

Géographie de l'internet

Éric Guichard*

2005

Note de référence : *Ce texte a été publié en 2007 dans l'ouvrage Lieux de Savoir (Dir. Ch. Jacob, Paris, Albin Michel), sous le titre «Géographie de l'internet» (pp. 989–1009). Il a été rédigé en 2005, et les différences entre cette version et l'imprimée sont mineures : bibliographie citée en note, ajout d'une carte du radar de l'internet (transmise par Matthieu Latapy) et attribution de majuscules aux mots «web» et «internet» caractérisent l'édition d'Albin Michel.*

1 Historique

1.1 Routes et flux

L'internet est un système de communication entre ordinateurs développé aux États-Unis dans les années 1970. Comme souvent avec l'informatique, le dispositif est un mélange de bon sens et d'algorithmes sophistiqués. En fait de bon sens, citons les «dialogues» entre les machines, qui sont en fait des textes formatés, de façon très protocolaire afin de pouvoir être interprétés par ces autres textes que sont les programmes locaux ou inter-ordinateurs. Par exemple, la convention de l'internet veut que les machines se «saluent» en trois «poignées de main» et se «disent au revoir» en quatre. Celle qui engage le contact est appelée le «client», la machine destinataire, le «serveur»¹.

Les deux protocoles de l'internet sur lesquels s'adosent ces échanges écrits sont TCP (Transport Control Protocol) et IP (Internet Protocol). Le premier garantit la bonne transmission des flux entre les ordinateurs, indépendamment de leurs systèmes d'exploitation, des réseaux par lesquels ils sont reliés, des diverses *gateways* (passerelles) qui permettent de basculer d'un réseau à l'autre. TCP découpe les flux en petits paquets appelés *datagrammes*, vérifie leur acheminement, les réordonne au besoin à l'arrivée. Encore faut-il que les machines soient bien identifiées et qu'existe une «route» permettant de les joindre. C'est là que le second protocole intervient. Il s'agit tout d'abord d'étiqueter chaque ordinateur. Dans la version historique de IP (Ipv4), chaque ordinateur dispose d'un numéro codé sur 4 octets, ce qui offre virtuellement un jeu de 256^4 possibilités, soit environ quatre milliards. *129.199.96.11* est un tel exemple d'étiquette.

*Maître de conférences à l'Enssib, directeur de programme au Collège international de philosophie, responsable de l'équipe *Réseaux, Savoirs & Territoires* de l'Ens.

1. Ces descriptions de ce qui n'est qu'une circulation d'électrons entre des architectures de silicium peuvent sembler très anthropo-centrées. Ce qui est prévisible, puisque les ordinateurs sont configurés à l'image de leurs concepteurs.

La route, elle, est choisie au coup par coup : hormis le cas où deux machines appartiennent à un même réseau local (souvent de type *Ethernet*, où toute communication entre deux ordinateurs est «vue» par tous les autres), on pourrait dire que la globalité de l'internet renvoie à une sorte de grand inconnu, juste jalonné par quelques serveurs particuliers, susceptibles d'aider à l'aiguillage d'un paquet un peu perdu cherchant son destinataire. Ces serveurs sont appelés «routeurs». Ils disposent de listes des numéros IP qu'ils peuvent contacter, ou pour leur adresser directement des paquets, ou pour leur demander à qui les adresser quand l'ordinateur destinataire leur est inconnu ; en quelque sorte, des panneaux indicateurs, dont la dernière pancarte proposerait «autres directions». Ainsi, le protocole IP, avec les étiquettes attribuées aux ordinateurs et les divers routeurs, choisit un chemin précis, et TCP joue le rôle d'un facteur aveugle, mais méticuleux.

TCP et IP se moquent (presque) du contenu du flux acheminé : courrier, requête web, calcul à distance, etc. sont pour l'un et l'autre des «applications de niveau supérieur», simplement repérées par des «ports», très vite intégrés dans les paquets, mais sans importance lors de leur transport.

1.2 Des machines aux usages

Ce dispositif est trop peu mnémotechnique pour satisfaire les humains. Imaginons-nous acheter un billet de train sur le site <http://129.189.13.2> ou envoyer un courriel à moncollegue@193.252.22.89 ? Les «alias» www.sncf.com ou wanadoo.fr semblent plus commodes.

C'est pourquoi il fut décidé, très tôt, d'attribuer des «noms» aux plages IP attribuées aux entités raccordées à l'internet. Par exemple, l'ensemble des numéros *129.199.0.0* à *129.199.255.255* (autrement dit, la plage *129.199.*.**, l'étoile symbolisant tout nombre entre 0 et 255) renvoient au «domaine» de l'École normale supérieure (ENS), qui a pour «nom» : *ens.fr*. Et au niveau inférieur, les machines sont aussi souvent dotées de nom. Par exemple, le serveur *129.199.133.50* «s'appelle» : *barthes.ens.fr*. Cette association entre noms et numéros est parfois considérée scandaleuse par les informaticiens, car elle contredit la décentralisation de l'internet et nécessite des tables de conversion entre noms et numéros. Ces dernières sont inscrites dans les DNS (*Domain Name Servers*). Il y en a un par domaine, et il ne décrit que les machines qui lui appartiennent. Ainsi, le DNS de l'ENS dispose du numéro IP de la machine «*barthes.ens.fr*» mais son gestionnaire n'a pas le droit d'y ajouter le numéro IP d'une machine de l'université de Rouen, même s'il le connaît. Ce qui fait que pour consulter le site web «*www.univ-rouen.fr*», il est nécessaire à un ordinateur de l'ENS d'interroger le DNS qui gère les noms de domaine «.fr» (donc français, hormis les «.com» de France), puis celui de l'Université de Rouen, alors que la connaissance du numéro IP associé à ce site (*193.52.144.50*) aurait permis d'éviter cette procédure supplémentaire et génératrice d'une hiérarchie encombrante.

Depuis, la notion de domaine s'est encore plus complexifiée, du fait de la commercialisation des noms associés.

1.3 Poids de l'histoire

L'internet s'appuie donc sur deux protocoles aux fonctions bien différentes. Dans les années quatre-vingts, on le dénomme «réseau TCP/IP» (TCP sur IP). Aujourd'hui encore, malgré diverses tentatives de remplacer TCP par UDP, un facteur plus laxiste, et malgré les efforts de nombreuses entreprises pour promouvoir des protocoles «propriétaires» —dont l'usage, par des tiers, aurait pu mener à des redevances—, 90% des applications de l'internet fonctionnent avec TCP. IP est en théorie faiblement centralisé, mais l'introduction des noms de domaine lui impose un fonctionnement hiérarchisé.

Ces protocoles s'appuient, bien sûr, sur des réseaux physiques, que nous ne pourrions négliger : câbles, relais hertziens, satellites, etc. L'internet n'a rien de virtuel. En tant que réseau de réseaux, il est ancré dans la matérialité, même si TCP/IP se joue des différentes contraintes de ces derniers.

Mais lors de la conception de l'internet le succès actuel n'était pas prévu : au milieu des années 1980, le réseau était constitué de 3600 nœuds, pour environ 100 fois plus de machines (en France, en 1990, il y avait 50 tels nœuds, pour 3500 ordinateurs connectés).

Ce succès est dû aux physiciens, qui ont inventé le web entre 1989 et 1991 : le web est un protocole de publication («http», ou HyperText Transfer Protocol), conçu au CERN² — en Europe, donc—, et dont les débuts furent laborieux : la légende veut que l'article de Tim Berners-Lee et Robert Cailliau qui en détaillait le fonctionnement faillit être refusé au motif qu'il n'apportait rien de nouveau, et en 1993, on ne dénombrait que 50 serveurs web dans le monde. Ce qui n'est rien au regard des quatre milliards de pages web que l'on a recensées, 10 ans plus tard.

Le caractère récent de cette «technique» donne à penser que l'internet reste dépendant de ses premiers usages : ce fut, avant tout, un outil de scientifiques pour des projets scientifiques. Les informaticiens s'en sont servi à des fins réflexives (développer l'internet) ; les mathématiciens pour développer leur réseau ; les physiciens avaient pour projet d'optimiser l'«efficacité cognitive» de leur communauté étouffée par les multiples données, articles, graphiques, schémas, etc., aussi indispensables à leur production scientifique qu'ils étaient difficiles à recueillir, à classer, à repérer.

Et si aujourd'hui, les usages de l'internet gagnent en variété (consultation de catalogues pour réaliser des achats par correspondance, abonnements à des publications peu ou pas rationnelles, visionnage de films, etc.), les personnes sont, aussi, de plus en plus nombreuses à reproduire les pratiques intellectuelles des usagers initiaux. L'intérêt que lui manifestent les étudiants, enseignants, chercheurs, mais aussi l'ensemble des utilisateurs cherchant à savoir, en témoigne. Encore faudra-t-il méditer sur la façon dont les modes d'élaboration de la rationalité propres aux «sciences exactes» ont déteint, avec l'internet, sur ceux des sciences humaines.

À cette inertie des usages s'ajoute le fonctionnement même de l'internet, qui, on l'a vu, relève de l'échange entre machines. Conçues par les humains, elles prolongent les outils en lesquels ceux-ci ont toute confiance. Dans la mesure où l'écriture est la «technologie de l'intellect» que nous avons privilégiée depuis quelques millénaires, ce sont des textes, et seulement des textes qui circulent entre ces machines. Ces deux points aident à comprendre la réalité de l'internet et son influence, en tant qu'élément de notre système d'écriture, sur nos représentations et nos cultures. D'autant que, malgré les utopies des tenants de la convergence et de la

2. Organisation européenne pour la recherche nucléaire, voir l'URL <http://cern.web.cern.ch/CERN/CERNName.html> pour l'origine de l'acronyme.

consommation instinctive, l'internet n'a aucune raison de transformer nos sociétés en nous apportant plus de liberté et de démocratie, ou en rendant caduques les notions d'effort intellectuel ou de détermination sociale. David Edgerton détaille très clairement les errances et les méfaits d'une telle idéologie (le déterminisme technique). Il rappelle que l'on mesure la maturité d'une technique à l'ampleur des détournements qu'elle subit (Edgerton, 1998). Ces signes sont lents à survenir (parfois de l'ordre d'un siècle), et témoignent du fait que c'est le social qui transforme la technique, et non l'inverse.

Reste que cet outillage mental contemporain «déboussole» : nous ne cessons d'évoquer réseaux, routes, transports et aiguillages, sans jamais préciser aucun lieu...

2 Dissolution de la géographie

Nous appelons territoire la superposition d'un espace physique et de pratiques sociales. Cette définition est ambiguë quant aux personnes qui «peuplent» ce territoire : elle fonctionne en creux, en négatif. Ce qui est logique, car les notions identitaires relatives au territoire se construisent beaucoup à partir de l'altérité : il est plus commode de définir le territoire des autres, et d'en déduire le sien, que de faire l'inverse.

Le territoire transporte, au moins, deux notions : celle de distance, liée à l'espace physique, et celle de culture, par le biais des identités collectives projetées ou revendiquées. Il n'est pas sûr que l'addition de ces deux ingrédients génère, à son tour, du territoire, et notre propos n'est pas de tester une telle hypothèse. Mais il nous paraît utile de montrer comment l'internet a, du fait de sa construction, ébranlé la notion commune de distance et sollicité, par réaction, celle de culture. Peut-être pourrions-nous alors confirmer la conjecture d'Henri Desbois, qui affirme l'existence d'un territoire de l'internet (Desbois, 2001).

La co-présence de deux identifiants que l'on croit souvent équivalents, le numéro IP et le nom d'une machine, et par extension, la plage IP et le nom de domaine génèrent deux effets complémentaires : les chiffres détruisent le territoire et fabriquent de l'espace ; les noms, du fait des pratiques sociales qui s'organisent autour d'eux, construisent de repères culturels et des territoires virtuels, qui peuvent facilement devenir réels.

2.1 Géométrie d'IP

Initialement, l'association entre noms et numéros était relativement biunivoque, et respectait une hiérarchie descendante. Et le non-respect de ces équivalences est passé du statut de possible à celui de norme : derrière un seul nom, peuvent se cacher des dizaines de milliers de machines (pensons aux ordinateurs derrière l'URL³ –Uniform Resource Locator– d'un moteur de recherche), et des centaines de noms achetés peuvent tous renvoyer aux fragments d'une même machine (par exemple, un hébergeur de sites individuels).

À la différence d'un nom de domaine, un numéro IP est trop important pour être livré aux lois du marché. Le découpage historique de ces numéros en «classes» est avant tout géométrique : le «cube» de dimension quatre (et de côté 256, rappelons que, un numéro IP est composé

3. L'URL correspond à une adresse ou à l'emplacement d'une ressource en ligne. Le plus souvent, un URL définit une page Web fixe. Ex. : <http://barthes.ens.fr/atelier/geo>. Un URL peut référer à des services plus complexes comme les réponses d'un robot à la requête qu'on lui adresse. Certains auteurs français attribuent à l'URL un genre féminin.

de quatre chiffres compris entre 0 et 255 et séparés par des points) de ces numéros a d'abord été coupé en 2 : d'abord la classe A, composée de 127 plages gigantesques —ou domaines— de 16 millions de numéros IP ($x.*.*.*$, x allant de 1 à 127) ; soit 127 cubes de dimension 3 (et de côté 256) réservés aux plus grandes institutions demandeuses dans les années 1980. La moitié restante de l'«hypercube» a aussi été coupée en deux parties égales : d'une part, 16 000 carrés ($x.y.*.*$, x et y fixés définissant le domaine, y choisi entre 0 et 255, x étant choisi dans la première moitié du segment [128,255], donc entre 128 et 191) de 65 000 numéros IP ; enfin, le quart du pauvre, soit 2 millions de segments de 256 numéros ($x.y.z.*$, x choisi entre 192 et 223).

Ce découpage est aussi politique : la distribution de ces numéros se fait sous le contrôle de l'IANA (Internet Assigned Numbers Authority, www.iana.org) et de certaines institutions continentales ou nationales auxquelles elle a donné délégation.

L'espace fini défini par cet «hypercube» construit par un simple jeu d'écriture n'est pas un territoire. Il est cependant indispensable au fonctionnement technique de l'internet. Bien entendu, des solutions ont été trouvées pour pallier les carences des numéros IP (et l'avenir nous promet une répartition plus généreuse de ces numéros, avec la mise en place d'IPv6 : numéros IP sur 16 octets au lieu de 4). De plus, à la suite d'abandons de privilège ou de négociations, des portions de segments, de carrés et de cubes ont été attribués à de nouvelles institutions.

De l'histoire de l'internet et des transactions liées à la redistribution des numéros IP, il s'ensuit qu'il n'existe aucune relation entre les numéros IP et la géographie : deux numéros «voisins» peuvent être attribués à des machines situés à des milliers de kilomètres l'une de l'autre, et inversement. En outre, aucun emboîtement ne prévaut : si la plage IP 129.199.*.* correspond à un domaine précis (en l'occurrence, celui de l'ENS), la plage 129.*.* ne correspond à rien du tout, et surtout pas au domaine *France*. Sur ce plan, l'Internet a une composante spatiale forte, bien qu'il soit destructeur de territoire.

2.2 Explosion des catégories

Cette structuration géométrique de l'internet fait penser à la fonction initiale des réseaux : irradier un espace de façon à y renforcer un pouvoir précis. Mais elle a aussi bousculé nos repères culturels, en faisant exploser toute classification, et par suite, la géographie : sans verser dans la mythologie, on peut imaginer un «Flower Power» à la fois californien et mondial, revendiquant comme Einstein le droit de tirer la langue au vieux monde, suite à une compréhension intime du rapport entre écriture, mathématique, et ordonnancement.

Pour nombre des informaticiens et mathématiciens à l'origine de l'internet, il pouvait s'avérer légitime de construire l'internet en négligeant les questions relatives aux catégories des sciences humaines, puisque ces questions apparaissaient insolubles par le biais de règles ou d'algorithmes. Par exemple, classifier les domaines IP par disciplines relève de la gageure, quand on constate leur réorganisation sous l'effet de la recherche : les informaticiens le savent bien, puisqu'ils ont rencontré d'énormes difficultés pour rendre autonome leur domaine face aux mathématiques et à l'électronique. L'histoire ne montre-t-elle pas que toute classification porte en elle les germes de ses contradictions futures ? Le regroupement d'Otlet, qui mettait la biologie et l'anthropologie dans le même registre, choque aujourd'hui, comme celui des successeurs de Dewey, qui confondent fascisme et communisme (Chevillotte, 2005).

Comment tenir compte de la géographie ? Là encore, il ne semble pas possible de poser une règle qui permette de prédéfinir les numéros IP de chaque pays, d'imaginer celle qui pourrait

s'appliquer aux futurs pays, ou au contraire, aux éventuelles fusions ; de même pour les parts relatives des multinationales, des organisations pacifistes et humanitaires, des gouvernements, etc.

Du fait que l'internet est d'abord une construction intellectuelle de scientifiques prédisposés à l'établissement de normes explicites et universelles, il s'ensuit une destruction des anciennes notions de distances physique et culturelle.

Il ne s'agit pas pour autant de promouvoir une histoire positiviste et apolitique de l'internet : par exemple, la désignation de noms de domaine de haut niveau pour les possessions des anciens empires (Réunion, Guyane, etc. pour la France) ne fut pas anodine, le refus de créer un nom de domaine pour l'Union Européenne non plus. Le positivisme, à comprendre comme une idéologie revendiquée par les fondateurs de l'internet, a pu aussi se développer du fait de leur contribution au nouvel essor de la technologie de l'intellect : en «écrivain» le monde et ses nouveaux réseaux, ils le sculptaient, le façonnaient. Cette idéologie s'est nourrie du fait que cette puissance créatrice de l'écriture était d'autant plus efficace qu'elle n'avait guère de concurrents. Au point que les simples processus de dénominations des machines et des domaines fabriquaient du réel, de l'économique.

2.3 Désignation, monnaie et culture

Autant la numérotation de l'internet a été classée «domaine réservé», et maintenue suffisamment absconse pour ne pas avoir d'effet symbolique, autant les noms de domaine en clair et les jeux langagiers qu'ils permettent ont été «livrés au marché».

Un nom de domaine, dont un exemple générique pourrait être «*monsitewebamoi.com*» s'achète, se vend, se négocie. Le prix de base pour un nom vierge est de 14 dollars l'unité (le symbole, voudrait-on dire). Les transactions peuvent sembler libres et sans entraves, mais là encore, l'autorité suprême est américaine. C'est l'ICANN (Internet Corporation For Assigned Names and Numbers), émanation du ministère américain du commerce, qui elle-même délègue à une poignée d'entreprises la gestion de ces noms. La plus connue est Verisign, responsable des «.com» (et des «.net»). Le coût pour l'entreprise s'élève à moins de 2 dollars. Le jeu d'écriture rapporte, donc, net 12 dollars. Et la manne jamais ne se tarira, car un nom de domaine ne s'achète pas. Il se loue : les tarifs précédents s'entendent à l'année.

On retrouve, ici, la capacité d'un système symbolique à produire de la valeur. Cette valeur n'a aucun rapport avec une quelconque productivité ; elle est intimement liée à l'écriture et à l'énonciation, laquelle sollicite autant la croyance que la raison. L'on est ici dans un registre quasi-performatif à la façon d'Austin, et l'on connaît depuis Searle les fonctionnements de ce système sur un plan économique et social (Searle, 1998), tout comme l'on dispose désormais d'exemples historiques de la relation forte entre monnaie et écriture (Herrenschmidt, 1999).

Pour exister, on s'inscrit, on s'affiche, on achète des mots, parfois très cher : en 2000, la société Vivendi décide de lancer un portail internet, «*vizzavi.fr*». Réalisant qu'une marque homonyme était déposée par les propriétaires d'un cybercafé parisien du 18^e arrondissement, elle leur a offert 24 millions de francs (environ 3,7 millions d'euros) pour qu'ils renoncent à toute plainte. La nouvelle a d'autant défrayé la chronique que l'un des trois propriétaires était bien loin du monde capitaliste : c'était Ababacar Diop, une des figures marquantes du mouvement de défense des «sans-papiers».

Que devient le territoire ? Si ces appellations sont susceptibles d'en produire, c'est certainement par l'intermédiaire des références culturelles qu'elles véhiculent. Quand on découvre des

sites, des domaines ou des machines identifiables par des noms comme «bbc.com» ou «univ-limoges.fr», on pense qu'ils sont «situés» à Londres dans le premier cas, à Limoges dans le second. Peu importe que ce soit vrai ou faux ; la dénomination génère un ancrage proprement géographique. Inversement, cette impression ne vaut pas pour des noms comme «bcd.com» ou «www.fsk.cn». La perte de la référence culturelle est alors totale, ce qui déstabilise et induit l'idée d'une dé-territorialisation de l'internet.

Les interrogations de nombreux universitaires et documentalistes quant à la validité ou la légitimité scientifique des contenus des sites internet semblent liées à ce potentiel qu'ont leurs noms de générer ou non une référence culturelle : il y a là un automatisme cognitif, rarement explicité dans le domaine bien balisé de l'imprimé (avec ses revues savantes, ses ouvrages, et leurs hiérarchies respectives) qui témoigne de la relation entre la culture et le territoire.

2.4 Territoires virtuels

Les numéros IP détruisent les territoires intellectuels en même temps qu'ils fabriquent de l'espace. Les noms de domaine se rapprochent des territoires car ils sont porteurs d'échanges sociaux (transactions, conflits...) et mettent en évidence la relation entre géographie et culture.

Mais cette relation entre l'internet et le territoire reste virtuelle. Elle va cependant pouvoir se réaliser par le truchement d'un mode très ancien de repérage géographique : l'itinéraire, au détriment de la carte. Car, s'il est nécessaire pour une institution (organisation, centre de recherche, entreprise, ministère) d'être aujourd'hui inscrite sur le web pour exister, l'enseigne ne suffit pas : encore faut-il que de multiples panneaux indicateurs puissent y conduire. Et, de façon surprenante, dans ce registre formel, intimement attaché à l'écriture, l'usage élaboré de la technologie contemporaine de l'intellect paie, au sens fort du terme ; de ce point de vue, les mondes lettrés ne sont pas les plus démunis.

3 Géographie propre à l'internet

Cette géographie peut se reconstruire en retrouvant les lieux de l'internet, puis en définissant des distances entre les contenus du web.

3.1 Géographie «physique»

3.1.1 Localisation statique des serveurs

Moyennant des efforts d'intensité variable, il est souvent possible de localiser une machine. Déjà, on peut tenter de retrouver son domaine à partir de son numéro IP, ou l'inverse (fonction *resolver*, qui interroge les DNS).

La plupart des informations légales figurent dans les bases «whois». Celles-ci, assez bien documentées pour les extensions nationales d'Europe («.fr», «.de», «.se», etc.), grâce au travail des RIPE (Réseaux IP Européens, délégation régionale de l'IANA), peuvent l'être moins pour leurs équivalents des autres continents. Depuis quelque temps, les «.com» sont relativement bien enregistrés par Verisign, mais leur localisation à la volée s'avère plus difficile que pour les machines gérées par les RIPE.

Dans tous les cas, les informations géographiques de ces bases sont pauvres et peu structurées : par exemple, pour les noms de domaines, on obtient une à trois adresses, dont on ne

peut déduire automatiquement laquelle réfère aux propriétaires du domaine, ou à l'un de leurs gestionnaires, secrétaires, prête-noms, etc., les uns pouvant être à des milliers de kilomètres des autres, et bien sûr des machines. On peut aussi vérifier de façon empirique si la machine visée (ou son domaine) dispose d'un site web, et si celui-ci contient une page avec une adresse géographique ou postale (explicitement ou par le biais de «métadonnées» aisément interprétables).

On comprend que la localisation d'un ordinateur ou d'un serveur est délicate, «coûteuse» en temps, et parfois infructueuse. Il est encore plus difficile d'obtenir des résultats en série, à moins de ne viser que des sites déjà connus même partiellement.

3.1.2 Localisation dynamique

Un vieil outil comme *traceroute*, qui décrit le chemin suivi par un paquet pour accéder à une machine donnée et les délais entre deux étapes consécutives, conserve un certain succès. Déjà, il permet de préciser le lieu approximatif d'une machine, sauf si celle-ci appartient à une organisation ou à un réseau qui refuse de répondre : c'est d'ailleurs le cas pour certaines entreprises (avec leurs intranet), mais aussi pour des fournisseurs comme AOL, dont les abonnés américains semblent tous situés de ce fait à New York ou à Seattle.

Au début des années 2000, des chercheurs ont eu l'idée de réaliser une «goniométrie» de l'internet avec *traceroute* : à partir de quatre ou cinq machines éloignées les unes des autres, et en calculant le temps d'accès de chacune à l'ordinateur dont ils recherchent la localisation, ils arrivent à estimer sa position à une centaine de kilomètres près (Ziviani et al., 2005). Cela vaut, bien entendu, si la vitesse des paquets est constante : pratiquement, dans une zone — d'Europe, d'Amérique du Nord, etc.— où l'on sait que les dorsales (*backbones*) de l'internet sont composées de fibre optique (la vitesse d'un paquet valant alors les deux tiers de celle de la lumière).

3.1.3 Une incessante exploration

On le voit, décrire la géographie de l'internet nécessite des méthodes variées, des hypothèses fortes, et relève partiellement du travail artisanal. Certaines entreprises se lancent maintenant dans ce travail de localisation. Signalons Geobytes (www.geobytes.com/IpLocator.htm?Get-Location), aussi précis pour les ordinateurs américains que lacunaire pour les ordinateurs européens —incapable de donner l'adresse du site www.ens.fr en juin 2005, mais sachant celle du site www.enst.fr—, et le moteur de recherche français Exalead (www.exalead.com), qui propose, quand c'est possible, de trier géographiquement les réponses à une requête donnée. Mais là encore, bien du travail reste à faire. Malgré les discours des technophiles, aucune norme, aucun encodage du futur proche n'imposent une localisation géographique des machines connectées d'une façon ou d'une autre à l'internet.

Paradoxalement, ces lacunes donnent aussi une dimension géographique à l'internet, et plus encore au web. On distingue les espaces institutionnels, qui s'affichent, bien balisés, au coût de localisation faible, qui apparaissent comme autant de pôles. Puis les espaces complètement privés, dont on arrive parfois à repérer les frontières et clôtures. En revanche, l'internet non exploré, mal cartographié, mélange de Pigalle et de mers asiatiques, s'esquisse comme une somme de non-lieux souvent agressifs, tirant parti du caractère écrit, donc aisément falsifiable, de l'internet.

3.2 Géographie culturelle

Pour toute personne ayant un projet éditorial (web, ressources audio-visuelles, etc.), l'important n'est pas tant de localiser ou d'être localisé que d'être joint. Là encore, il est délicat de décrire un tel phénomène sans abuser de références spatiales et dynamiques, même si les questions du lieu et de l'itinéraire s'affranchissent de la géographie classique. Avec l'internet, c'est l'ensemble de nos catégories analogiques qui explose : IP est un cube numérique sans ordre. TCP est un facteur aveugle qui n'a pas de souvenir de sa tournée. On ne sait pas où sont les ordinateurs. Leurs noms n'ont pas de sens, et rien ne structure leurs contenus. On imagine alors que les notions de proximité et de distance à inventer auront peu de rapport avec la géodésie ou avec nos classifications intellectuelles, mais qu'elles spatialiseront de fait l'internet, et accroîtront encore sa virtualité territoriale.

Dans ce contexte, les maîtres mots sont «graphe» et «moteur de recherche». Le graphe du web, tout comme le fonctionnement des moteurs de recherche, fait l'objet d'une très abondante littérature⁴.

3.2.1 Le graphe du web

Un graphe est composé de nœuds (ou sommets), et d'arcs (ou flèches) les reliant. Ces liens peuvent être orientés ou non (arêtes). L'internet, avec les ordinateurs et les câbles qui les relient deux à deux, est un graphe. Le web en est aussi un : les nœuds sont les pages web et les arcs sont les renvois hypertextuels qui les relient. Les liens qui pointent vers une page donnée sont appelés ses «degrés entrants», ceux qu'elle contient, à destination d'autres pages web, sont ses «degrés sortants».

On peut décrire le graphe du web avec chaque type de tels liens, ou avec les deux. Pour faire simple, le graphe que nous évoquerons sera celui construit à partir des degrés entrants, et sera considéré dans un premier temps comme non-orienté (cela réduit le web aux pages accessibles : les pages web vers lesquelles aucune autre page ne pointe «n'existent pas», sauf si on part d'elles). Pour construire le graphe du web, on part de quelques pages, et on suit les chemins qu'elles proposent. De proche en proche, on finit par obtenir un réseau qui, si on s'y prend bien, est d'un seul tenant.

De façon intuitive, comment évaluer l'«importance» d'un sommet dans un tel graphe ? On pense à l'équivalent d'une gare de triage, d'un nœud dont la suppression oblige à des longs détours, ou coupe le graphe en deux. C'est effectivement ainsi que l'«importance» d'une page va être formalisée. D'ores et déjà, les moteurs de recherche, les annuaires, et les pages de pointeurs sont «importants». En effet, de très nombreuses pages qui n'ont rien à voir ensemble proposent un renvoi vers les moteurs de recherche : ceux-ci composent donc le trait d'union entre des sous-graphes (on dit aussi *clusters*) distincts.

Le graphe du web a d'étranges particularités : il est à la fois dense (beaucoup de pages sont fortement interconnectées) et vide (vu dans son ensemble). Il est gigantesque (plusieurs milliards de pages) et pour les physiciens, il relève des «petits mondes». Par exemple, on passe d'une page à l'autre en une poignée d'étapes. Enfin, la statistique de ses arêtes renvoie à des lois dites «de puissance». Celles-ci rendent toute estimation (ou prévision) quasi impossible (variance théorique infinie). Toutefois, elles témoignent, sur le plan intellectuel, d'un double phénomène : d'une part, une uniformisation de la culture du fait qu'un très grand nombre de

4. Pour une première approche, cf. (Guichard, 2004).

pages renvoient vers les mêmes sites et que la quasi-totalité des pages sont très faiblement repérées (quelques liens vers elles) ; d'autre part, une mise en valeur de la singularité et de l'originalité du fait que, structurellement, les lois de puissance signifient une forte interdépendance de l'ensemble des objets qu'elles décrivent — que ceux-ci apparaissent fréquemment, rarement, ou entre les deux ⁵.

Tout cela fait qu'on ne sait pas très bien à quoi ressemble ce graphe, qu'on a bien du mal à le modéliser (et à estimer l'efficacité de ses modèles), à imaginer son évolution. D'autant que le web est aussi constitué de plus en plus de pages «dynamiques», produites à la volée en fonction de certaines requêtes, et donc non repérables. Les experts considèrent que les meilleurs moteurs de recherche ne recensent que 20% des pages web.

Là encore, nous sommes «déroutés». Nos références spatiales sont bien mises à mal : il n'y a plus de surface ni de volume, rien d'autre qu'un maillage complexe qu'on ne sait pas décrire.

3.2.2 Les moteurs de recherche

Moyennant des puissances de calcul colossales, et des réseaux à très hauts débits, on peut imaginer parcourir tout le web accessible et l'«aspérer» pour en faire la carte (des liens) et pour en indexer le contenu. C'est peu ou prou ce que font les moteurs de recherche. Mais, quand une requête arrive, quels résultats renvoient-ils, et dans quel ordre ? L'absence de classification du web, la non-intelligence des machines, et la faible structuration des contenus (métadonnées qui de plus peuvent être copiées et détournées) induisent une réaction des machines en deux temps.

D'abord, toutes les pages web contenant les mots de la requête sont sélectionnées ; ce qui renvoie avant tout à des méthodes lexicométriques, lesquelles peuvent sembler fort décevantes : par exemple, si on saisit la requête «emploi Paris», des millions de pages contiennent ces deux mots, et sont donc admissibles, indépendamment (ou presque) de leur thème ou de leur sens. C'est ensuite que ces pages sont triées en fonction de leur «importance» dans le graphe du web.

Il est intéressant de savoir comment les concepteurs des moteurs décrivent ce calcul d'importance : une fois obtenu le graphe du web, on attribue à chaque page le même fragment d'une quantité d'argent, que l'on fait circuler. Une page citée par d'autres reçoit la somme qu'elles ont accumulée (au prorata des pages qu'elles citent). À son tour, elle distribue son «capital» aux pages vers lesquelles elle pointe. Au bout de plusieurs itérations, le système se stabilise (la quantité d'argent est constante), et l'importance d'une page se mesure à son «capital», ou à la vitesse à laquelle elle reconstitue ce capital (puisque ce dernier ne cesse de circuler).

Cette représentation est critiquée car elle semble privilégier la popularité au détriment de la qualité. Pourtant, elle revient à celle de passage obligé (et mathématiquement, les deux appliquent un même théorème du point fixe). Elle est revendiquée par le moteur en position de monopole et, conformément au vocabulaire marchand qu'elle sollicite, favorise les médias, la consommation et l'actualité au détriment de l'érudition. Mais, d'une part, cet avantage est le fruit des usages, et non pas de l'algorithme, et d'autre part, l'économie symbolique du monde universitaire n'est pas pour autant complètement désorganisée par cette méthode. Deux phénomènes opposés en témoignent.

1. Usages : des requêtes simples

La majorité des requêtes des internautes se compose de quelques mots courants (ce qui ne signifie pas que ce type de question définisse une pratique «moyenne», puisque là

5. La notion de moyenne n'a pas de sens ; et, en l'occurrence, il est impossible d'associer les phénomènes majoritaires à une pratique normative, tout comme il est impossible de marginaliser

encore, on a affaire à des lois de puissance (Guichard, 2002). Or, les sites universitaires étant fort rares, rapportés à l'ensemble des autres, la probabilité pour qu'ils apparaissent comme des nœuds essentiels du graphe de la portion de web contenant ces mots communs est faible. À ce fait s'en ajoute un autre dicté aussi par l'usage : les propositions d'itinéraires que font les moteurs n'incitent pas particulièrement à explorer les pages qui ne sont pas parmi les premières qu'ils signalent. La raison n'en est pas une prétendue fainéantise des internautes, mais une réelle surcharge cognitive peu évoquée.

Ainsi les sites érudits restent-ils à l'écart de l'univers des questions communes.

2. Capital social et culturel

À l'inverse, il reste assez aisé, en usant de mots-clé spécifiques, de cerner les sites universitaires. Et il n'est pas rare que, entre un site inconnu et celui d'une institution de recherche, les professions intellectuelles choisissent le second, par un effet de confiance spontanée.

La notion d'importance renvoie alors à des pratiques sociales mieux cernées : un site «essentiel» est relaté par d'autres, qui sont souvent des portails spécialisés ou des pages de liens ; ces derniers sont déjà de meilleurs garants scientifiques que dans le cas général : ces intermédiaires semblent autant de «caisses d'épargne» locales qui assoient la notoriété d'une institution ou d'une personne. Ce qui, dans une certaine mesure, n'est pas foncièrement distinct de la fonction des comptes rendus d'ouvrages dans les revues savantes (d'autant que, pour le web universitaire comme pour l'imprimé, les lois lentes de la renommée s'appliquent : il faut du temps pour qu'un travail soit évoqué).

L'effet pervers est manifeste, quand certains décideurs de la recherche utilisent les moteurs de recherche comme d'un indicateur de la qualité scientifique d'une personne. Mais il n'est pas nécessairement plus biaisé que ceux de la «scientométrie» actuelle et les premiers à critiquer cet effet sont parfois ceux qui refusent d'appliquer à l'édition électronique les modalités de la concurrence qu'ils acceptent de l'imprimé. L'un des intérêts du web en tant que dispositif de publication est qu'il oblige aussi les institutions à «tenir leur rang», en apparaissant comme incontournables dans les domaines de recherche dans lesquelles elles prétendent exceller.

Il est d'autant plus facile d'être repéré sur le web que l'on est bien référencé, non par son moteur de recherche préféré, mais par d'autres éditeurs. Et la qualité rapporte, tout comme la singularité. Ainsi, pour l'éditeur en ligne, la géographie de l'internet passe avant tout par les chemins qui mènent à son site. Elle est culturelle et sociale, et de ce fait, reste très proche des formes classiques de la géographie du monde intellectuel.

3.2.3 Absence de boussole et fracture cognitive

La situation semble différente pour l'utilisateur, le consultant du web, qui, étonnamment, peut vite se perdre alors qu'il ne suit que des chemins bien balisés, de site en site, de lien en lien, à partir de moteurs de recherche, de portails, ou d'onglets : le web n'est jamais qu'une succession d'adresses, que seuls les spécialistes de la métrologie tentent de joindre de façon purement aléatoire.

Mais l'internaute est soumis à rude épreuve quand il recherche une information. Quand il s'adresse à un moteur de recherche, il lui faut choisir une série de mots-clé, les affiner, les traduire au besoin ; lire en diagonale les résultats, évaluer la notoriété institutionnelle de leurs

auteurs, multiplier les allers et retours entre les réponses. Même en cas de question unique et précise, il faut parfois plusieurs heures pour obtenir un résultat satisfaisant. On comprend que l'exploration d'un thème entier, par exemple pour réaliser un état de l'art, puisse prendre des mois de travail, et que soient alors sollicités des annuaires, des banques de données, payantes ou gratuites, ou des listes de pointeurs réalisées par des spécialistes. Et on imagine alors que la totalité de ce temps passé ou perdu finisse par réduire celui qui pourrait être passé dans une bibliothèque d'imprimés.

Ce qui manque le plus aux lecteurs de l'internet, même avertis, ce sont des cartes de l'intellect : comme au Moyen-Âge, ils ne disposent que d'itinéraires ou de guides. Des cartes mentales du web ont bien été proposées, mais elles ne sont pas encore très satisfaisantes sur le plan ergonomique.

Certes, il existe de multiples programmes pour assister les internautes en ligne, voire pour leur proposer un enseignement personnalisé, qu'il serait dangereux d'ignorer. Mais, sur le web, à chaque fois qu'une réelle difficulté apparaît, sa dimension cognitive est masquée au profit du déterminisme technique : l'apprentissage, l'art de se repérer dans la forêt documentaire, d'opérer des sélections et des synthèses sont trop souvent décrits comme des inconvénients passagers qui disparaîtront «naturellement» avec l'accession à des ordinateurs encore plus puissants, reliés par des réseaux plus rapides. C'est dans ce contexte que la «fracture numérique» est évoquée, voire prétendument mesurée⁶, alors que les personnes disposant d'une culture à la fois informatique et encyclopédique (on voudrait dire critique...) arrivent aisément à trouver réponse à leurs questions sur les réseaux avec de vieux ordinateurs et des débits fort réduits.

Toutefois, la foison des informations que l'on rencontre sur le web, et les formes de l'expérience (qui renvoient le plus souvent au capital social) rendent au final possible de tirer parti des réseaux et de leurs contenus. Mais, effet de la nouveauté du système technique ou, au contraire, preuve qu'il est profondément inscrit dans les dispositifs intellectuels conçus par les humains, il est pas garanti que l'internet, et principalement le web, soit aujourd'hui bien balisé pour la majorité des internautes.

Cette surcharge cognitive, ajoutée à l'opacité du fonctionnement des moteurs de recherche, ne finit-elle pas par signifier une écrasante domination intellectuelle, dont trop peu de personnes cherchent à préciser les contours et les enjeux ?

4 Le pouvoir des géomètres

La géographie de l'internet échappe peut-être définitivement à ses utilisateurs. La construction de celle-ci s'évalue tout d'abord au nombre de cartes de l'internet, qui sont le fait des métrologues, arpenteurs modernes qui se définissent comme mathématiciens, physiciens ou informaticiens. Ils produisent leurs cartes pour mieux comprendre les phénomènes complexes qu'ils analysent : flux de l'internet, congestions, typologie des usages, topologie des réseaux, etc. Ces cartes font donc partie de leur outillage mental, au même titre que les graphiques et schémas qui aident les physiciens et leurs collègues à asseoir un raisonnement, à produire une synthèse ou encore à préciser une hypothèse.

Reste qu'elles sont rarement ancrées dans la géographie terrestre, alors même que leurs auteurs ne dédaignent pas leurs aspects esthétiques. Le monde qu'ils décrivent n'a plus de frontières ni de pôles, et encore moins de fleuves. C'est un réseau, avec ses *clusters*, ou ses arbres,

6. Cf. (DOT Force, 2002) pour un tel discours et (Guichard, 2003) pour une critique.

hauts en couleurs. Et cette cartographie tout à fait spécifique finit par produire du territoire : on y découvre, présentés comme des continents, des déserts, des archipels, de nouveaux acteurs : sites web, propriétaires de réseaux, mais aussi «communautés» de pratiques, au point que la sociologie du web et de ses usages semble ne plus être l'affaire que des physiciens.

Pour finir, à voir la prolifération de ces cartes et leurs effets sur nos représentations du monde, on se demande quelle est la marge d'autonomie de cette production graphique *a priori* neutre. Ces cartographes des temps modernes n'obéissent-ils qu'à une logique purement technique ou scientifique, ou leur production est-elle le résultat de leur imaginaire ? Autrement dit, sauraient-ils dessiner autre chose que ce que leur culture leur dicte ? Une approche sociologique du monde scientifique et un regard curieux sur la production littéraire (science-fiction et bandes dessinées incluses) et cinématographique des dernières décennies, qui ne néglige pas non plus les formes urbanistiques de villes comme Los Angeles et Tokyo (et leur description au prisme des formes précédentes), invitent à ne pas rejeter la seconde hypothèse : la façon dont les scientifiques dessinent et cartographient l'internet doit peut-être beaucoup à la culture dans laquelle ils ont été immergés.

Ce faisant, l'internet favorise la diffusion de cette production géo-graphique. Ainsi se révèle une étroite dépendance entre science et culture, entre écriture et territoire.

Références

- CHEVILLOTTE, S. (2005). Les langages documentaires. *In Repère*. Enssib. <http://repere.enssib.fr/frontOffice/afficheArticle.asp?idTheme=12>.
- DESBOIS, H. (2001). Les territoires de l'internet : suggestions pour une cybergéographie. *In* GUICHARD, É., éditeur : *Comprendre les usages de l'Internet*, pages 253–263. Éditions Rue d'Ulm, Paris.
- DOT FORCE (2002). Les technologies de l'information au service de tous. Digital Opportunity Task Force : Rapport du Groupe d'experts sur l'accès aux nouvelles technologies. <http://www.dotforce.org/reports>.
- EDGERTON, D. (1998). De l'innovation aux usages. Dix thèses éclectiques sur l'histoire des techniques. *Annales Histoire, Sciences Sociales*, 4–5:815–837. Voir aussi *Des Sciences et des Techniques : un débat* (éd. de l'Ehess, 1998, pp. 259–287).
- GUICHARD, É. (2002). *L'internet : mesures des appropriations d'une technique intellectuelle*. Thèse de doctorat, École des hautes études en sciences sociales, Paris. <http://barthes.ens.fr/articles/theseEG>.
- GUICHARD, É. (2003). Does the 'Digital Divide' Exist? *In* van SETERS, P., de Gaay FORTMAN, B. et de RUIJTER, A., éditeurs : *Globalization and its new divides : malcontents, recipes, and reform*, pages 69–77. Dutch University Press, Amsterdam. Traduction française à l'URL : <http://barthes.enssib.fr/articles/Guichard-Digital-Divide.html>.
- GUICHARD, É., éditeur (2004). *Mesures de l'internet*. Les Canadiens en Europe, Paris.
- HERRENSCHMIDT, C. (1999). Écriture, monnaie, réseaux. inventions des anciens, inventions des modernes. *Le Débat*, 106:37–65. <http://barthes.enssib.fr/articles/Herrenschmidt-ecr-monnaie-reseaux.html>.
- SEARLE, J. (1998). *La construction de la réalité sociale*. Gallimard, Paris.

ZIVIANI, A. et AL. (2005). Vers la localisation géographique d'hôtes dans l'internet basée sur la multilatération. http://www-rp.lip6.fr/site_npa/site_rp/_publications/590-CFIP05.pdf.