec Internet.