

Comment les entreprises surveillent notre quotidien

Vous croyez tout savoir déjà sur l'exploitation de nos données personnelles ? Parcourez plutôt quelques paragraphes de ce très vaste dossier...

Il s'agit du remarquable travail d'enquête procuré [par Craked Labs](#), une organisation sans but lucratif qui se caractérise ainsi :

... un institut de recherche indépendant et un laboratoire de création basé à Vienne, en Autriche. Il étudie les impacts socioculturels des technologies de l'information et développe des innovations sociales dans le domaine de la culture numérique.

... Il a été créé en 2012 pour développer l'utilisation participative des technologies de l'information et de la communication, ainsi que le libre accès au savoir et à l'information – indépendamment des intérêts commerciaux ou gouvernementaux. Cracked Labs se compose d'un réseau interdisciplinaire et international d'experts dans les domaines de la science, de la théorie, de l'activisme, de la technologie, de l'art, du design et de l'éducation et coopère avec des parties publiques et privées.

Bien sûr, vous connaissez les GAFAM omniprésents aux avant-postes pour nous engluer au point que s'en déprendre complètement est difficile... Mais connaissez-vous Acxiom et LiveRamp, Equifax, Oracle, Experian et TransUnion ? Non ? Pourtant il y a des chances qu'ils nous connaissent bien...

Il existe une industrie très rentable et très performante des données « client ».

Dans ce long article documenté et qui déploie une vaste gamme

d'exemples dans tous les domaines, vous ferez connaissance avec les coulisses de cette industrie intrusive pour laquelle il semble presque impossible de « passer inaperçu », où notre personnalité devient un profil anonyme mais tellement riche de renseignements que nos nom et prénom n'ont aucun intérêt particulier.

L'article est long, vous pouvez préférer le lire à votre rythme en format .PDF (2,3 Mo)

[-> framablog.org-Comment-les-entreprises-surveillent-notre-quotidien-NEW](http://framablog.org-Comment-les-entreprises-surveillent-notre-quotidien-NEW)

L'équipe de Framalang s'est largement mobilisée pour vous procurer cette longue traduction : Abel, mo, Moutmout, Penguin, Opsylac, Luc, Lyn., hello, Jérochat, QS, Jérochat, Asta, Mannik, roptat, audionuma, Opsylac, Lumibd, linuxmario, goofy et un anonyme.

Des entreprises mettent notre quotidien sous surveillance

Source : <http://crackedlabs.org/en/corporate-surveillance>

Par Wolfie Christl

avec les contributions de : Katharina Kopp, Patrick Urs Riechert / Illustrations de Pascale Osterwalder.

Comment des milliers d'entreprises surveillent, analysent et influencent la vie de milliards de personnes. Quels sont les principaux acteurs du pistage numérique aujourd'hui ? Que peuvent-ils déduire de nos achats, de nos appels téléphoniques, de nos recherches sur le Web, de nos *Like* sur Facebook ? Comment les plateformes en ligne, les entreprises technologiques et les courtiers en données font-ils pour collecter, commercialiser et exploiter nos données personnelles ?

Ces dernières années, des entreprises dans de nombreux secteurs se sont mises à surveiller, pister et suivre les gens dans **pratiquement tous les aspects de leur vie**. Les comportements, les déplacements, les relations sociales, les centres d'intérêt, les faiblesses et les moments les plus intimes de milliards de personnes sont désormais continuellement enregistrés, évalués et analysés en temps réel. L'exploitation des données personnelles est devenue une industrie pesant plusieurs milliards de dollars. Pourtant, de ce pistage numérique omniprésent, on ne voit que **la partie émergée de l'iceberg** ; la majeure partie du processus se déroule dans les coulisses et reste opaque pour la plupart d'entre nous.

Ce rapport de Cracked Labs examine le fonctionnement interne et les pratiques en vigueur dans cette industrie des données personnelles. S'appuyant sur des années de recherche et sur un précédent rapport de 2016, l'enquête donne à voir la circulation cachée des données entre les entreprises. Elle cartographie la structure et l'étendue de **l'écosystème numérique de pistage et de profilage** et explore tout ce qui s'y rapporte : les technologies, les plateformes, les matériels ainsi que les dernières évolutions marquantes.

Le rapport complet (93 pages, en anglais) est disponible [en téléchargement au format PDF](#), et cette publication web en présente un résumé **en dix parties**.

Sommaire

- [I. Analyser les individus](#)
- [II. Analyser les individus dans la finance, les assurances et la santé](#)
- [III. Collecte et utilisation massives de données client](#)
- [IV. Les courtiers en données/ et le marché des données personnelles](#)
- [V. La surveillance en temps réel des comportements quotidiens](#)

- [VI. Relier, faire correspondre et combiner des profils numériques](#)
- [VII. Gérer les clients et les comportements : personnalisation et contrôle](#)
- [VIII. Dans les mailles du filet : vie quotidienne, données commerciales et analyse du risque](#)
- [IX. Cartographie de l'écosystème du pistage et du profilage commercial](#)
- [X. Vers une société du contrôle social numérique généralisé ?](#)

En 2007, Apple a lancé le smartphone, Facebook a atteint les 30 millions d'utilisateurs, et des entreprises de publicité en ligne ont commencé à cibler les internautes en se basant sur des données relatives à leurs préférences individuelles et leurs centres d'intérêt. Dix ans plus tard, **un large ensemble d'entreprises dont le cœur de métier est les données (les data-companies ou entreprises de données en français) a émergé**, on y trouve de très gros acteurs comme Facebook ou Google mais aussi des milliers d'autres entreprises, qui sans cesse, **se partagent et se vendent les unes aux autres des profils numériques**. Certaines entreprises ont commencé à combiner et à relier des données du web et des smartphones avec les données clients et les informations hors-ligne qu'elles avaient accumulées pendant des décennies.

La machine omniprésente de surveillance en temps réel qui a été développée pour la publicité en ligne s'étend rapidement à d'autres domaines, **de la tarification à la communication politique en passant par le calcul de solvabilité et la gestion des risques**. Des plateformes en ligne énormes, des entreprises de publicité numérique, des courtiers en données et des entreprises de divers secteurs peuvent maintenant identifier, trier, catégoriser, analyser, évaluer et classer les utilisateurs via les plateformes et les matériels. **Chaque clic sur un site web** et chaque mouvement du doigt sur un smartphone peut activer un large éventail de mécanismes de

partage de données distribuées entre plusieurs entreprises, ce qui, en définitive, **affecte directement les choix offerts aux gens**. Le pistage numérique et le profilage, en plus de la personnalisation ne sont pas seulement utilisés pour surveiller, mais aussi pour influencer les comportements des personnes.

Vous devez vous battre pour votre vie privée, sinon vous la perdrez.

Eric Schmidt, Google/Alphabet, 2013

Analyser les individus

Des études scientifiques démontrent que de nombreux aspects de la personnalité des individus peuvent être déduits des données générées par des recherches sur Internet, des historiques de navigation, des comportements lors du visionnage d'une vidéo, des activités sur les médias sociaux ou des achats. Par exemple, des **données personnelles sensibles** telles que l'origine ethnique, les convictions religieuses ou politiques, la situation amoureuse, l'orientation sexuelle, ou l'usage d'alcool, de cigarettes ou de drogues peuvent être assez précisément déduites des **Like sur Facebook d'une personne**. L'analyse des profils de réseaux sociaux peut aussi prédire des traits de personnalité comme la stabilité émotionnelle, la satisfaction individuelle, l'impulsivité, la dépression et l'intérêt pour le sensationnel.

Analyser les Like Facebook, les données des téléphones, et les styles de frappe au clavier



Pour plus de détails, se référer à Christl and Spiekermann 2016 (p. 14-20). Sources : Kosinski et al 2013, Chittaranjan et al 2011, Epp et al 2011.

De la même façon, il est possible de déduire certains traits de caractères d'une personne à partir de données sur les sites Web qu'elle a visités, sur les appels téléphoniques qu'elle a passés, et sur les applis qu'elle a utilisées. **L'historique de navigation** peut donner des informations sur la profession et le niveau d'étude. Des chercheurs canadiens ont même réussi à évaluer des états émotionnels comme la confiance, la nervosité, la tristesse ou la fatigue en analysant la **façon dont on tape sur le clavier de l'ordinateur**.

Analyser les individus dans la finance, les assurances et la santé

Les résultats des méthodes actuelles d'extraction et d'analyse des données reposent sur des **corrélations statistiques** avec un certain niveau de probabilité. Bien qu'ils soient significativement plus fiables que le hasard dans la prédiction des caractéristiques ou des traits de caractère d'un individu, ils ne sont évidemment pas toujours exacts. Néanmoins, ces méthodes sont déjà mises en œuvre pour **trier, catégoriser, étiqueter, évaluer, noter et classer les personnes**, non seulement dans une approche marketing mais aussi pour prendre des décisions dans des domaines riches en conséquence comme la finance, l'assurance, la santé, pour ne citer qu'eux.

L'évaluation de crédit basée sur les données de

comportement numérique

Des startups comme Lenddo, Kreditech, Cignifi et ZestFinance utilisent déjà les données récoltées sur les réseaux sociaux, lors de recherches sur le web ou sur les téléphones portables pour **calculer la solvabilité d'une personne** sans même utiliser de données financières. D'autres se basent sur la façon dont quelqu'un va remplir un formulaire en ligne ou naviguer sur un site web, sur la grammaire et la ponctuation de ses textos, ou sur l'état de la batterie de son téléphone. Certaines entreprises incluent même des données sur les amis avec lesquels une personne est connectée sur un réseau social pour évaluer sa solvabilité.

Cignifi, qui calcule la solvabilité des clients en fonction **des horaires et de la fréquence des appels téléphoniques**, se présente comme « la plateforme ultime de monétisation des données pour les opérateurs de réseaux mobiles ». De grandes entreprises, notamment MasterCard, le fournisseur d'accès mobile Telefonica, les agences d'évaluation de solvabilité Experian et Equifax, ainsi que le géant chinois de la recherche web Baidu, ont commencé à nouer des partenariats avec des startups de ce genre. L'application à plus grande échelle de services de cette nature est particulièrement en croissance dans les pays du Sud, ainsi qu'auprès de groupes de population vulnérables dans d'autres régions.

Réciproquement, les données de crédit nourrissent le marketing en ligne. Sur Twitter, par exemple, les annonceurs peuvent cibler leurs publicités en fonction de la **solvabilité supposée des utilisateurs de Twitter** sur la base des données client fournies par le courtier en données Oracle. Allant encore plus loin dans cette logique, **Facebook** a déposé un brevet pour une évaluation de crédit basée sur la cote de solvabilité de vos amis sur un réseau social. Personne ne sait s'ils ont l'intention de réellement mettre en application cette intégration totale des réseaux sociaux, du marketing et de l'évaluation des risques.

On peut dire que toutes les données sont des données sur le crédit, mais il manque encore la façon de les utiliser.

Douglas Merrill, fondateur de ZestFinance et ancien directeur des systèmes d'informations chez Google, 2012

Prédire l'état de santé à partir des données client

Les entreprises de données et les assureurs travaillent sur des programmes qui utilisent les informations sur la vie quotidienne des consommateurs pour prédire leurs risques de santé. Par exemple, l'assureur **Aviva**, en coopération avec la société de conseil Deloitte, a utilisé des données clients achetées à un courtier en données et habituellement utilisées pour le marketing, pour prédire les **risques de santé individuels** (comme le diabète, le cancer, l'hypertension et la dépression) de 60 000 personnes souhaitant souscrire une assurance.

La société de conseil McKinsey a aidé à **prédire les coûts hospitaliers** de patients en se basant sur les données clients d'une « grande compagnie d'assurance » santé américaine. En utilisant les informations concernant la démographie, la structure familiale, les achats, la possession d'une voiture et d'autres données, McKinsey a déclaré que ces « renseignements peuvent aider à identifier des sous-groupes stratégiques de patients avant que des périodes de coûts élevés ne surviennent ».

L'entreprise d'analyse santé **GNS Healthcare** a aussi calculé les risques individuels de santé de patients à partir d'un large champ de données tel que la génétique, les dossiers médicaux, les analyses de laboratoire, les appareils de santé mobiles et le comportement du consommateur. Les sociétés partenaires des assureurs tels que Aetna donnent une note qui identifie « les personnes susceptibles de subir une opération » et proposent de **prédire l'évolution de la maladie**

et les résultats des interventions. D'après un rapport sectoriel, l'entreprise « classe les patients suivant le retour sur investissement » que l'assureur peut espérer s'il les cible pour des interventions particulières.

LexisNexis Risk Solutions, à la fois, un important courtier en données et une société d'analyse de risque, fournit un produit d'évaluation de santé qui calcule les risques médicaux ainsi que les frais de santé attendus individuellement, en se basant sur une importante quantité de données consommateurs, incluant les achats.

Collecte et utilisation massives de données client

Les plus importantes plates-formes connectées d'aujourd'hui, **Google et Facebook** en premier lieu, ont des informations détaillées sur la vie quotidienne de milliards de personnes dans le monde. Ils sont les plus visibles, les plus envahissants et, hormis les entreprises de renseignement, les publicitaires en ligne et les services de détection des fraudes numériques, peut-être les acteurs les plus avancés de **l'industrie de l'analyse et des données personnelles**. Beaucoup d'autres agissent en coulisse et hors de vue du public.

Le cœur de métier de la **publicité en ligne** consiste en un écosystème de milliers d'entreprises concentrées sur la traque constante et le profilage de milliards de personnes. À chaque fois qu'une publicité est affichée sur un site web ou une application mobile, un profil d'utilisateur vient juste d'être vendu au plus gros enchérisseur dans les millisecondes précédentes. Contrairement à ces nouvelles pratiques, les **agences d'analyse de solvabilité** et les **courtiers en données clients** exploitent des données personnelles depuis des décennies. Ces dernières années, ils ont commencé à combiner les très nombreuses données dont ils disposent sur la vie hors-ligne des personnes avec les bases de données

utilisateurs et clients utilisées par de grandes plateformes, par des entreprises de publicité et par une multitude d'autres entreprises dans de nombreuses secteurs.

Les entreprises de données ont des informations détaillées sur des milliards de personnes

Plateformes en ligne grand public

Facebook dispose des profils de	1,9 milliards d'utilisateurs de Facebook 1,2 milliards d'utilisateurs de Whatsapp 600 millions d'utilisateurs d'Instagram
Google dispose des profils de	2 milliards d'utilisateurs d'Android + d'un milliard d'utilisateurs de Gmail + d'un milliard d'utilisateurs de Youtube
Apple dispose des profils de	1 milliard d'utilisateurs d'iOS

Sociétés d'analyse de la solvabilité

Experian	dispose des données de solvabilité de 918 millions de personnes dispose des données marketing de 700 millions de personnes a un "aperçu" sur 2,3 milliards de personnes
Equifax	dispose des données de 820 millions de personnes et d'1 milliard d'appareils

TransUnion	dispose des données d'1 milliard de personnes
------------	---

Courtiers en données clients

Acxiom	dispose des données de	700 millions de personnes
		1 milliard de cookies et d'appareils mobiles
		3,7 milliards de profils clients
Oracle	dispose des données de	1 milliard d'utilisateurs d'appareils mobiles
		1,7 milliards d'internautes
	donne accès à	5 milliards d'identifiants uniques client

Facebook utilise au moins 52 000 caractéristiques personnelles pour trier et classer ses 1,9 milliard d'utilisateurs suivant, par exemple, leur orientation politique, leur origine ethnique et leurs revenus. Pour ce faire, la plateforme analyse leurs messages, leurs *Likes*, leurs partages, leurs amis, leurs photos, leurs mouvements et beaucoup d'autres comportements. De plus, Facebook **acquiert à d'autres entreprises des données sur ses utilisateurs**. En 2013, la plateforme démarre son partenariat avec les quatre courtiers en données Acxiom, Epsilon, Datalogix et BlueKai, les deux derniers ont ensuite été rachetés par le géant de l'informatique Oracle. Ces

sociétés aident Facebook à **pister et profiler** ses utilisateurs bien mieux qu'il le faisait déjà en lui fournissant des données collectées en dehors de sa plateforme.

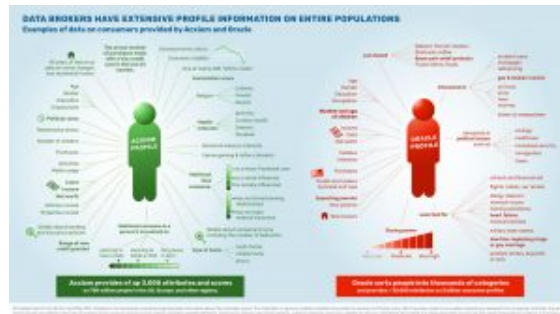
Les courtiers en données et le marché des données personnelles

Les courtiers en données client ont un rôle clé dans le marché des données personnelles actuel. Ils agrègent, combinent et échangent des quantités astronomiques d'informations sur des populations entières, collectées depuis des sources en ligne et hors-ligne. Les courtiers en données collectent de l'information disponible publiquement et achètent le droit d'utiliser les données clients d'autres entreprises. Leurs données proviennent en général de **sources qui ne sont pas les individus eux-mêmes**, et sont collectées en grande partie sans que le consommateur soit au courant. Ils analysent les données, en font des déductions, construisent des catégories de personnes et fournissent à leurs clients des informations sur des milliers de caractéristiques par individu.

Dans les profils individuels créés par les courtiers en données, on trouve non seulement des informations à propos de l'éducation, de l'emploi, des enfants, de la religion, de l'origine ethnique, de la position politique, des loisirs, des centres d'intérêts et de l'usage des médias, mais aussi à propos du comportement en ligne, par exemple les recherches sur Internet. Sont également collectées les données sur les achats, l'usage de carte bancaire, le revenu et l'endettement, la gestion bancaire et les polices d'assurance, la propriété immobilière et automobile, et tout un tas d'autres types d'information. Les courtiers en données calculent et **attribuent aussi des notes aux individus afin de prédire leur comportement futur**, par exemple en termes de stabilité économique, de projet de grossesse ou de changement d'emploi.

Quelques exemples de données clients fournies par Acxiom et

Oracle



exemples de données clients
fournies par Acxiom et
Oracle (en avril/mai 2017)
– sources : voir le rapport

Acxiom, un important courtier en données

Fondée en 1969, Acxiom gère l'une des plus grandes bases de données client commerciales au monde. Disposant de milliers de sources, l'entreprise fournit jusqu'à 3000 types de données sur 700 millions de personnes réparties dans de nombreux pays, dont les États-Unis, le Royaume-Uni et l'Allemagne. Née sous la forme d'une entreprise de marketing direct, Acxiom a développé ses bases de données client centralisées à la fin des années 1990.

À l'aide de son système **Abilitek Link**, l'entreprise tient à jour une sorte de registre de la population dans lequel chaque personne, chaque foyer et chaque bâtiment reçoit un identifiant unique. En permanence, l'entreprise met à jour ses bases de données sur la base d'informations concernant **les naissances et les décès, les mariages et les divorces, les changements de nom ou d'adresse** et aussi bien sûr de nombreuses autres données de profil. Quand on lui demande des renseignements sur une personne, Acxiom peut par exemple donner une appartenance religieuse parmi l'une des 13 retenues comme « catholique », « juif », ou « musulman » et une appartenance ethnique sur quasiment 200 possibles.

Acxiom commercialise l'accès aux profils détaillés des consommateurs et aide ses clients à trouver, cibler, identifier, analyser, trier, noter et classer les gens. L'entreprise gère aussi directement pour ses propres clients **15 000 bases de données clients** représentant des milliards de profils consommateurs. Les clients d'Acxiom sont des grandes banques, des assureurs, des services de santé et des organismes gouvernementaux. En plus de son activité de commercialisation de données, Acxiom fournit également des services de vérification d'identité, de gestion du risque et de détection de fraude.

Acxiom et ses fournisseurs de données, ses partenaires et ses services



Acxiom et ses fournisseurs de données, ses partenaires et ses clients (en avril/mai 2017) – sources : voir le rapport

Depuis l'acquisition en 2014 de la société de données en ligne **LiveRamp**, Acxiom a déployé d'importants efforts pour connecter son dépôt de données – couvrant une dizaine d'années – au monde numérique. Par exemple, Acxiom était parmi les premiers courtiers en données à fournir de l'information additionnelle à Facebook, Google et Twitter afin d'aider ces plateformes à mieux pister ou catégoriser les utilisateurs en fonction de leurs achats mais aussi en fonction d'autres comportements

qu'ils ne savaient pas encore eux-mêmes pister.

LiveRamp de Acxiom **connecte et combine les profils numériques** issus de centaines d'entreprises de données et de publicité. Au centre se trouve son système IdentityLink, qui aide à reconnaître les individus et à relier les informations les concernant, dans les bases de données, les plateformes et les appareils en se basant sur leur adresse de courriel, leur numéro de téléphone, l'identifiant de leur téléphone, ou d'autres identifiants. Bien que l'entreprise assure que les correspondances et les associations se fassent de manière « anonyme » et « dé-identifiée », elle dit aussi pouvoir « connecter des données hors-ligne et en ligne sur un seul identifiant ».

Parmi les entreprises qui ont récemment été reconnues comme étant des fournisseurs de données par LiveRamp, on trouve les géants de l'analyse de solvabilité **Equifax, Experian et TransUnion**. De plus, de nombreux services de pistage numérique collectant des données par Internet, par les applications mobiles, et même par **des capteurs placés dans le monde réel**, fournissent des données à LiveRamp. Certains d'entre eux utilisent les base de données de LiveRamp, qui permettent aux entreprises « d'acheter et de vendre des données client précieuses ». D'autres fournissent des données afin que Acxiom et LiveRamp puissent reconnaître des individus et relier les informations enregistrées avec les profils numériques d'autres provenances. Mais le plus préoccupant, c'est sans doute le partenariat entre Acxiom et Crossix, une entreprise avec **des données détaillées sur la santé** de 250 millions de consommateurs américains. Crossix figure parmi les fournisseurs de données de LiveRamp.

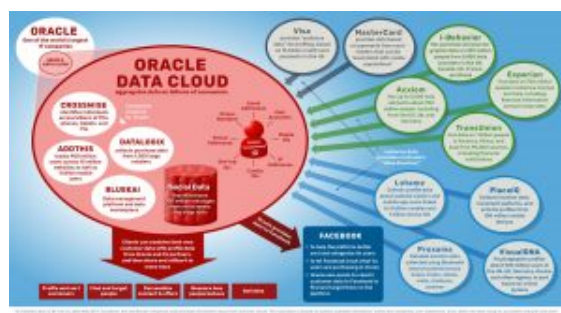
Quiconque enregistrant des données sur les consommateurs peut potentiellement être un fournisseur de données. »

Travis May, Directeur général de Acxiom-LiveRamp

Oracle, un géant des technologies de l'information pénètre le marché des données client

En faisant l'acquisition de plusieurs entreprises de données telles que Datalogix, BlueKai, AddThis et CrossWise, Oracle, un des premiers fournisseurs de logiciels d'entreprises et de bases de données dans le monde, est également récemment devenu **un des premiers courtiers en données clients**. Dans son « cloud », Oracle rassemble 3 milliards de profils utilisateurs issus de 15 millions de sites différents, les données d'un milliard d'utilisateurs mobiles, des milliards d'historiques d'achats dans des chaînes de supermarchés et 1500 détaillants, ainsi que 700 millions de messages par jour issus des réseaux sociaux, des blogs et des sites d'avis de consommateurs.

Oracle rassemble des données sur des milliards de consommateurs



Oracle et ses fournisseurs de données, ses partenaires et ses clients (en avril/mai 2017) – sources : voir le rapport

Oracle catalogue près de 100 fournisseurs de données dans son répertoire de données, parmi lesquels figurent Acxiom et des agences d'analyse de solvabilité telles que **Experian et TransUnion**, ainsi que des entreprises qui tracent les visites de sites Internet, l'utilisation d'applications mobiles et les

déplacements, ou qui collectent des données à partir de questionnaires en ligne. **Visa et MasterCard** sont également référencés comme fournisseurs de données. En coopération avec ses partenaires, Oracle fournit plus de 30 000 catégories de données différentes qui peuvent être attribuées aux consommateurs. Réciproquement, l'entreprise partage des données avec Facebook et aide **Twitter** à calculer la solvabilité de ses utilisateurs.

Le **Grphe d'Identifiants** Oracle détermine et combine des profils utilisateur provenant de différentes entreprises. Il est le « trait d'union entre les interactions » à travers les différentes bases de données, services et appareils afin de « créer un profil client adressable » et « d'identifier partout les clients et les prospects ». D'autres entreprises peuvent envoyer à Oracle, des clés de correspondance construites à partir **d'adresses courriel, de numéros de téléphone, d'adresse postale ou d'autres identifiants**, Oracle les synchronisera ensuite à son « réseau d'identifiants utilisateurs et statistiques, connectés ensemble dans le Grphe d'Identifiants Oracle ». Bien que l'entreprise promette de n'utiliser que des identifiants utilisateurs anonymisés et des profils d'utilisateurs anonymisés, ceux-ci font tout de même référence à certains individus et peuvent être utilisés pour les reconnaître et les cibler dans de nombreux contextes de la vie.

Le plus souvent, les clients d'Oracle peuvent **télécharger dans le « cloud » d'Oracle leurs propres données concernant : leurs clients, les visites sur leur site ou les utilisateurs d'une application** ; ils peuvent les combiner avec des données issues de nombreuses autres entreprises, puis les transférer et les utiliser en temps réel sur des centaines d'autres plateformes de commerce et de publicité. Ils peuvent par exemple les utiliser pour **trouver et cibler** des personnes sur tous les appareils et plateformes, **personnaliser leurs interactions**, et le cas échéant **mesurer** la réaction des clients qui ont été

personnellement ciblés.

La surveillance en temps réel des comportements quotidiens

Les plateformes en ligne, les fournisseurs de technologies publicitaires, les courtiers en données, et les négociants de toutes sortes d'industries peuvent maintenant surveiller, reconnaître et analyser des individus dans de nombreuses situations. Ils peuvent étudier ce qui intéresse les gens, ce qu'ils ont fait aujourd'hui, ce qu'ils vont sûrement faire demain, et leur valeur en tant que client.

Les données concernant les vies en ligne et hors ligne des personnes

Une large spectre d'entreprises collecte des informations sur les personnes depuis des décennies. Avant l'existence d'Internet, les agences de crédit et les agences de marketing direct servaient de point d'intégration principal entre les données provenant de différentes sources. Une première étape importante dans **la surveillance systématique des consommateurs** s'est produite **dans les années 1990**, par la commercialisation de bases de données, les programmes de fidélité et l'analyse poussée de solvabilité. Après l'essor d'Internet et de la publicité en ligne **au début des années 2000**, et la montée des réseaux sociaux, des smartphones et de la publicité en ligne **à la fin des années 2000**, on voit maintenant **dans les années 2010** l'industrie des données clients s'intégrer avec le nouvel écosystème de pistage et de profilage numérique.

Cartographie de la collecte de données clients



Différents niveaux, domaines et sources de collecte de données clients par les entreprises

De longue date, les courtiers en données clients et d'autres entreprises acquièrent des informations sur **les abonnés à des journaux et à des magazines**, sur les membres de clubs de lecture et de ciné-clubs, sur les acheteurs de catalogues de vente par correspondance, sur les personnes réservant dans les agences de voyage, sur les participants à des séminaires et à des conférences, et sur les consommateurs qui remplissent les cartes de garantie pour leurs achats. La collecte de données d'achats grâce à des **programmes de fidélité** est, de ce point de vue, une pratique établie depuis longtemps.

En complément des données provenant directement des individus, sont utilisées, par exemple les informations concernant **le type quartiers et d'immeubles** où résident les personnes afin de décrire, étiqueter, trier et catégoriser ces personnes. De même, les entreprises utilisent maintenant des profils de consommateurs s'appuyant sur les métadonnées concernant **le type de sites Internet fréquentés**, les vidéos regardées, les applications utilisées et les zones géographiques visitées. Au cours de ces dernières années, l'échelle et le niveau de détail des flux de données comportementales générées par toutes sortes d'activités du quotidien, telles que l'utilisation d'Internet, des réseaux sociaux et des équipements, ont rapidement augmenté.

Ce n'est pas un téléphone, c'est mon mouchard /pisteur/. New York Times, 2012

Un pistage et un profilage omniprésents

Une des principales raisons pour lesquelles le pistage et le profilage commerciaux sont devenus si généralisés c'est que quasiment tous les sites Internet, les fournisseurs d'applications mobiles, ainsi que de nombreux vendeurs d'équipements, partagent activement des données comportementales avec d'autres entreprises.

Il y a quelques années, la plupart des sites Internet ont commencé à inclure dans leur propre site **des services de pistage qui transmettent des données à des tiers**. Certains de ces services fournissent des fonctions visibles aux utilisateurs. Par exemple, lorsqu'un site Internet montre un bouton Facebook « j'aime » ou une vidéo YouTube encapsulée, des données utilisateur sont transmises à Facebook ou à Google. En revanche, de nombreux autres services ayant trait à la publicité en ligne demeurent cachés et, pour la plupart, ont pour seul objectif de collecter des données utilisateur. **Le type précis de données utilisateur** partagées par les éditeurs numériques et la façon dont les tierces parties utilisent ces données reste largement méconnus. Une partie de ces activités de pistage peut être analysée par n'importe qui ; par exemple en installant l'extension pour navigateur Lightbeam, il est possible de visualiser le réseau invisible des trackers des parties tierces.

Une étude récente a examiné un million de sites Internet différents et a trouvé plus de **80 000 services tiers** recevant des données concernant les visiteurs de ces sites. Environ 120 de ces services de pistage ont été trouvés sur plus de 10 000 sites, et six entreprises surveillent les utilisateurs sur plus de 100 000 sites, dont Google, Facebook, Twitter et BlueKai d'Oracle. Une étude sur 200 000 utilisateurs allemands

visitant 21 millions de pages Internet a montré que les trackers tiers étaient présents **sur 95 %** des pages visitées. De même, la plupart des applications mobiles partagent des informations sur leurs utilisateurs avec d'autres entreprises. Une étude menée en 2015 sur les applications à la mode en Australie, en Allemagne et aux États-Unis a trouvé qu'**entre 85 et 95 % des applications gratuites**, et même **60 % des applications payantes** se connectaient à des tierces parties recueillant des données personnelles.

Une **carte interactive des services cachés de pistage tiers** sur les applications Android créée par des chercheurs européens et américains peut être explorée à l'adresse suivante : haystack.mobi/panopticon



Copie d'écran du ISCI
Haystack Panopticon
disponible sur
haystack.mobi/panopticon, ©
mis à disposition
gracieusement par ISCI,
Université de Berkeley

En matière d'appareils, ce sont peut-être **les smartphones** qui actuellement contribuent le plus au recueil omniprésent données. L'information enregistrée par les téléphones portables fournit un aperçu détaillé de la personnalité et de la vie quotidienne d'un utilisateur. Puisque les consommateurs ont en général besoin d'un **compte Google, Apple ou Microsoft** pour les utiliser, une grande partie de l'information est déjà

reliée à l'identifiant d'une des principales plateformes.

La vente de données utilisateurs ne se limite pas aux éditeurs de sites Internet et d'applications mobiles. Par exemple, l'entreprise d'intelligence commerciale SimilarWeb reçoit des données issues non seulement de centaines de milliers de sources de mesures directes depuis les sites et les applications, mais aussi des logiciels de bureau et des extensions de navigateur. Au cours des dernières années, de nombreux autres appareils avec des capteurs et des connexions réseau ont intégré la vie de tous les jours, **cela va des liseuses électroniques et autres accessoires connectés aux téléphones intelligents, compteurs, thermostats, détecteurs de fumée, imprimantes, réfrigérateurs, brosses à dents, jouets et voitures**. À l'instar des smartphones, ces appareils donnent aux entreprises un accès sans précédent au comportement des consommateurs dans divers contextes de leur vie.

Publicité programmatique et technologie marketing

La plus grande partie de la publicité numérique prend aujourd'hui la forme d'enchères en temps réel hautement automatisées entre les éditeurs et les publicitaires ; on appelle cela la **publicité programmatique**. Lorsqu'une personne se rend sur un site Internet, les données utilisateur sont envoyées à une kyrielle de services tiers, qui cherchent ensuite à reconnaître la personne et extraire l'information disponible sur le profil. Les publicitaires souhaitant livrer une publicité à cet individu, en particulier du fait de certains attributs ou comportements, placent une enchère. En quelques millisecondes, le publicitaire le plus offrant gagne et place la pub. Les publicitaires peuvent de la même façon **enchérir sur les profils utilisateurs** et le placement de publicités au sein des applications mobiles.

Néanmoins, ce processus ne se déroule pas, la plupart du temps, entre **les éditeurs et les publicitaires**. L'écosystème est constitué d'une pléthore de toutes sortes de données

différentes et de fournisseurs de technologies en interaction les uns avec les autres, parmi lesquels des réseaux publicitaires, des marchés publicitaires, des plateformes côté vente et des plateformes côté achat. Certains se spécialisent dans le pistage et la publicité suivant les résultats de recherche, dans la publicité généraliste sur Internet, dans la pub sur mobile, dans les pubs vidéos, dans les pubs sur les réseaux sociaux, ou dans les pubs au sein des jeux. D'autres se concentrent sur l'approvisionnement en données, en analyse ou en services de personnalisation.

Pour tracer le portrait des utilisateurs d'Internet et d'applications mobiles, toutes les parties impliquées ont développé des méthodes sophistiquées pour accumuler, regrouper et relier les informations provenant de différentes entreprises afin **de suivre les individus dans tous les aspects de leur vie**. Nombre d'entre elles recueillent et utilisent des profils numériques sur des centaines de millions de consommateurs, leurs navigateurs Internet et leurs appareils.

De nombreux secteurs rejoignent l'économie de pistage

Au cours de ces dernières années, des entreprises dans plusieurs secteurs ont commencé à partager et à utiliser à très grande échelle des données concernant leurs utilisateurs et clients.

La plupart des **détaillants** vendent des formes agrégées de **données sur les habitudes d'achat** auprès des entreprises d'études de marchés et des courtiers en données. Par exemple, l'entreprise de données IRI accède aux données de plus de 85 000 magasins ('alimentation, grande distribution, médicaments, d'alcool et d'animaux de compagnie, magasin à prix unique et magasin de proximité). Nielsen déclare recueillir les informations concernant les ventes de 900 000 magasins dans le monde dans plus de 100 pays. L'enseigne de grande distribution britannique **Tesco** sous-traite son

programme de fidélité et ses activités en matière de données auprès d'une filiale, Dunhumby, dont le slogan est « transformer les données consommateur en régal pour le consommateur ». Lorsque Dunhumby a fait l'acquisition de l'entreprise technologique de publicité allemande Sociomantic, il a été annoncé que Dunhumby « conjuguerait ses connaissances étendues au sujet sur les préférences d'achat de 400 millions de consommateurs » avec les « données en temps réel de plus de 700 millions de consommateurs en ligne » de Sociomantic afin de personnaliser et d'évaluer les publicités.

Cartographie de l'écosystème du pistage et du profilage commercial



Aujourd'hui de nombreux industriels dans divers secteurs ont rejoint l'écosystème de pistage et de profilage numérique, aux côtés des grandes plateformes en ligne et des professionnels de l'analyse des données clients.

De grands **groupes médiatiques** sont aussi fortement intégrés dans l'écosystème de pistage et de profilage numérique actuel. Par exemple, Time Inc. a fait l'acquisition d'Adelphic, une importante société de pistage et de technologies publicitaires multi-support, mais aussi de Viant, une entreprise qui déclare

avoir accès à plus de 1,2 milliard d'utilisateurs enregistrés. La plateforme de streaming **Spotify** est un exemple célèbre d'éditeur numérique qui vend les données de ses utilisateurs. Depuis 2016, la société partage avec le département données du géant du marketing WPP des informations à propos de ce que les utilisateurs écoutent, sur leur humeur ainsi que sur leur comportement et leur activité en termes de playlist. WPP a maintenant accès « aux préférences et comportements musicaux des 100 millions d'utilisateurs de Spotify ».

De nombreuses grandes **entreprises de télécom et de fournisseurs d'accès Internet** ont fait l'acquisition d'entreprises de technologies publicitaires et de données. Par exemple, Millennial Media, une filiale d'**AOL-Verizon**, est une plateforme de publicité mobile qui collecte les données de plus de 65 000 applications de différents développeurs, et prétend avoir accès à environ 1 milliard d'utilisateurs actifs distincts dans le monde. **Singtel**, l'entreprise de télécoms basée à Singapour, a acheté Turn, une plateforme de technologies publicitaires qui donne accès aux distributeurs à 4,3 milliards d'appareils pouvant être ciblés et d'identifiants de navigateurs et à 90 000 attributs démographiques, comportementaux et psychologiques.

Comme les compagnies aériennes, les hôtels, les commerces de détail et les entreprises de beaucoup d'autres secteurs, le secteur des **services financiers** a commencé à agréger et utiliser des données clients supplémentaires grâce à des programmes de fidélité dans les années 80 et 90. Les entreprises dont la clientèle cible est proche et complémentaires partagent depuis longtemps certaines de leurs données clients entre elles, un processus souvent géré par des intermédiaires. Aujourd'hui, l'un de ces intermédiaires est Cardlytics, une entreprise qui gère des programmes de fidélité pour plus de 1 500 institutions financières, telles que **Bank of America** et MasterCard. Cardlytics s'engage auprès des institutions financières à « générer des nouvelles sources de

revenus en exploitant le pouvoir de [leurs] historiques d'achat ». L'entreprise travaille aussi en partenariat avec LiveRamp, la filiale d'Acxiom qui combine les données en ligne et hors ligne des consommateurs.

Pour **MasterCard**, la vente de produits et de services issus de l'analyse de données pourrait même devenir son **cœur de métier**, sachant que la production d'informations, dont la vente de données, représentent une part considérable et croissante de ses revenus. **Google** a récemment déclaré qu'il capture environ 70 % des transactions par carte de crédit aux États-Unis via « partenariats tiers » afin de tracer les achats, mais n'a pas révélé ses sources.

Ce sont vos données. Vous avez le droit de les contrôler, de les partager et de les utiliser comme bon vous semble.

C'est ainsi que le courtier en données Lotame s'adresse sur son site Internet à ses entreprises clientes en 2016.

Relier, faire correspondre et combiner des profils numériques

Jusqu'à récemment, les publicitaires, sur Facebook, Google ou d'autres réseaux de publicité en ligne, ne pouvaient cibler les individus qu'en analysant leur comportement en ligne. Mais depuis quelques années, grâce aux moyens offerts par les entreprises de données, les profils numériques issus de différentes plateformes, de différentes bases de données clients et du monde de la publicité en ligne peuvent désormais être associés et combinés entre eux.

Connecter les identités en ligne et hors ligne

Cela a commencé en 2012, quand Facebook a permis aux entreprises de télécharger leurs propres **listes d'adresses de courriel et de numéros de téléphone** sur la plateforme. Bien que les adresses et numéros de téléphone soient convertis en

pseudonyme, Facebook est en mesure de relier directement ces données client provenant d'entreprises tierces avec ses propres comptes utilisateur. Cela permet par exemple aux entreprises de **trouver et de cibler très précisément sur Facebook** les personnes dont elles possèdent les adresses de courriel ou les numéros de téléphone. De la même façon, il leur est éventuellement possible d'exclure certaines personnes du ciblage de façon sélective, ou de déléguer à la plateforme le repérage des personnes qui ont des caractéristiques, centres d'intérêts, et comportements communs.

C'est une **fonctionnalité puissante**, peut-être plus qu'il n'y paraît au premier abord. Elle permet en effet aux entreprises d'associer systématiquement leurs données client avec les données Facebook. Mieux encore, d'autres publicitaires et marchands de données peuvent également synchroniser leurs bases avec celles de la plateforme et en exploiter les ressources, ce qui équivaut à fournir une sorte de **télécommande en temps réel pour manipuler l'univers des données Facebook**. Les entreprises peuvent maintenant capturer en temps réel des données comportementales extrêmement précises comme un clic de souris sur un site, le glissement d'un doigt sur une application mobile ou un achat en magasin, et demander à Facebook de trouver et de cibler aussitôt les personnes qui viennent de se livrer à ces activités. Google et Twitter ont mis en place des fonctionnalités similaires en 2015.

Les plateformes de gestion de données

De nos jours, la plupart des entreprises de technologie publicitaire croisent en continu plusieurs sources de codage relatives aux individus. Les **plateformes de gestion de données** permettent aux entreprises de tous les domaines **d'associer et de relier leurs propres données clients**, comprenant des informations en temps réel sur les achats, les sites web consultés, les applications utilisées et les réponses aux courriels, avec des profils numériques fournis par une

multitude de fournisseurs tiers de données. Les données associées peuvent alors être analysées, triées et classées, puis utilisées pour envoyer un message donné à des personnes précises via des réseaux ou des appareils particuliers. Une entreprise peut, par exemple, cibler un groupe de clients existants ayant visité une page particulière sur son site ; ils sont alors **perçus comme pouvant devenir de bons clients**, bénéficiant alors de contenus personnalisés ou d'une réduction, que ce soit sur Facebook, sur une appli mobile ou sur le site même de l'entreprise.

L'émergence des plateformes de gestion de données **marque un tournant** dans le développement d'un envahissant pistage des comportements d'achat. Avec leur aide, les entreprises dans tous les domaines et partout dans le monde peuvent **très facilement associer et relier** les données qu'elles ont collectées depuis des années sur leurs clients et leurs prospects avec les milliards de profils collectés dans le monde numérique. Les principales entreprises faisant tourner ces plateformes sont : Oracle, Adobe, Salesforce (KruX), Wunderman (KBM Group/Zipline), Neustar, Lotame et Cxense.

Nous vous afficherons des publicités basées sur votre identité, mais cela ne veut pas dire que vous serez identifiable.

Erin Egan, Directeur de la protection de la vie privée chez Facebook, 2012

Identifier les gens et relier les profils numériques

Pour surveiller et suivre les gens dans les différentes situations de leur vie, pour leur associer des profils et toujours les reconnaître comme un seul et même individu, les entreprises amassent une grande variété de types de données qui, en quelque sorte, les identifient.

Parce qu'il est ambigu, le nom d'une personne a toujours été un mauvais identifiant pour un recueil de données. **L'adresse postale**, par contre, a longtemps été et est encore, une indication clé qui permet d'associer et de relier des données de différentes origines sur les consommateurs et leur famille. Dans le monde numérique, les identifiants les plus pertinents pour relier les profils et les comportements sur les différentes bases de données, plateformes et appareils sont : **l'adresse de courriel, le numéro de téléphone**, et le code propre à chaque smartphone ou autre appareil.

Les **identifiants de compte utilisateur** sur les immenses plateformes comme Google, Facebook, Apple et Microsoft jouent aussi un rôle important dans le suivi des gens sur Internet. Google, Apple, Microsoft et Roku attribuent un « **identifiant publicitaire** » aux individus, qui est maintenant largement utilisé pour faire correspondre et relier les données d'appareils tels que les smartphones avec les autres informations issues du monde numérique. Verizon utilise son propre identifiant pour pister les utilisateurs sur les sites web et les appareils. Certaines grandes entreprises de données comme **Acxiom, Experian et Oracle** disposent, au niveau mondial, d'un identifiant unique par personne qu'elles utilisent pour relier des dizaines d'années de données clients avec le monde numérique. Ces **identifiants d'entreprise** sont constitués le plus souvent de deux identifiants ou plus qui sont attachés à différents aspects de la vie en ligne et hors ligne d'une personne et qui peuvent être d'une certaine façon reliés l'un à l'autre.

Des Identifiants utilisés pour pister les gens sur les sites web, les appareils et les lieux de vie



Comment les entreprises identifient les consommateurs et les reliant à des informations de profils – sources : voir le rapport

Les entreprises de pistage utilisent également des identifiants plus ou moins temporaires, comme les cookies qui sont attachés aux utilisateurs surfant sur le web. Depuis que les utilisateurs peuvent ne pas autoriser ou supprimer les cookies dans leur navigateur, elles ont développé des méthodes sophistiquées permettant de calculer une empreinte numérique unique basée sur diverses caractéristiques du navigateur et de l'ordinateur d'une personne. De la même manière, les entreprises amassent les empreintes sur les appareils tels que les smartphones. Les cookies et les empreintes numériques sont **continuellement synchronisés** entre les différents services de pistage et ensuite reliés à des identifiants plus permanents.

D'autres entreprises fournissent des services de pistage multi-appareils qui utilisent le *machine learning* ([voir Wikipédia](#)) pour analyser de grandes quantités de données. Par exemple, **Tapad**, qui a été acheté par le géant des télécoms norvégiens Telenor, analyse les données de deux milliards d'appareils dans le monde et utilise des modèles basés sur les comportements et les relations pour trouver la probabilité qu'un ordinateur, une tablette, un téléphone ou un autre appareil appartienne à la même personne.

Un profilage « anonyme » ?

Les entreprises de données suppriment les noms dans leurs profils détaillés et utilisent des fonctions de hachage ([voir Wikipedia](#)) pour convertir les adresses de courriel et les numéros de téléphone en **code alphanumérique** comme “e907c95ef289”. Cela leur permet de déclarer sur leur site web et dans leur politique de confidentialité qu’elles recueillent, partagent et utilisent uniquement des données clients « anonymisées » ou « dé-identifiées ».

Néanmoins, comme la plupart des entreprises utilisent les mêmes process déterministes pour calculer ces codes alphanumériques, on devrait les considérer comme des pseudonymes qui sont en fait bien plus pratiques que les noms réels pour identifier les clients dans le monde numérique. Même si une entreprise partage des profils contenant uniquement des adresses de courriels ou des numéros de téléphones chiffrés, **une personne peut toujours être reconnue** dès qu’elle utilise un autre service lié avec la même adresse de courriel ou le même numéro de téléphone. De cette façon, bien que chaque service de pistage impliqué ne connaissent qu’une partie des informations du profil d’une personne, les entreprises peuvent **suivre et interagir avec les gens** au niveau individuel via les services, les plateformes et les appareils.

Si une entreprise peut vous suivre et interagir avec vous dans le monde numérique – et cela inclut potentiellement votre téléphone mobile ou votre télé – alors son affirmation que vous êtes anonyme n’a aucun sens, en particulier quand des entreprises ajoutent de temps à autre des informations hors-ligne aux données en ligne et masquent simplement le nom et l’adresse pour rendre le tout « anonyme ».

Joseph Turow, spécialiste du marketing et de la vie privée dans son livre « The Daily You », 2011

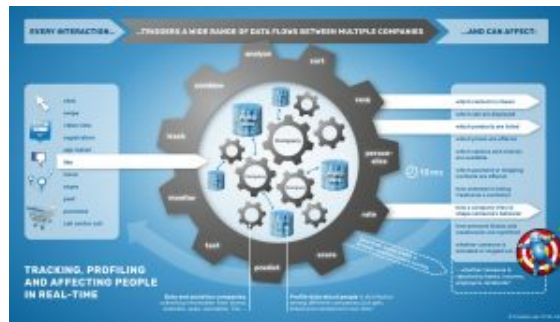
Gérer les clients et les comportements : personnalisation et évaluation

S'appuyant sur les méthodes sophistiquées d'interconnexion et de combinaison de données entre différents services, les entreprises de tous les secteurs d'activité peuvent utiliser les **flux de données comportementales actuellement omniprésents** afin de surveiller et d'analyser une large gamme d'activités et de comportements de consommateurs pouvant être pertinents vis-à-vis de leurs intérêts commerciaux.

Avec l'aide des vendeurs de données, les entreprises tentent d'entrer en contact avec les clients tout au long de leurs parcours autant de fois que possible, à travers les achats en ligne ou en boutique, le publipostage, les pubs télé et les appels des centres d'appels. Elles tentent **d'enregistrer et de mesurer chaque interaction** avec un consommateur, y compris sur les sites Internet, plateformes et appareils qu'ils ne contrôlent pas eux-mêmes. Elles peuvent recueillir en continu une abondance de données concernant leurs clients et d'autres personnes, les améliorer avec des informations provenant de tiers, et utiliser les profils améliorés au sein de l'écosystème de commercialisation et de technologie publicitaire. À l'heure actuelle, les plateformes de gestion des données clients permettent **la définition de jeux complexes de règles** qui régissent la façon de réagir automatiquement à certains critères tels que des activités ou des personnes données ou une combinaison des deux.

Par conséquent, les **individus ne savent jamais** si leur comportement a déclenché une réaction de l'un de ces réseaux de pistage et de profilage constamment mis à jour, interconnectés et opaques, ni, le cas échéant, comment cela influence les options qui leur sont proposées à travers les canaux de communication et dans les situations de vie.

Tracer, profiler et influencer les individus en temps réel



Chaque interaction enclenche un large éventail de flux de données entre de nombreuses entreprises.

Personnalisation en série

Les flux de données échangés entre les publicitaires en ligne, les courtiers en données, et les autres entreprises ne sont pas seulement utilisés pour diffuser de la publicité ciblée sur les sites web ou les applis mobiles. Ils sont de plus en plus utilisés pour **personnaliser les contenus, les options et les choix** offerts aux consommateurs sur le site d'une entreprise par exemple. Les entreprises de technologie des données, comme par exemple Optimizely, peuvent aider à personnaliser un site web spécialement pour les personnes qui le visitent pour la première fois, en s'appuyant sur les profils numériques de ces visiteurs fournis par Oracle.

Les boutiques en ligne, par exemple, personnalisent l'accueil des visiteurs : **quels produits** seront mis en évidence, **quelles promotions** seront proposées, et même le prix et des produits ou des services peuvent être différents selon la personne qui visite le site. Les services de détection de la fraude évaluent les utilisateurs en temps réel et décident **quels moyens de paiement et de transport** peuvent être proposés.

Les entreprises développent des technologies pour calculer et évaluer en continu le potentiel de valeur à long terme d'un client en s'appuyant sur son historique de navigation, de

recherche et de localisation, mais aussi sur son usage des applis, sur les produits achetés et sur ses amis sur les réseaux sociaux. Chaque clic, chaque glissement de doigt, chaque *Like*, chaque partage est susceptible d'influencer **la manière dont une personne est traitée en tant que client**, combien de temps elle va attendre avant que la *hotline* ne lui réponde, ou si elle sera complètement exclue des relances et des services marketing.

L'Internet des riches n'est pas le même que celui des pauvres.

Michael Fertik, fondateur de reputation.com, 2013

Trois types de **plateformes technologiques** jouent un rôle important dans cette sorte de personnalisation instantanée. Premièrement, les entreprises utilisent des systèmes de gestion de la relation client pour gérer leurs données sur les clients et les prospects. Deuxièmement, elles utilisent des plateformes de gestion de données pour connecter leurs propres données à l'écosystème de publicité numérique et obtiennent ainsi des informations supplémentaires sur le profil de leurs clients. Troisièmement, elles peuvent utiliser des **plateformes de marketing prédictif** qui les aident à produire le bon message pour la bonne personne au bon moment, calculant comment convaincre quelqu'un en exploitant ses faiblesses et ses préjugés.

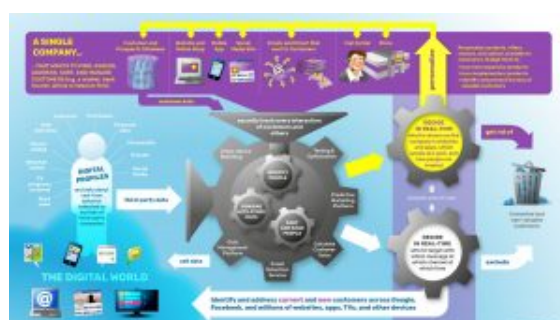
Par exemple, l'entreprise de données **RocketFuel** promet à ses clients de « leur apporter des milliers de milliards de signaux numériques ou non pour créer des profils individuels et pour fournir aux consommateurs une expérience personnalisée, toujours actualisée et toujours pertinente » s'appuyant sur les 2,7 milliards de profils uniques de son dépôt de données. Selon RocketFuel, il s'agit « de noter chaque signal selon sa propension à influencer le consommateur ».

La plateforme de marketing prédictif **TellApart**, qui appartient à **Twitter**, associe une valeur à chaque couple client/produit acheté, une « synthèse entre la probabilité d'achat, l'importance de la commande et la valeur à long terme », s'appuyant sur « des centaines de signaux en ligne et en magasin sur un consommateur anonyme unique ». En conséquence, TellApart regroupe automatiquement du contenu tel que « l'image du produit, les logos, les offres et toute autre métadonnée » pour construire des publicités, des courriels, des sites web et des offres personnalisées.

Tarifs personnalisés et campagnes électorales

Des méthodes identiques peuvent être utilisées pour personnaliser les tarifs dans les boutiques en ligne, par exemple, en prédisant le niveau d'achat d'un client à long terme ou le montant qu'il sera probablement prêt à payer un peu plus tard. Des preuves sérieuses suggèrent que les boutiques en ligne affichent déjà des **tarifs différents selon les consommateurs**, ou même des prix différents pour le même produit, en s'appuyant sur leur comportement et leurs caractéristiques. Un champ d'action similaire est la personnalisation lors des campagnes électorales. Le ciblage des électeurs avec des messages personnalisés, adaptés à leur personnalité, et à leurs opinions politiques sur des problèmes donnés a fait monter les débats sur une possible **manipulation politique**.

Utiliser les données, les analyser et les personnaliser pour gérer les consommateurs



Actuellement, dans tous les

domaines, les entreprises peuvent mobiliser les réseaux de suivi et de profilage pour trouver, évaluer, contacter, trier et gérer les consommateurs

Tests et expériences sur les personnes

La personnalisation s'appuyant sur de riches informations de profil et sur du suivi invasif en temps réel est devenue **un outil puissant pour influencer le comportement du consommateur** quand il visite une page web, clique sur une pub, s'inscrit à un service, s'abonne à une newsletter, télécharge une application ou achète un produit.

Pour améliorer encore cela, les entreprises ont commencé à faire des expériences en continu sur les individus. Elles procèdent à **des tests en faisant varier** les fonctionnalités, le design des sites web, l'interface utilisateur, les titres, les boutons, les images ou mêmes les tarifs et les remises, surveillent et mesurent avec soin comment les différents groupes d'utilisateurs interagissent avec ces modifications. De cette façon, les entreprises **optimisent sans arrêt leur capacité à encourager les personnes** à agir comme elles veulent qu'elles agissent.

Les organes de presse, y compris à grand tirage comme le Washington Post, utilisent **différentes versions des titres de leurs articles** pour voir laquelle est la plus performante. Optimizely, un des principaux fournisseurs de technologies pour ce genre de tests, propose à ses clients la capacité de « faire des tests sur l'ensemble de l'expérience client sur n'importe quel canal, n'importe quel appareil, et n'importe quelle application ». Expérimenter sur des usagers qui l'ignorent **est devenu la nouvelle norme**.

En 2014, **Facebook** a déclaré faire tourner « plus d'un millier d'expérimentations chaque jour » afin « d'optimiser des résultats précis » ou pour « affiner des décisions de design sur le long terme ». En 2010 et 2012, la plateforme a mené des expérimentations sur des millions d'utilisateurs et montré qu'en manipulant l'interface utilisateur, les fonctionnalités et le contenu affiché, Facebook **pouvait augmenter significativement le taux de participation électorale** d'un groupe de personnes. Leur célèbre expérimentation sur l'humeur des internautes, portant sur 700 000 individus, consistait à manipuler secrètement la quantité de messages émotionnellement positifs ou négatifs présents dans les fils d'actualité des utilisateurs : il s'avéra que cela avait **un impact sur le nombre de messages positifs ou négatifs** que les utilisateurs postaient ensuite eux-mêmes.

Suite à la critique massive de Facebook par le public concernant cette expérience, la plateforme de rendez-vous **OkCupid** a publié un article de blog provocateur défendant de telles pratiques, déclarant que « nous faisons des expériences sur les êtres humains » et « c'est ce que font tous les autres ». OkCupid a décrit une expérimentation dans laquelle a été manipulé le pourcentage de « compatibilité » montré à des paires d'utilisateurs. Quand on affichait un taux de 90 % entre deux utilisateurs qui en fait étaient peu compatibles, les utilisateurs échangeaient nettement plus de messages entre eux. OkCupid a déclaré que quand elle « dit aux gens » qu'ils « vont bien ensemble », alors ils « agissent comme si c'était le cas ».

Toutes ces expériences qui posent de vraies questions éthiques montrent le **pouvoir de la personnalisation basée sur les données pour influencer sur les comportements.**

Dans les mailles du filet : vie quotidienne, données commerciales et analyse du risque

Les données concernant les comportements des personnes, les liens sociaux, et les moments les plus intimes sont de plus en plus utilisées dans des contextes ou à des fins complètement différents de ceux dans lesquels elles ont été enregistrées. Notamment, elles sont de plus en plus utilisées pour prendre des **décisions automatisées au sujet d'individus** dans des domaines clés de la vie tels que la finance, l'assurance et les soins médicaux.

Données relatives aux risques pour le marketing et la gestion client

Les agences d'évaluation de la solvabilité, ainsi que d'autres acteurs clés de l'évaluation du risque, principalement dans des domaines tels que la **vérification des identités, la prévention des fraudes, les soins médicaux et l'assurance** fournissent également des solutions commerciales. De plus, la plupart des courtiers en données s'échangent divers types d'informations sensibles, par exemple des informations concernant la situation financière d'un individu, et ce à des fins commerciales. **L'utilisation de l'évaluation de solvabilité à des fins de marketing** afin soit de cibler soit d'exclure des ensembles vulnérables de la population a évolué pour devenir des produits qui associent le marketing et la gestion du risque.

L'agence d'évaluation de la solvabilité **TransUnion** fournit, par exemple, un **produit d'aide à la décision piloté par les données** à destination des commerces de détail et des services financiers qui leur permet « de mettre en œuvre des stratégies de marketing et de gestion du risque sur mesure pour atteindre les objectifs en termes de clients, canaux de vente et résultats commerciaux », il inclut des données de crédit et

promet « un aperçu inédit du comportement, des préférences et des risques du consommateur. » Les entreprises peuvent alors laisser leurs clients « choisir parmi une gamme complète d'offres sur mesure, répondant à leurs besoins, leurs préférences et leurs profils de risque » et « évaluer leurs clients sur divers produits et canaux de vente et leur présenter uniquement la ou les offres les plus pertinentes pour eux et les plus rentables » pour l'entreprise. De même, **Experian** fournit un produit qui associe « crédit à la consommation et informations commerciales, fourni avec plaisir par Experian. »

En matière de surveillance, il n'est pas question de connaître vos secrets, mais de gérer des populations, de gérer des personnes.

Katarzyna Szymielewicz, Vice-Présidente EDRi, 2015

Vérification des identités en ligne et détection de la fraude

Outre la machine de surveillance en temps réel qui a été développée au travers de la publicité en ligne, d'autres formes de pistage et de profilage généralisées ont émergé dans les domaines de **l'analyse de risque, de la détection de fraudes et de la cybersécurité.**

De nos jours, les services de détection de fraude en ligne utilisent des technologies hautement intrusives afin **d'évaluer des milliards de transactions numériques.** Ils recueillent d'énormes quantités d'informations concernant les appareils, les individus et les comportements. Les fournisseurs habituels dans l'évaluation de solvabilité, la vérification d'identité, et la prévention des fraudes ont commencé à surveiller et à évaluer la façon dont les personnes surfent sur le web et utilisent leurs appareils mobiles. En outre, ils ont entrepris

de relier les données comportementales en ligne avec l'énorme quantité d'information hors-connexion qu'ils recueillent depuis des dizaines d'années.

Avec l'émergence de services passant par l'intermédiaire d'objets technologiques, la vérification de l'identité des consommateurs et la prévention de la fraude sont devenues de plus en plus importantes et de plus en plus contraignantes, notamment au vu de la cybercriminalité et de la fraude automatisée. Dans un même temps, les systèmes actuels d'analyse du risque ont agrégé **des bases de données gigantesques contenant des informations sensibles** sur des pans entiers de population. Nombre de ces systèmes répondent à un grand nombre de cas d'utilisation, parmi lesquels la preuve d'identité pour les services financiers, l'évaluation des réclamations aux compagnies d'assurance et des demandes d'indemnités, de l'analyse des transactions financières et l'évaluation de milliards de transactions en ligne.

De tels systèmes d'analyse du risque peuvent décider si une requête ou une transaction est **acceptée ou rejetée** ou décider des options de livraison disponibles pour une personne lors d'une transaction en ligne. Des services marchands de vérification d'identité et d'analyse de la fraude sont également employés dans des domaines tels que **les forces de l'ordre et la sécurité nationale**. La frontière entre les applications commerciales de l'analyse de l'identité et de la fraude et celles utilisées par les agences gouvernementales de renseignement est de plus en plus floue.

Lorsque des individus sont ciblés par des systèmes aussi opaques, ils peuvent être **signalés comme étant suspects et nécessitant un traitement particulier ou une enquête**, ou bien ils peuvent être rejetés sans plus d'explication. Ils peuvent recevoir un courriel, un appel téléphonique, une notification, un message d'erreur, ou bien le système peut tout simplement ne pas indiquer une option, sans que l'utilisateur ne connaisse son existence pour d'autres. Des évaluations

erronées peuvent se propager d'un système à l'autre. Il est souvent **difficile, voire impossible de faire recours** contre ces évaluations négatives qui excluent ou rejettent, notamment à cause de la difficulté de s'opposer à quelque chose dont on ne connaît pas l'existence.

Exemples de détection de fraude en ligne et de service d'analyse des risques

L'entreprise de cybersécurité **ThreatMetrix** traite les données concernant 1,4 milliard de « comptes utilisateur uniques » sur des « milliers de sites dans le monde. » Son Digital Identity Network (Réseau d'Identité Numérique) enregistre des « millions d'opérations faites par des consommateurs chaque jour, notamment des connexions, des paiements et des créations de nouveaux comptes », et cartographie les « associations en constante évolution entre les individus et leurs appareils, leurs positions, leurs identifiants et leurs comportements » à des fins de vérification des identités et de prévention des fraudes. L'entreprise collabore avec Equifax et TransUnion. Parmi ses clients se trouvent Netflix, Visa et des entreprises dans des secteurs tels que le jeu vidéo, les services gouvernementaux et la santé.

De façon analogue, l'entreprise de données **ID Analytics**, qui a récemment été achetée par **Symantec**, exploite un Réseau d'Identifiants fait de « 100 millions de nouveaux éléments d'identité quotidiens issus des principales organisations interprofessionnelles. ». L'entreprise agrège des données concernant 300 millions de consommateurs, sur les prêts à haut risque, les achats en ligne et les demandes de carte de crédit ou de téléphone portable. Son Indice d'Identité, ID Score, prend en compte les appareils numériques ainsi que les noms, les numéros de sécurité sociale et les adresses postales et courriel.

Trustev, une entreprise en ligne de détection de la fraude dont le siège se situe en Irlande et qui a été rachetée par

l'agence d'évaluation de la solvabilité **TransUnion** en 2015, juge des transactions en ligne pour des clients dans les secteurs des services financiers, du gouvernement, de la santé et de l'assurance en s'appuyant sur l'analyse des comportements numériques, les identités et les appareils tels que les téléphones, les tablettes, les ordinateurs portables, les consoles de jeux, les télévisions et même les réfrigérateurs. L'entreprise propose aux entreprises clientes la possibilité d'analyser **la façon dont les visiteurs cliquent et interagissent avec les sites Internet et les applications**. Elle utilise une large gamme de données pour évaluer les utilisateurs, y compris les numéros de téléphone, les adresses courriel et postale, les empreintes de navigateur et d'appareil, les vérifications de la solvabilité, les historiques d'achats sur l'ensemble des vendeurs, les adresses IP, les opérateurs mobiles et la géolocalisation des téléphones. Afin d'aider à « accepter les transactions futures », chaque appareil se voit attribuer une empreinte digitale d'appareil unique. Trustev propose aussi une technologie de marquage d'empreinte digitale sociale qui analyse le contenu des réseaux sociaux, notamment une « analyse de la liste d'amis » et « l'identification des schémas ». TransUnion a intégré la technologie Trustev dans ses propres solutions identifiantes et anti-fraude.

Selon son site Internet, Trustev utilise une large gamme de données pour évaluer les personnes



Capture d'écran du site Internet de Trustev, 2 juin 2016

De façon similaire, l'agence d'évaluation de la solvabilité Equifax affirme qu'elle possède des données concernant près de 1 milliard d'appareils et peut affirmer « l'endroit où se situe en fait un appareil et s'il est associé à d'autres appareils utilisés dans des fraudes connues ». En associant ces données avec « des milliards d'identités et d'événements de crédit pour trouver les activités douteuses » dans tous les secteurs, et en utilisant des informations concernant **la situation d'emploi et les liens entre les ménages, les familles et les partenaires**, Equifax prétend être capable « de distinguer les appareils ainsi que les individus ».

Je ne suis pas un robot

Le produit **reCaptcha** de Google fournit en fait un service similaire, du moins en partie. Il est incorporé dans des millions de sites Internets et aide les fournisseurs de sites Internets à décider si un visiteur est un être humain ou non. Jusqu'à récemment, les utilisateurs devaient résoudre diverses sortes de défis rapides tels que le déchiffrement de lettres dans une image, la sélection d'images dans une grille, ou simplement en cochant la case « Je ne suis pas un robot ». En 2017, **Google** a présenté une version invisible de reCaptcha, en expliquant qu'à partir de maintenant, les utilisateurs humains pourront passer « **sans aucune interaction utilisateur**, contrairement aux utilisateurs douteux et aux robots ». L'entreprise ne révèle pas le type de données et de comportements utilisateurs utilisés pour reconnaître les humains. Des analyses laissent penser que Google, outre les adresses IP, les empreintes de navigateur, la façon dont l'utilisateur frappe au clavier, déplace la souris ou utilise l'écran tactile « avant, pendant et après » une interaction reCaptcha, utilise plusieurs témoins Google. **On ne sait pas exactement** si les individus sans compte utilisateur sont désavantagés, si Google est capable d'identifier des individus particuliers plutôt que des « humains » génériques, ou si

Google utilise les données enregistrées par reCaptcha à d'autres fins que la détection de robots.

Le pistage numérique à des fins publicitaires et de détection de la fraude ?

Les flux omniprésents de données comportementales enregistrées pour la publicité en ligne s'écoulent vers les systèmes de détection de la fraude. Par exemple, **la plateforme de données commerciales** Segment propose à ses clients des moyens faciles d'envoyer des données concernant leurs clients, leur site Internet et les utilisateurs mobiles à une kyrielle de services de technologies commerciales, ainsi qu'à des **entreprises de détection de fraude**. Castle est l'une d'entre-elles et utilise « les données comportementales des consommateurs pour prédire les utilisateurs qui présentent vraisemblablement un risque en matière de sécurité ou de fraude ». Une autre entreprise, Smyte, aide à « prévenir les arnaques, les messages indésirables, le harcèlement et les fraudes par carte de crédit ».

La grande agence d'analyse de la solvabilité **Experian** propose un service de pistage multi-appareils qui fournit de la reconnaissance universelle d'appareils, sur mobile, Internet et les applications **pour le marketing numérique**. L'entreprise s'engage à concilier et à associer les « identifiants numériques existants » de leurs clients, y compris des « témoins, identifiants d'appareil, adresses IP et d'autres encore », fournissant ainsi aux commerciaux un « lien omniprésent, cohérent et permanent sur tous les canaux ».

La technologie d'identification d'appareils provient de **41st parameter** (le 41e paramètre), une entreprise de détection de la fraude rachetée par Experian en 2013. En s'appuyant sur la technologie développée par 41st parameter, Experian propose aussi une solution d'intelligence d'appareil pour **la détection de la fraude** au cours des paiements en ligne. Cette solution qui « crée un identifiant fiable pour l'appareil et recueille

des données appareil abondantes » « identifie en quelques millisecondes chaque appareil à chaque visite » et « fournit une visibilité jamais atteinte de l'individu réalisant le paiement ». On ne sait pas exactement si Experian utilise les mêmes données pour ses services d'identification d'appareils pour détecter la fraude que pour le marketing.

Cartographie de l'écosystème du pistage et du profilage commercial

Au cours des dernières années, les pratiques déjà existantes de surveillance commerciale ont rapidement muté en un large éventail d'acteurs du secteur privé qui surveillent en permanence des populations entières. Certains des acteurs de l'écosystème actuel de pistage et de profilage, tels que **les grandes plateformes** et d'autres **entreprises avec un grand nombre de clients**, tiennent une position unique en matière d'étendue et de niveau de détail de leurs profils de consommateurs. Néanmoins, les données utilisées pour prendre des décisions concernant les individus sur de nombreux sujets ne sont généralement pas centralisées en un lieu, mais plutôt assemblées en temps réel à partir de plusieurs sources selon les besoins.

Un large éventail **d'entreprises de données et de services d'analyse** en marketing, en gestion client et en analyse du risque recueillent, analysent, partagent et échangent de façon uniforme des données client et les associent avec des informations supplémentaires issues de milliers d'autres entreprises. Tandis que l'industrie des données et des services d'analyse fournissent les moyens pour déployer ces puissantes technologies, **les entreprises dans de nombreuses industries** contribuent à augmenter la quantité et le niveau de détail des données collectées ainsi que la capacité à les utiliser.

Cartographie de l'écosystème du pistage et du profilage

commercial numérique



En plus des grandes plateformes en ligne et de l'industrie des données et des services d'analyse des consommateurs, des entreprises dans de nombreux secteurs ont rejoint les écosystèmes de pistage et de profilage numérique généralisé.

Google et Facebook, ainsi que d'autres grandes plateformes telles que Apple, Microsoft, Amazon et Alibaba ont un accès sans précédent à des données concernant les vies de milliards de personnes. Bien qu'ils aient des modèles commerciaux différents et jouent par conséquent des rôles différents dans l'industrie des données personnelles, ils ont le pouvoir de dicter dans une large mesure les paramètres de base des marchés numériques globaux. Les grandes plateformes limitent principalement la façon dont les autres entreprises peuvent obtenir leurs données. Ainsi, ils les obligent à utiliser les **données utilisateur de la plateforme dans leur propre écosystème** et recueillent des données au-delà de la portée de la plateforme.

Bien que **les grandes multinationales** de différents secteurs ayant des interactions fréquentes avec des centaines de millions de consommateurs soient en quelque sorte dans une

situation semblable, elles ne font pas qu'acheter des données clients recueillies par d'autres, elles en fournissent aussi. Bien que certaines parties des secteurs des services financiers et des télécoms ainsi que des **domaines sociétaux critiques** tels que la santé, l'éducation et l'emploi soient soumis à une réglementation plus stricte dans la plupart des juridictions, un large éventail d'entreprises a commencé à utiliser ou fournissent des données aux réseaux actuels de surveillance commerciale.

Les détaillants et d'autres entreprises qui vendent des produits et services aux consommateurs vendent pour la plupart les données concernant les achats de leurs clients. **Les conglomérats médiatiques et les éditeurs numériques** vendent des données au sujet de leur public qui sont ensuite utilisées par des entreprises dans la plupart des autres secteurs. **Les fournisseurs de télécoms et d'accès haut débit** ont entrepris de suivre leurs clients sur Internet. Les grandes groupes de distribution, de médias et de télécoms ont acheté ou achètent **des entreprises de données, de pistage et de technologie publicitaire**. Avec le rachat de NBC Universal par Comcast et le rachat probable de Time Warner par AT&T, les grands groupes de télécoms aux États-Unis sont aussi en train de devenir des éditeurs gigantesques, créant par là même des portefeuilles puissants de contenu, de données et de capacité de pistage. Avec l'acquisition de AOL et de Yahoo, Verizon aussi est devenu une « plateforme ».

Les institutions financières ont longtemps utilisé des données sur les consommateurs pour la gestion du risque, notamment dans l'évaluation de la solvabilité et la détection de fraude, ainsi que pour le marketing, l'acquisition et la rétention de clientèle. Elles complètent leurs propres données avec des données externes issues d'agences d'évaluation de la solvabilité, de courtiers en données et d'entreprises de données commerciales. **PayPal**, l'entreprise de paiements en ligne la plus connue, partage des informations personnelles

avec plus de 600 tiers, parmi lesquels d'autres fournisseurs de paiements, des agences d'évaluation de la solvabilité, des entreprises de vérification de l'identité et de détection de la fraude, ainsi qu'avec les acteurs les plus développés au sein de l'écosystème de pistage numérique. Tandis que les réseaux de cartes de crédit et les banques ont partagé des informations financières sur leurs clients avec les fournisseurs de données de risque depuis des dizaines d'années, ils ont maintenant commencé à **vendre des données sur les transactions** à des fins publicitaires.

Une myriade d'entreprises, grandes ou petites, **fournissant des sites Internets, des applications mobiles, des jeux et d'autres solutions** sont étroitement liées à l'écosystème de données commerciales. Elles utilisent des services qui leur permettent de facilement transmettre à des services tiers des données concernant leurs utilisateurs. Pour nombre d'entre elles, la vente de flux de données comportementales concernant leurs utilisateurs constitue un élément clé de leur *business model*. De façon encore plus inquiétante, les entreprises qui fournissent des services tels que les **enregistreurs d'activité physique** intègrent des services qui transmettent les données utilisateurs à des tierces parties.

L'envahissante machine de surveillance en temps réel qui a été développée pour **la publicité en ligne** est en train de s'étendre vers d'autres domaines dont la politique, la tarification, la notation des crédits et la gestion des risques. Partout dans le monde, **les assureurs** commencent à proposer à leurs clients des offres incluant du suivi en temps réel de leur comportement : comment ils conduisent, quelles sont leurs activités santé ou leurs achats alimentaires et quand ils se rendent au club de gym. Des nouveaux venus dans **l'analyse assurantielle et les technologies financières** prévoient les risques de santé d'un individu en s'appuyant sur les données de consommation, mais évaluent aussi la solvabilité à partir de données de comportement via les appels

téléphoniques ou les recherches sur Internet.

Les courtiers en données sur les consommateurs, les entreprises de gestion de clientèle et les agences de publicité comme Acxiom, Epsilon, Merkle ou Wunderman/WPP jouent un rôle prépondérant en rassemblant et reliant les données entre les plateformes, les multinationales et le monde de la technologie publicitaire. **Les agences d'évaluation de crédit** comme Experian qui fournissent de nombreux services dans des domaines très sensibles comme l'évaluation de crédit, la vérification d'identité et la détection de la fraude jouent également un rôle prépondérant dans l'actuel envahissant écosystème de la commercialisation des données.

Des entreprises particulièrement importantes qui fournissent des données, des analyses et des solutions logicielles sont également appelées « plateforme ». **Oracle**, un fournisseur important de logiciel de base de données est, ces dernières années, devenu un courtier en données de consommation. **Salesforce**, le leader sur le marché de la gestion de la relation client qui gère les bases de données commerciales de millions de clients qui ont chacun de nombreux clients, a récemment acquis Krux, une grande entreprise de données, connectant et combinant des données venant de l'ensemble du monde numérique. L'entreprise de logiciels **Adobe** joue également un rôle important dans le domaine des technologies de profilage et de publicité.

En plus, les principales grandes entreprises **du conseil, de l'analyse et du logiciel commercial**, comme IBM, Informatica, SAS, FICO, Accenture, Capgemini, Deloitte et McKinsey et même des **entreprises spécialisées dans le renseignement et la défense** comme Palantir, jouent également un rôle significatif dans la gestion et l'analyse des données personnelles, de la gestion de la relation client à celle de l'identité, du marketing à l'analyse de risque pour les assureurs, les banques et les gouvernements.

Vers une société du contrôle social numérique généralisé ?

Ce rapport montre qu'aujourd'hui, les réseaux entre plateformes en ligne, fournisseurs de technologies publicitaires, courtiers en données, et autres peuvent **suivre, reconnaître et analyser des individus dans de nombreuses situations de la vie courante**. Les informations relatives aux comportements et aux caractéristiques d'un individu sont reliées entre elles, assemblées, et utilisées en temps réel par des entreprises, des bases de données, des plateformes, des appareils et des services. Des acteurs uniquement motivés par des buts économiques ont fait naître un environnement de données dans lequel les individus sont constamment **sondés et évalués, catégorisés et regroupés, notés et classés, numérotés et comptés, inclus ou exclus**, et finalement traités de façon différente.

Ces dernières années, plusieurs **évolutions importantes** ont donné de nouvelles capacités sans précédent à la surveillance omniprésente par les entreprises. Cela comprend l'augmentation des médias sociaux et des appareils en réseau, le pistage et la mise en relation en temps réel de flux de données comportementales, le rapprochement des données en ligne et hors ligne, et la consolidation des données commerciales et de gestion des risques. L'envahissant pistage et profilage numériques, mélangé à la personnalisation et aux tests, ne sont pas seulement utilisés pour surveiller, mais aussi pour influencer systématiquement **le comportement des gens**. Quand les entreprises utilisent les données sur les situations du quotidien pour prendre des décisions parfois triviales, parfois conséquente sur les gens, cela peut conduire à **des discriminations**, et renforcer voire aggraver des inégalités existantes.

Malgré leur omniprésence, seul le haut de l'iceberg des données et des activités de profilage est visible pour les particuliers. La plupart d'entre elles restent opaques et à

peine compréhensible par la majorité des gens. Dans le même temps, les gens ont **de moins en moins de solutions pour résister au pouvoir de cet écosystème de données** ; quitter le pistage et le profilage envahissant, est devenu synonyme de quitter la vie moderne. Bien que les responsables des entreprises affirment que la vie privée est morte (tout en prenant soin de préserver leur propre vie privée), Mark Andrejevic suggère que les gens perçoivent en fait l'asymétrie du pouvoir dans le monde numérique actuel, mais se sentent « frustrés par un sentiment d'impuissance face à une collecte et à une exploitation de données de plus en plus sophistiquées et exhaustives. »

Au regard de cela, ce rapport se concentre sur le fonctionnement interne et les pratiques en vigueur dans l'actuelle industrie des données personnelles. Bien que l'image soit devenue plus nette, de larges portions du système restent encore dans le noir. **Renforcer la transparence sur le traitement des données par les entreprises** reste un prérequis indispensable pour résoudre le problème de l'asymétrie entre les entreprises de données et les individus. Avec un peu de chance, les résultats de ce rapport encourageront des travaux ultérieurs de la part de journalistes, d'universitaires, et d'autres personnes concernés par les libertés civiles, la protection des données et celle des consommateurs ; et dans l'idéal des travaux des législateurs et des entreprises elles-mêmes.

En 1999, Lawrence Lessig, avait bien prédit que, laissé à lui-même, le cyberspace, deviendrait un parfait outil de contrôle façonné principalement par la « main invisible » du marché. Il avait dit qu'il était possible de « construire, concevoir, ou programmer le cyberspace pour protéger les valeurs que nous croyons fondamentales, ou alors de construire, concevoir, ou programmer le cyberspace pour permettre à toutes ces valeurs de disparaître. » De nos jours, la deuxième option est presque devenue réalité au vu des milliards de dollars investis dans

le capital-risque pour financer des modèles économiques s'appuyant sur une exploitation massive et sans scrupule des données. L'insuffisance de régulation sur la vie privée aux USA et l'absence de son application en Europe ont réellement gêné l'émergence **d'autres modèles d'innovation numérique**, qui seraient fait de pratiques, de technologies, de modèles économiques qui protègent la liberté, la démocratie, la justice sociale et la dignité humaine.

À un niveau plus global, la législation sur la protection des données ne pourra pas, à elle seule, atténuer les conséquences qu'un monde « conduit par les données » a sur les individus et la société que ce soit aux USA ou en Europe. Bien que le **consentement et le choix** soient des principes cruciaux pour résoudre les problèmes les plus urgents liés à la collecte massive de données, ils peuvent également mener à une illusion de volontarisme. En plus d'instruments de régulation supplémentaires sur la non-discrimination, la protection du consommateur, les règles de concurrence, il faudra en général **un effort collectif important** pour donner une vision positive d'une future société de l'information. Sans quoi, on pourrait se retrouver bientôt dans une société avec un envahissant contrôle social numérique, dans la laquelle la vie privée deviendrait, si elle existe encore, un luxe pour les riches. Tous les éléments en sont déjà en place.

Lectures pour approfondir le sujet

– L'article ci-dessus en format .PDF (376,2 Ko)

–> [framablog.org-Comment les entreprises surveillent notre quotidien](https://framablog.org/Comment-les-entreprises-surveillent-notre-quotidien)

– Un essai plus exhaustif sur les questions abordées par la publication ci-dessus ainsi que des références et des sources peuvent être trouvés dans [le rapport complet, disponible au téléchargement en PDF.](#)

– Le rapport de 2016 « Les réseaux du contrôle » par Wolfie Christl et Sarah Spiekermann sur lequel le présent rapport est

largement fondé est [disponible au téléchargement en PDF ainsi qu'en format papier.](#)

La production de ce rapport, matériaux web et illustrations a été soutenue par Open Society Foundations.

Bibliographie

Christl, W. (2017, juin). Corporate surveillance in everyday life. Cracked Labs.

Christl, W., & Spiekermann, S. (2016). *Networks of Control, a Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy* (p. 14-20). Consulté à l'adresse https://www.privacylab.at/wp-content/uploads/2016/09/Christl-Networks__K_o.pdf

Epp, C., Lippold, M., & Mandryk, R. L. (2011). Identifying emotional states using keystroke dynamics (p. 715). ACM Press. <https://doi.org/10.1145/1978942.1979046>

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802-5805. <https://doi.org/10.1073/pnas.1218772110>

Turow, J. (s. d.). *Daily You* | Yale University Press. Consulté 25 septembre 2017, à l'adresse

Des routes et des ponts (17) – une vue d'ensemble

Nous arrivons bientôt au terme de notre traduction semaine après semaine de l'ouvrage *Des routes et des ponts* de Nadia Eghbal ([version originale en PDF](#)). Pour remonter vers les épisodes précédents il suffit de cliquer [sur ce lien](#).

Aujourd'hui nous vous proposons un chapitre qui recense les moyens dont les institutions et entreprises pourraient contribuer au développement et à la pérennité des projets *open source* qui constituent la colonne vertébrale de l'infrastructure numérique.

Traduction Framalang : Penguin, lyn, goofy, Opsylac, xXx

Une esquisse du tableau

Il est trop tôt pour dire à quoi devrait ressembler le soutien institutionnel à long terme d'un point de vue prospectif, mais il y a plusieurs domaines de travail critiques qui peuvent nous aider à le déterminer. Les propositions suivantes sont rattachées à trois domaines :

- **Traiter les infrastructures numériques comme un bien commun** essentiel et les élever au rang d'acteur intersectoriel clé ;
- **Travailler avec des projets** pour améliorer les standards, la sécurité et les flux de production ;

- **Augmenter la taille du groupe de contributeurs** de manière à ce que davantage de personnes, et davantage de personnes de types différents, puissent élaborer et soutenir ensemble les logiciels publics.

Conscientiser et éduquer les acteurs clés

Comme nous l'avons relevé dans ce rapport, beaucoup d'acteurs clés – dont les *startups*, les gouvernements, et les sociétés de capital risque – pensent à tort que les logiciels publics « fonctionnent, tout simplement » et ne requiert pas de maintenance supplémentaire. Pour entretenir correctement l'écosystème de nos infrastructures numériques, ces populations devraient être les premières à être informées du problème. Les infrastructures numériques ont besoin de porte-paroles qui soient affranchis de toute contrainte politique ou commerciale et qui puissent comprendre et communiquer les besoins de l'écosystème.

Traiter les infrastructures numériques comme des biens publics essentiels pourrait également motiver l'investissement direct dans la construction de meilleurs systèmes à partir de zéro. Par exemple, aux États-Unis, les autoroutes inter-états et le réseau de bibliothèques publiques furent dès l'origine conçus comme des ressources publiques. Les unes et les autres ont eu leur champion (respectivement le Président Dwight Eisenhower et le philanthrope Andrew Carnegie) qui ont clairement argumenté en faveur du bénéfice social et financier qui résulterait de tels projets.

Un réseau national d'autoroutes ne sert pas uniquement à nous relier en tant qu'individus, en facilitant les déplacements d'un endroit à un autre, mais il apporte aussi la prospérité financière dans tous les coins du pays, grâce à l'usage commercial des voies rapides pour acheminer les marchandises. Dans les bibliothèques Andrew Carnegie, publiques et gratuites, les livres étaient accessibles et non stockés en magasin, pour permettre aux gens de les feuilleter et d'y

trouver eux-mêmes l'information sans avoir recours à un·e bibliothécaire. Cette pratique a aidé à démocratiser l'information et à donner aux gens les moyens de s'éduquer eux-mêmes.

Une meilleure éducation et une meilleure prise de conscience pourraient s'étendre jusqu'aux gouvernements, dont certains ont rendu, par la loi, les infrastructures numériques difficiles à soutenir et qui ne sont peut-être pas familiers des normes culturelles et de l'histoire de l'*open source*. Aux USA, l'IRS [NdT : Internal Revenue Service – organisme qui collecte l'impôt] a une définition très restrictive des activités caritatives, et comme l'*open source* est mal comprise, son impact positif sur la société demeure ignoré. Cela complique l'institutionnalisation de plus gros projets à travers une fondation ou une association professionnelle.

Mesurer l'utilisation et l'impact de l'infrastructure numérique

L'impact de l'infrastructure numérique est encore très difficile à mesurer. Les indicateurs habituels sont soit très imprécis, soit simplement indisponibles. Ce n'est pas un problème facile à résoudre. Mais sans données relatives aux outils utilisés et à notre dépendance vis-à-vis d'eux, il est difficile de se faire une idée nette de ce qui manque de financement.

Avec de meilleurs indicateurs, nous pourrions décrire l'impact économique de l'infrastructure numérique, identifier les projets essentiels qui manquent de financement, et comprendre les dépendances entre les projets et entre les personnes. Pour le moment, il est impossible de dire qui utilise un projet *open source* à moins que l'utilisateur, individu ou entreprise, ne le révèle. Pour déterminer quel projet a besoin de plus de soutien, nous ne disposons que d'informations anecdotiques.

De meilleures statistiques pourraient également nous aider à

identifier les « contributeurs clé de voûte ». En biologie environnementale, une « espèce clé » est une espèce animale qui a un impact disproportionné sur son environnement au regard de ses effectifs. Dans la même idée, un « contributeur clé » pourrait être un développeur qui contribue à plusieurs projets essentiels, qui est le seul responsable d'un projet crucial, ou qui est généralement perçu comme influent et digne de confiance. Les « contributeurs clés » sont des défenseurs essentiels, les valoriser en leur fournissant les ressources dont ils ont besoin pourrait améliorer le système dans son ensemble. Comprendre les relations entre les communautés *open source* et les « contributeurs clés » pourrait aider à identifier rapidement les secteurs qui auront besoin de soutien supplémentaire.

On dispose également de peu de données sur les contributeurs eux-mêmes : qui contribue à l'*open source*, quelles conditions leur permettent de le faire, et quelles sont les contributions effectuées. Les femmes, les non-anglophones, et les nouveaux contributeurs à l'*open source* sont des exemples de population qui devraient être suivies dans le temps, en particulier pour mesurer l'impact des programmes de soutien.

Les seules statistiques disponibles sur les dépôts GitHub sont le nombre de personnes ayant étoilé (action semblable à *liker*), vu (c'est-à-dire qu'elles reçoivent des nouvelles du projet) ou « forké » un projet. Ces chiffres permettent de fournir des indicateurs concernant la popularité, mais ils peuvent être trompeurs. Beaucoup de personnes peuvent étoiler un projet, parce qu'il a une conception intéressante par exemple, sans toutefois l'intégrer à leur propre code.

Certains gestionnaires de paquets tels npm (qui est celui de Node.js) suivent les téléchargements. Le « *popularity contest* » de Debian piste les téléchargements du système d'exploitation libre Debian. Néanmoins, chaque gestionnaire de paquets est limité à un écosystème particulier, et aucun de ces gestionnaires ne peut donner une image du système dans son

ensemble. Plusieurs projets ne sont pas inclus dans un gestionnaire de paquets et ne sont pas suivis. Libraries.io, un site web créé par Andrew Nesbitt, est une tentative pour agréger des données des projets *open source* en fonction de leur usage, il piste environ 1,3 millions de bibliothèques *open source* sur 32 gestionnaires de paquets.

Travailler avec les projets pour moderniser l'organisation de travail

Beaucoup de projets sont en difficulté et pas seulement à cause d'un manque de financement, mais aussi parce qu'il est difficile d'y contribuer, ou encore parce qu'il existe un goulot d'étranglement au niveau des mainteneurs qui traitent les demandes de modification (*pull requests*) de la communauté. C'est vrai, en particulier, pour les plus anciens projets qui ont été bâtis avec des outils de développement, des langages et des processus qui ne sont plus populaires (ceux qui par exemple utilisent un système de contrôle de version autre que Git, dont la popularité croît chez les développeurs).

On peut faire beaucoup de choses pour faciliter la contribution à un projet, depuis la migration vers un flux de travail (*workflow*) plus moderne, le nettoyage du code, la fermeture des *pull request* délaissées, jusqu'à la mise en place d'une politique claire pour les contributions. Certains projets expérimentent pour rendre les contributions plus simples. Le développeur Felix Geisendörfer, par exemple, a suggéré que chaque personne qui soumet une modification du code devrait avoir une permission de *commit* afin de réduire l'engorgement au niveau de l'unique mainteneur vérifiant et approuvant ces changements. Felix a estimé que « cette approche est un fantastique moyen d'éviter que le projet ne se ratatine en transformant le projet d'un seul homme en celui d'une communauté. »

Le règlement de contribution de Node.js, qui peut être adopté

par les autres projets Node, met l'accent sur l'augmentation du nombre de contributeurs et sur leur autonomisation dans la prise de décision, plutôt que de désigner les mainteneurs comme seule autorité approbatrice. Leurs règles de contribution expliquent comment soumettre et valider des *pull requests*, comment consigner des bugs, etc. Les mainteneurs Node.js ont constaté qu'adopter de meilleures règles les avait aidés à gérer leur charge de travail et à faire évoluer leur communauté vers un projet plus sain et actif.



Photo par [Jez Nicholson](#) (CC BY-SA 2.0)

Dans un premier temps, il y a des recherches à faire pour déterminer quels projets doivent avancer. Autrement dit, à quoi ressemble un « projet à succès », aussi bien en termes de financement et de modèles de gouvernance, que dans l'équilibre à trouver entre mainteneurs, contributeurs et usagers ! La réponse peut varier en fonction des différents types de

projets et de leur ampleur.

Encourager les standards communs dans les projets open source

Bien que GitHub soit en train de devenir une plateforme standard pour la collaboration sur le code, de nombreux aspects des projets *open source* ne sont pas encore standardisés, notamment l'ampleur et la richesse de la documentation, des licences et des guides de contribution, ainsi que le style de code et le formatage.

Encourager l'adoption de standards de projets pourrait faciliter, pour les mainteneurs, la gestion des contributions, tout en réduisant pour les contributeurs les obstacles à la participation.

Parmi les exemples de standardisation croissante, on trouve le code de conduite, qui est un règlement détaillant les attentes en termes d'attitude et de communication.

Ces dernières années, des codes de conduite ont été adoptés par un nombre croissant de communautés de projets, notamment Node.js, Django et Ruby. Bien que le processus d'adoption ait pu donner lieu à d'intenses débats au sein de certaines communautés, leur prolifération révèle un intérêt croissant pour la responsabilisation du comportement des communautés.

Augmenter le nombre de contributeurs et contributrices open source

Comme nous l'avons évoqué dans un chapitre précédent de ce rapport, l'industrie du logiciel est florissante, avec un nombre croissant de nouveaux développeurs mais aussi d'autres talents variés : il y a du travail à faire pour encourager ces nouveaux arrivants à contribuer à l'*open source*. Augmenter le nombre de contributeurs permet aux projets *open source* d'être plus durables, car davantage de personnes participent à leur développement. Permettre à davantage de personnes de

contribuer à l'*open source* accroît également l'empathie et la communication entre les « utilisateurs » de l'*open source* et les projets dont ils dépendent.

« *Your First PR* » (« votre première PR », PR pour Pull Request, NdT) est un exemple d'initiative, développée par la programmeuse Charlotte Spencer, qui aide les nouveaux venus à effectuer leur première contribution à l'*open source*. « *First Timers Only* » (Réservé aux débutants) et « *Make a Pull Request* » (Faites une *pull request*) sont deux autres exemples de ressources populaires qui introduisent les nouveaux venus à l'*open source*. Certains projets *open source* utilisent également des étiquettes telles que « *first bug* » ou « *contributor friendly* » pour signaler les problèmes susceptibles d'être résolus par des contributeurs moins expérimentés. Il serait également bénéfique d'encourager les contributions à l'*open source* autres que le code, comme la rédaction de documentation technique, la gestion des tâches et des flux de travail, ou la création d'un site internet pour le projet.

En plus de l'augmentation de la proportion de techniciens talentueux contribuant à l'*open source* existe la possibilité de puiser dans un groupe de contributeurs plus large. Faire en sorte que les non-anglophones se sentent bienvenus dans les communautés *open source*, par exemple, pourrait rendre la technologie plus accessible à travers le monde. Et comme beaucoup de recruteurs utilisent les travaux *open source* comme un portfolio au moment d'embaucher un développeur, une communauté *open source* plus diverse encouragerait l'apparition d'un personnel technique globalement plus inclusif.

Améliorer les relations entre projets et acteurs extérieurs

Les entreprises sont une pièce incontournable de l'écosystème *open source*, et leur rôle ne fait que gagner en importance à

mesure que davantage d'entre elles adoptent les logiciels *open source*. Faciliter la collaboration entre entreprises et projets, ainsi qu'aider les entreprises à comprendre les besoins des communautés *open source*, pourrait débloquent le soutien des entreprises susceptibles de devenir mécènes ou promoteurs de l'*open source*.

Selon l'étude annuelle des entreprises *open source* réalisée par Black Duck, seulement 27 % des entreprises ont un règlement formel concernant les contributions de leurs employés à l'*open source*. Clarifier la possibilité ou non pour les employés de contribuer à l'*open source* sur leur temps de travail, et les encourager à le faire, pourrait grandement améliorer le soutien des entreprises aux projets *open source*.

En 2014, un groupement d'entreprises a fondé le TODO Group, pour partager les bonnes pratiques autour de la participation corporative à l'*open source*. Parmi les membres de ce groupe, on trouve Box, Facebook, Dropbox, Twitter et Stripe. En mars 2016, le TODO Group a annoncé qu'il serait hébergé par la Fondation Linux en tant que projet collaboratif.

Les entreprises peuvent également fournir un soutien financier aux projets, mais il est parfois difficile pour elles de trouver comment formaliser leur mécénat. Créer des budgets dédiés au mécénat en direction des équipes d'ingénieurs ou des employés, ou encore créer des documents permettant aux projets de « facturer » plus facilement leurs services aux entreprises, sont autant d'initiatives qui pourraient augmenter les contributions financières à l'*open source*.

Poul-Henning Kamp, par exemple, travaille sur un projet *open source* nommé Varnish, utilisé par un dixième des sites les plus visités d'internet, notamment Facebook, Twitter, Tumblr, The New York Times et The Guardian. Pour financer ce travail, il a créé la Varnish Moral License pour faciliter la sponsorisation du projet par les entreprises.

Même si en pratique la relation est un mécénat, Poul Henning utilise une terminologie familière aux entreprises, avec des termes tels que « facturation » et « licences », pour réduire les obstacles à la participation.

Augmenter le soutien aux compétences diverses et aux fonctions hors-codage

Dans un passé pas si lointain, les *startups* de logiciels étaient fortement centrées sur les compétences techniques. Les autres rôles, comme le marketing et le design, étaient considérés comme secondaires par rapport au code. Aujourd'hui, avec la création et la consommation rapide de logiciels, cette conception ne tient plus. Les *startups* sont en concurrence pour capter l'attention de leurs clients. L'identité de la marque est devenue l'un des principaux facteurs de différenciation.

Ces cinq dernières années ont été celles de l'essor du développeur *full stack* (polyvalent) : des développeurs plus généralistes que spécialisés, capables de travailler sur plusieurs domaines d'un logiciel complexe, et qui sont susceptibles d'avoir des compétences dans la conception et la production. Les équipes de développement sont plus soudées, elles utilisent les méthodes agiles avec des approches de conception d'architecture logicielle (où le livrable est élaboré en faisant des navettes entre les équipes de techniciens, designers et commerciaux), plutôt qu'une approche en cascade (où chaque équipe apporte sa pièce à l'édifice avant de la transmettre au groupe de travail suivant).

Les projets *open source* ont connu peu d'évolutions de ce genre, malgré notre dépendance croissante à leurs logiciels. On comprend aisément que le code soit au cœur d'un projet *open source*, il est en quelque sorte le « produit final » ou le livrable. Les fonctions relatives à la gestion de la communauté, à la documentation, ou à la promotion du projet,

qui sont la marque d'une organisation saine et durable, sont moins valorisées. Il en découle que les projets sont déséquilibrés. Beaucoup de choses pourraient être entreprises pour financer et soutenir les contributions autres que le code, des dons en nature pour payer les serveurs par exemple, ou des avantages comme une assurance maladie. Disposer de soutiens de ce type permettrait de réduire notablement la charge des développeurs.

Des routes et des ponts (15) – les institutions et l'open source

Voici le plus long des chapitres de [Des routes et des ponts](#) de [Nadia Ehgbal](#) que nous traduisons pour vous semaine après semaine ([si vous avez raté les épisodes précédents](#)). Il s'agit cette fois-ci des institutions (ici nord-américaines) qui par diverses formes de mécénat, contribuent au développement et au maintien des projets d'infrastructure numérique open source parce qu'elles y trouvent leur intérêt. Pas sûr qu'en Europe et en France ces passerelles et ces coopérations bien comprises entre entreprises et open source soient aussi habituelles...

Traduction Framalang : Opsylac, Luc, jums, xi, serici, lyn, mika, AFS, Goofy

Les efforts institutionnels pour

financer les infrastructures numériques

Il existe des institutions qui s'efforcent d'organiser collectivement les projets *open source* et aider à leur financement. Il peut s'agir de fondations indépendantes liées aux logiciels, ou d'entreprises de logiciels elles-mêmes qui apportent leur soutien.

Soutien administratif et mécénat financier

Plusieurs fondations fournissent un soutien organisationnel, comme le mécénat financier, aux projets *open source* : en d'autres termes, la prise en charge des tâches autres que le code, dont beaucoup de développeurs se passent volontiers. L'Apache Software Foundation, constituée en 1999, a été créée en partie pour soutenir le développement du serveur Apache HTTP, qui dessert environ 55 % de la totalité des sites internet dans le monde.

Depuis lors, la fondation Apache est devenue un foyer d'ancrage pour plus de 350 projets *open source*. Elle se structure comme une communauté décentralisée de développeurs, sans aucun employé à plein temps et avec presque 3000 bénévoles. Elle propose de multiples services aux projets qu'elle héberge, consistant principalement en un soutien organisationnel, juridique et de développement. En 2011, Apache avait un budget annuel de plus de 500 000 \$, issu essentiellement de subventions et de donations.

Le Software Freedom Conservancy, fondée en 2006, fournit également des services administratifs non-lucratifs à plus de 30 projets libres et *open source*. Parmi les projets que cette fondation soutient, on retrouve notamment Git, le système de contrôle de versions dont nous avons parlé plus haut et sur lequel GitHub a bâti sa plateforme, et Twisted, une librairie

Python déjà citée précédemment.

On trouve encore d'autres fondations fournissant un soutien organisationnel, par exemple The Eclipse Foundation et Software in the Public Interest. La Fondation Linux et la Fondation Mozilla soutiennent également des projets *open source* externes de diverses façons (dont nous parlerons plus loin dans ce chapitre), bien que ce ne soit pas le but principal de leur mission.

Il est important de noter que ces fondations fournissent une aide juridique et administrative, mais rarement financière. Ainsi, être sponsorisé par Apache ou par le Software Freedom Conservancy ne suffit pas en soi à financer un projet ; les fondations ne font que faciliter le traitement des dons et la gestion du projet.

Un autre point important à noter, c'est que ces initiatives soutiennent le logiciel libre et open source d'un point de vue philosophique, mais ne se concentrent pas spécifiquement sur ses infrastructures. Par exemple, OpenTripPlanner, projet soutenu par le Software Freedom Conservancy, est un logiciel pour planifier les voyages : même son code est *open source*, il s'agit d'une application destinée aux consommateurs, pas d'une infrastructure.



La coopération est nécessaire pour construire et maintenir une infrastructure – photo par [An en Alain](#) (licence CC By 2.0)

Créer une fondation pour aider un projet

Certains projets sont suffisamment importants pour être gérés à travers leurs propres fondations. Python, Node.js, Django et jQuery sont tous adossés à des fondations.

Il y a deux conditions fondamentales à remplir pour qu'une fondation fonctionne : accéder au statut d'exemption fiscale et trouver des financements.

Réussir à accéder au [statut 501\(c\)](#), la loi américaine qui définit les organismes sans but lucratif, peut s'avérer difficile pour ces projets, à cause du manque de sensibilisation autour de la technologie *open source* et de la tendance à voir l'*open source* comme une activité non-caritative. En 2013, une controverse a révélé que l'IRS

(Internal Revenue Service, service des impôts américain) avait dressé une liste de groupes postulant au statut d'exemption fiscale qui nécessiteraient davantage de surveillance : l'*open source* en faisait partie. Malheureusement, ces contraintes ne facilitent pas l'institutionnalisation de ces projets.

Par exemple, Russell Keith-Magee, qui était jusqu'à une époque récente président de la Django Software Foundation, a expliqué que la fondation ne pouvait pas directement financer le développement logiciel de Django, sans prendre le risque de perdre son statut 501(c). La fondation soutient plutôt le développement via des activités communautaires.

En juin 2014, la Fondation Yorba, qui a créé des logiciels de productivité qui tournent sous Linux, s'est vu refuser le statut 501(c) après avoir attendu la décision pendant presque quatre ans et demi. Jim Nelson, son directeur exécutif, a été particulièrement inquieté par le raisonnement de l'IRS : parce que leur logiciel pouvait potentiellement être utilisé par des entités commerciales, le travail de Yorba ne pouvait pas être considéré comme caritatif. Une lettre de l'IRS explique :

« Se contenter de publier sous une licence open source tous usages ne signifie pas que les pauvres et les moins privilégiés utiliseront effectivement les outils. [...] On ne peut pas savoir qui utilise les outils, et encore moins quel genre de contenus sont créés avec ces outils. »

Nelson a pointé les failles de ce raisonnement dans un billet de blog, comparant la situation à celle d'autres biens publics :

« Il y a une organisation caritative ici à San Francisco qui plante des arbres pour le bénéfice de tous. Si l'un de leurs arbres... rafraîchit les clients d'un café pendant qu'ils profitent de leur espresso, cela signifie-t-il que l'organisation qui plante des arbres n'est plus caritative ? »

Les projets qui accèdent au statut 501(c) ont tendance à insister sur l'importance de la communauté, comme la Python Software Foundation, dont l'objet est de « promouvoir, protéger et faire progresser le langage de programmation Python, ainsi que de soutenir et faciliter la croissance d'une communauté diversifiée et internationale de programmeurs Python ».

En parallèle, certains projets candidatent pour devenir une association de commerce au sens du [statut 501\(c\)\(6\)](#). La Fondation jQuery en est un exemple, se décrivant comme « une association de commerce à but non-lucratif pour développeurs web, financée par ses membres ». La Fondation Linux est également une association de commerce.

Le deuxième aspect de la formalisation de la gouvernance d'un projet à travers une fondation est la recherche de la source de financement adéquate. Certaines fondations sont financées par des donations individuelles, mais ont proportionnellement de petits budgets.

La Django Software Foundation, par exemple, gère Django, le plus populaire des frameworks web écrits en Python, utilisé par des entreprises comme Instagram et Pinterest. La Fondation est dirigée par des bénévoles, et reçoit moins de 60 000 \$ de donations par an. L'année dernière, la Django Software Foundation a reçu une subvention ponctuelle de la part de la Fondation Mozilla.

Parmi les autres sources habituelles de financement on trouve les entreprises mécènes. En effet, les entreprises privées sont bien placées pour financer ces projets logiciels, puisqu'elles les utilisent elles-mêmes. La Fondation Linux est l'un de ces cas particuliers qui rencontrent le succès, et ce grâce la valeur fondamentale du noyau Linux pour les activités de quasiment toutes les entreprises. La Fondation Linux dispose de 30 millions de dollars d'un capital géré sur une base annuelle, alimenté par des entreprises privées comme IBM,

Intel, Oracle et Samsung – et ce chiffre continue d'augmenter.

Créer une fondation pour soutenir un projet est une bonne idée pour les projets d'infrastructure très conséquents. Mais cette solution est moins appropriée pour de plus petits projets, en raison de la quantité de travail, des ressources, et du soutien constant des entreprises, nécessaires pour créer une organisation durable.

Node.js est un exemple récent d'utilisation réussie d'une fondation pour soutenir un gros projet. Node.js est un *framework* JavaScript, développé en 2009 par Ryan Dahl et différents autres développeurs employés par Joyent, une entreprise privée du secteur logiciel. Ce *framework* est devenu extrêmement populaire, mais a commencé à souffrir de contraintes de gouvernance liées à l'encadrement par Joyent, que certaines personnes estimaient incapable de représenter pleinement la communauté enthousiaste et en pleine croissance de Node.js.

En 2014, un groupe de contributeurs de Node.js menaça de *forker* le projet. Joyent essaya de gérer ces problèmes de gouvernance en créant un conseil d'administration pour le projet, mais la scission eut finalement lieu, le nouveau *fork* prenant le nom d'io.js. En février 2015 fut annoncée l'intention de créer une organisation 501(c) (6) en vue d'extraire Node.js de la mainmise de Joyent. Les communautés Node.js et io.js votèrent pour travailler ensemble sous l'égide de cette nouvelle entité, appelée la Fondation Node.js. La Fondation Node.js, structurée suivant les conseils de la Fondation Linux, dispose d'un certain nombre d'entreprises mécènes qui contribuent financièrement à son budget, notamment IBM, Microsoft et PayPal. Ces sponsors pensent retirer une certaine influence de leur soutien au développement d'un projet logiciel populaire qui fait avancer le web, et ils ont des ressources à mettre à disposition.



Montage d'une yourte, photo par [Armel](#) (CC BY SA 2.0)

Un autre exemple prometteur est Ruby Together, une organisation initiée par plusieurs développeurs Ruby pour soutenir des projets d'infrastructure Ruby. Ruby Together est structuré en tant qu'association commerciale, dans laquelle chaque donateur, entreprise ou individu, investit de l'argent pour financer le travail à temps plein de développeurs chargés d'améliorer le cœur de l'infrastructure Ruby. Les donateurs élisent un comité de direction bénévole, qui aide à décider chaque mois sur quels projets les membres de Ruby Together devraient travailler.

Ruby Together fut conçue par deux développeurs et finance leur travail de : André Arko et David Radcliffe. Aujourd'hui, en avril 2016, est également rémunéré le travail de quatre autres mainteneurs d'infrastructure. Le budget mensuel en mars 2016 était d'un peu plus de 18 000 dollars par mois, couvert entièrement par des dons. La création de Ruby Together fut annoncée en mars 2015 et reste un projet récent, mais pourrait bien servir de base à un modèle davantage orienté vers la communauté pour financer la création d'autres projets d'infrastructure.

Programmes d'entreprises

Les éditeurs de logiciels soutiennent les projets d'infrastructure de différentes manières.

En tant que bénéficiaires des projets d'infrastructures, ils contribuent en faisant remonter des dysfonctionnements et des bugs, en proposant ou soumettant de nouvelles fonctionnalités ou par d'autres moyens. Certaines entreprises encouragent leurs employés à contribuer à des projets d'une importance critique sur leur temps de travail. De nombreux employés contribuent ainsi de manière significative à des projets *open source* extérieurs à l'entreprise. Pour certains employés, travailler sur de l'*open source* fait clairement partie de leur travail. L'allocation de temps de travail de leurs salariés est une des plus importantes façons de contribuer à l'*open source* pour les entreprises.

Les grandes entreprises comme Google ou Facebook adhèrent avec enthousiasme à l'*open source*, de façon à inspirer confiance et renforcer leur influence ; elles sont de fait les seuls acteurs institutionnels assez importants qui peuvent assumer son coût sans avoir besoin d'un retour financier sur investissement. Les projets *open source* aident à renforcer l'influence d'une entreprise, que ce soit en publiant son propre projet *open source* ou en embauchant des développeurs de premier plan pour qu'ils travaillent à plein temps sur un projet *open source*.

Ces pratiques ne sont pas limitées aux entreprises purement logicielles. Walmart, par exemple, qui est un soutien majeur de l'*open source*, a investi plus de deux millions de dollars dans un projet *open source* nommé hapi. Eran Hammer, développeur senior à Walmart Labs, s'est empressé de préciser que « l'*open source*, ce n'est pas du caritatif » et que les ressources d'ingénierie gratuites sont proportionnelles à la taille des entreprises qui utilisent hapi. Dion Almaer, l'ancien vice-président en ingénierie de Walmart Labs, a

remarqué que leur engagement envers l'*open source* les aidait à recruter, à construire une solide culture d'entreprise, et à gagner « une série d'effets de levier ».

En termes de soutien direct au maintien du projet, il arrive que des entreprises embauchent une personne pour travailler à plein temps à la maintenance d'un projet *open source*. Les entreprises donnent aussi occasionnellement à des campagnes de financement participatif pour un projet particulier. Par exemple, récemment, une campagne sur Kickstarter pour financer un travail essentiel sur Django a reçu 32 650 £ (environ 40 000 €) ; Tom Christie, l'organisateur de la campagne, a déclaré que 80 % du total venait d'entreprises. Cependant, ces efforts sont toujours consacrés à des projets spécifiques et les infrastructures numériques ne sont pas encore vues communément comme une question de responsabilité sociale par les entreprises de logiciel à but lucratif. Cela laisse encore beaucoup de marge aux actions de défense et promotion.

L'un des programmes d'entreprise les plus connus est le *Summer of Code* de Google (été de programmation, souvent nommé GSoC), déjà mentionné dans ce livre, qui offre de l'argent à des étudiant·e·s pour travailler sur des projets *open source* pendant un été. Les étudiant·e·s sont associé·e·s à des mentors qui vont les aider à se familiariser avec le projet. Le Summer of Code est maintenu par le bureau des programmes *open source* de Google, et il a financé des milliers d'étudiant·e·s.

Le but du Summer of Code est de donner à des étudiants la possibilité d'écrire du code pour des projets *open source*, non de financer les projets eux-mêmes.

L'an dernier, Stripe, une entreprise de traitement des paiements, a annoncé une « retraite *open source* », offrant un salaire mensuel d'un maximum de 7500 dollars pour une session de trois mois dans les locaux de Stripe. À l'origine, l'entreprise voulait uniquement offrir deux bourses, mais

après avoir reçu 120 candidatures, le programme a été ouvert à quatre bénéficiaires.

Ces derniers ont été enchantés par cette expérience. L'un d'entre eux, Andrey Petrov, continue de maintenir la bibliothèque Python `urllib3` dont nous avons déjà parlé, et qui est largement utilisée dans l'écosystème Python.

À propos de cette expérience, Andrey a écrit :

« La publication et la contribution au code open source vont continuer que je sois payé pour ou non, mais le processus sera lent et non ciblé. Ce qui n'est pas un problème, car c'est ainsi que l'open source a toujours fonctionné. Mais on n'est pas obligé d'en rester là. [...] »

Si vous êtes une entreprise liée à la technologie, allouez s'il vous plaît un budget pour du financement et des bourses dans le domaine de l'open source. Distribuez-le sur Gittip [Note : Gittip est maintenant dénommé Gratipay. Le produit a été quelque peu modifié depuis la publication originelle du billet d'Andrew] si vous voulez, ou faites ce qu'a fait Stripe et financez des sprints ambitieux pour atteindre des objectifs de haute valeur.

Considérez ceci comme une demande solennelle de parrainage : s'il vous plaît, aidez au financement du développement d'urllib3. »

La retraite *open source* de Stripe peut servir de modèle aux programmes de soutien. Stripe a décidé de reconduire le programme pour une deuxième année consécutive en 2015. Malgré la popularité de leur programme et la chaude réception qu'il a reçue chez les développeurs et développeuses, cette pratique n'est toujours pas répandue dans les autres entreprises.

Les entreprises montrent un intérêt croissant pour l'*open source*, et personne ne peut prédire au juste ce que cela

donnera sur le long terme. Les entreprises pourraient régler le problème du manque de support à long terme en consacrant des ressources humaines et un budget aux projets *open source*. Des programmes de bourse formalisés pourraient permettre de mettre en contact des entreprises avec des développeurs *open source* ayant besoin d'un soutien à plein temps. Alors que les équipes de contributeurs à un projet étaient souvent composées d'une diversité de développeurs venant de partout, peut-être seront-elles bientôt composées par un groupe d'employés d'une même entreprise. Les infrastructures numériques deviendront peut-être une série de « jardins clos », chacun d'entre eux étant techniquement ouvert et bénéficiant d'un soutien solide, mais en réalité, grâce à ses ressources illimitées, une seule entreprise et de ses employés en assureront le soutien.

Mais si on pousse la logique jusqu'au bout, ce n'est pas de très bon augure pour l'innovation. Jeff Lindsay, un architecte logiciel qui a contribué à mettre en place l'équipe de Twilio, une entreprise performante de solutions de communication dans le *cloud*, livrait l'an dernier ses réflexions dans une émission :

« À Twilio, on est incité à améliorer le fonctionnement de Twilio, à Amazon on est incité à améliorer le fonctionnement d'Amazon. Mais qui est incité à mieux les faire fonctionner ensemble et à offrir plus de possibilités aux usagers en combinant les deux ? Il n'y a personne qui soit vraiment incité à faire ça. »

Timothy Fuzz, un ingénieur système, ajoute :

« Pour Bruce Schneier, cette situation tient du servage. Nous vivons dans un monde où Google est une cité-état, où Apple est une cité-état et... si je me contente de continuer à utiliser les produits Google, si je reste confiné dans l'environnement Google, tout me paraît bénéfique. Mais il est quasi impossible de vivre dans un monde où je change

d'environnement : c'est très pénible, vous tombez sur des bugs, et aucune de ces entreprises ne cherche vraiment à vous aider. Nous sommes dans ce monde bizarre, mais si vous regardez du côté des cités-états, l'un des problèmes majeurs c'est le commerce inter-étatique : si on doit payer des droits de douane parce qu'on cherche à exporter quelque chose d'Austin pour le vendre à Dallas, ce n'est pas un bon modèle économique. On pâtit de l'absence d'innovation et de partage des idées. On en est là, aujourd'hui. »

Bien que l'argument du « servage » se réfère généralement aux produits d'une entreprise, comme l'addiction à l'iPhone ou à Android, il pourrait être tout aussi pertinent pour les projets *open source* parrainés. Les améliorations prioritaires seront toujours celles qui bénéficient directement à l'entreprise qui paie le développeur. Cette remarque ne relève pas de la malveillance ou de la conspiration : simplement, être payé par une entreprise pour travailler à un projet qui ne fait pas directement partie de ses affaires est une contrainte à prendre en compte.

Mais personne, pas plus Google que la Fondation Linux ou qu'un groupe de développeurs indépendants, ne peut contrôler l'origine d'un bon projet *open source*. Les nouveaux projets de valeur peuvent germer n'importe où, et quand ils rendent un service de qualité aux autres développeurs, ils sont largement adoptés. C'est une bonne chose et cela alimente l'innovation.

Aide spécifique de fondation

Deux fondations ont récemment fait part de leur décision de financer plus spécifiquement l'infrastructure numérique : la Fondation Linux et la Fondation Mozilla.

Après la découverte de la faille Heartbleed, la Fondation Linux a annoncé qu'elle mettait en place l'Initiative pour les infrastructures essentielles (Core Infrastructure Initiative,

CII) pour éviter que ce genre de problème ne se reproduise. Jim Zemlin, le directeur-général de la Fondation Linux, a réuni près de 4 millions de dollars en promesses de dons provenant de treize entreprises privées, dont Amazon Web Services, IBM et Microsoft, pour financer des projets liés à la sécurité des infrastructures pour les trois ans à venir. La Fondation Linux s'occupe également d'obtenir des financements gouvernementaux, y compris de la Maison-Blanche.

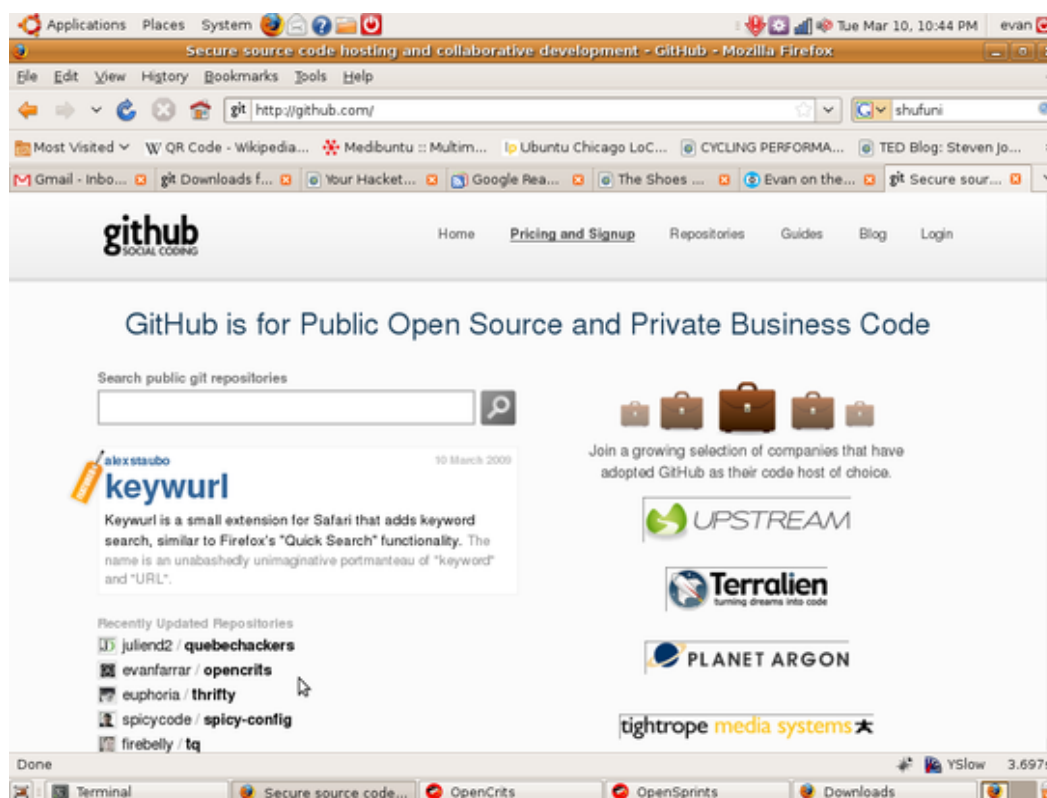
La CII est officiellement un projet de la fondation Linux. Depuis sa création en avril 2014, la CII a sponsorisé du travail de développement d'un certain nombre de projets, dont OpenSSL, NTP, GnuPG (un système de chiffrement des communications) et OpenSSH (un ensemble de protocoles relatifs à la sécurité). La CII se concentre en priorité sur une partie de l'infrastructure numérique : les projets relatifs à la sécurité.

Au mois d'octobre 2015, Mitchell Baker, la présidente de la Fondation Mozilla, a annoncé la création du Programme de soutien à l'*open source* de Mozilla (Mozilla Open Source Support Program, MOSS) et a promis de consacrer un million de dollars au financement de logiciels libres et *open source*. Selon Baker, ce programme aura deux volets : un volet « rétribution » pour les projets qu'utilise Mozilla et un volet « contribution » pour les projets libres et *open source* en général. Grâce aux suggestions de la communauté, Mozilla a sélectionné neuf projets pour la première série de bourses. Ils se disent également prêts à financer des audits de sécurité pour les projets *open source* importants.

Enfin, certaines fondations contribuent ponctuellement à des projets de développement logiciel. Par exemple, la Python Software Foundation propose aux individus et aux associations des bourses modestes destinées pour la plupart aux actions pédagogiques et de sensibilisation.

Autres acteurs institutionnels

Il existe plusieurs autres acteurs qui apportent diverses formes de soutien aux infrastructures numériques : Github, le capital-risque et le monde universitaire. Si Facebook est un « utilitaire social » et Google un « utilitaire de recherche », tous deux régulant *de facto* les corps dans leur domaine respectif – alors Github a une chance de devenir « l'utilitaire *open source* ». Son modèle économique l'empêche de devenir un mastodonte financier (contrairement à Facebook ou Google dont le modèle est basé sur la publicité, alors que Github se monétise par l'hébergement de code pour les clients professionnels, et par l'hébergement individuel de code privé), mais Github est toujours un endroit où aujourd'hui encore l'*open source* est créée et maintenue.



GitHub s'adresse aux entreprises aussi – Image par [Evan](#) (licence CC BY 2.0)

GitHub s'est doté de grandes aspirations avec une levée de fonds de capital-risque de 350 millions de dollars, même si l'entreprise était déjà rentable. Si Github assume pleinement

son rôle d'administrateur du code *open source*, l'organisation peut avoir une énorme influence sur le soutien apporté à ces projets. Par exemple, elle peut créer de meilleurs outils de gestion de projets *open source*, défendre certaines catégories de licences, ou aider les gestionnaires de projets à gérer efficacement leurs communautés.

Github a subi de grosses pressions venant des développeurs qui gèrent certains projets, ces pressions incluent [une lettre ouverte collective](#) intitulée « Cher Github », principalement issue de la communauté Javascript. Cette lettre explique : « Beaucoup sont frustrés. Certains parmi nous qui déploient des projets très populaires sur Github se sentent totalement ignoré par vous ». La lettre inclut une liste de requêtes pour l'amélioration de produits, qui pourrait les aider à gérer plus efficacement leurs projets.

Github se confronte de plus en plus à des difficultés largement documentées dans les médias. Auparavant, l'entreprise était connue pour sa hiérarchie horizontale, sans aucun manager ni directive venant d'en haut. Les employés de Github avaient aussi la liberté de choisir de travailler sur les projets qu'ils souhaitaient. Ces dernières années, tandis que Github s'est développée pour atteindre presque 500 employés, l'entreprise a réorienté sa stratégie vers une orientation plus commerciale en recrutant des équipes de vente et des dirigeants, insérés dans un système hiérarchique plus traditionnel. Cette transition d'une culture décentralisée vers plus de centralité s'est faite dans la douleur chez Github : au moins 10 dirigeants ont quitté l'organisation durant les quelques mois de l'hiver 2015-2016, ces départs incluant l'ingénieur en chef, le directeur des affaires financières, le directeur stratégique et le directeur des ressources humaines. En raison de ces conflits internes, Github n'a toujours pas pris position publiquement pour jouer un rôle de promoteur de l'*open source* et assumer un leadership à même de résoudre les questions pressantes autour de l'*open*

source, mais le potentiel est bel et bien là.

Pour le capital-risque, abordé précédemment, il y a un enjeu particulier dans l'avenir des infrastructures numériques. Comme les outils des développeurs aident les entreprises du secteur technologique à créer plus rapidement et plus efficacement, meilleurs sont les outils, meilleures sont les *startups*, meilleure sera la rentabilité du capital-risque. Néanmoins, l'infrastructure, d'un point de vue capitaliste, n'est en rien limitée à l'*open source* mais plus largement focalisée sur les plateformes qui aident d'autres personnes à créer. C'est pour cela que les investissements dans Github ou npm, qui sont des plateformes qui aident à diffuser du code source, ont un sens, mais tout aussi bien les investissements dans Slack, une plateforme de travail collaboratif que les développeurs peuvent utiliser pour construire des applications en ligne de commande connectées à la plateforme (à ce propos, le capital-risque a constitué un fonds de 80 millions dédié au support de projets de développement qui utilisent Slack). Même si le capital-risque apprécie les mécaniques sous-jacentes de l'infrastructure, il est limité dans ses catégories d'actifs : un capitaliste ne peut pas investir dans un projet sans modèle économique.

Enfin, les institutions universitaires ont joué un rôle historique éminent dans le soutien aux infrastructures numériques, tout particulièrement le développement de nouveaux projets. Par exemple, LLVM, un projet de compilateur pour les langages C et C++, a démarré en tant que projet de recherche au sein de l'Université de l'Illinois, à Urbana-Champaign. Il est maintenant utilisé par les outils de développement de MacOS X et iOS d'Apple, mais aussi dans le kit de développement de la Playstation 4 de Sony.

Un autre exemple, R, un langage de programmation répandu dans la statistique assistée par ordinateur et l'analyse de données, a été d'abord écrit par Robert Gentleman et Ross Ihaka à l'Université d'Auckland. R n'est pas uniquement

utilisé par des entreprises logicielles comme Facebook ou Google, mais aussi par la Bank of America, l'Agence américaine des produits alimentaires et médicamenteux et le Service météorologique national américain, entre autres.

Quelques universités emploient également des programmeurs qui ont alors la liberté de travailler à des projets *open source*. Par exemple, le protocole d'heure réseau ou NTP (Network Time Protocol) utilisé pour synchroniser le temps via Internet, fut d'abord développé par David Mills, maintenant professeur émérite de l'université du Delaware – le projet continuant à être maintenu par un groupe de volontaires conduit par Harlan Stenn. Bash, l'outil de développement dont nous parlions dans un chapitre précédent, est actuellement maintenu par Chet Ramsay, qui est employé par le département des technologies de l'information de l'université Case Western.

Les institutions universitaires ont le potentiel pour jouer un rôle important dans le soutien de nouveaux projets, parce que cela coïncide avec leurs missions et types de donation, mais elles peuvent aussi manquer de la réactivité nécessaire pour attirer les nouveaux programmeurs *open source*. NumFOCUS est un exemple d'une fondation 501(c)(3) qui soutient les logiciels scientifiques *open source* à travers des donations et parrainages financiers. Le modèle de la fondation externe peut aider à fournir le soutien dont les logiciels scientifiques ont besoin dans un contexte d'environnement universitaire. Les fondations Alfred P. Sloan et Gordon & Betty Moore expérimentent aussi des manières de connecter les institutions universitaires avec les mainteneurs de logiciels d'analyse des données, dans le but de soutenir un écosystème ouvert et durable.