

## À QUI PROFITE LE BIG DATA ?

Dans un passé encore récent, on inscrivait ses rendez-vous sur un agenda ou sur un calendrier. En vacances, on envoyait des cartes postales. Nos vaccins étaient consignés sur un carnet de santé...L'arrivée de l'informatique et la numérisation ont chamboulé ces usages.

### Ressources

- Dominique Cardon, *À quoi rêvent les algorithmes, nos vies à l'heure des big data*, Éditions du Seuil, 2015.
- Gilles Dowek, Serge Abitboul, *Le temps des algorithmes*, Éditions du Pommier, 2017.  
Ouvrage pédagogique sur le sujet  
<https://www.editions-lepommier.fr/le-temps-des-algorithmes>
- Michaël Keller, Josh Neufeld, *Dans l'ombre de la peur*, Éditions Ça et Là, 2017.  
Une BD sur les données personnelles et leurs utilisations  
<http://www.caetla.fr/Dans-l-ombre-de-la-peur-le-Big-Data-et-nous>
- Cathy O'Neil, *Weapons of Math Destruction*, Crown Random House, 2016 (non traduit).  
Ouvrage sur l'évolution des algorithmes et leur dangers  
<https://weaponsofmath-destructionbook.com>
- *Le P'tit Libé* sur les données personnelles, n° 52, avril 2018  
<https://ptitlibe.liberation.fr/les-donnees-personnelles,100943>
- *Terra Data, nos vies à l'ère numérique*, 2018.  
Catalogue très intéressant d'une exposition sur la DATA  
<http://www.cite-sciences.fr/fr/au-programme/expos-temporaires/terra-data/>

Aujourd'hui, de plus en plus d'informations sont stockées dans des serveurs, via d'immenses ordinateurs connectés les uns aux autres. Non seulement les informations que nous consignons auparavant sur du papier, mais aussi les traces de notre activité sur Internet : messages sur les réseaux sociaux, visites de sites, photographies partagées... On appelle toutes ces informations des données.

Le nombre de données créées et utilisées chaque année est de plus en plus important. On en crée sans s'en rendre compte, avec des téléphones qui comptent les pas ou quand on emprunte un vélo en libre-service. De nouveaux métiers sont apparus pour gérer toutes ces données, mais aussi pour les exploiter. On a en effet découvert qu'on pouvait apprendre beaucoup de leur étude. Ainsi, en 2008, des chercheurs ont analysé les recherches faites sur Google par un échantillon d'Américains et ont essayé de prédire quand la grippe allait arriver. Ils ont développé un algorithme, une formule informatique, grâce à laquelle ils ont pu comparer l'évolution de la grippe des années précédentes avec celle des recherches sur Google. Ils ont ainsi repéré quand les conditions semblaient identiques. Pendant plusieurs années, Google Flu Trends, le nom de l'algorithme, a permis de prévoir la grippe avant les médecins<sup>1</sup>.

Les algorithmes sont aujourd'hui présents partout. Beaucoup de décisions sont prises à partir de ces formules informatiques. Certaines permettent de découvrir le profil des internautes en fonction de leur activité en ligne, afin de leur proposer des publicités ciblées. D'autres sont utilisées par Netflix pour recommander de nouvelles séries à ses utilisateurs, adaptant même l'image de présenta-

tion à leur profil<sup>2</sup>. En France, des algorithmes développés par la gendarmerie nationale tentent de prédire où auront lieu les prochains cambriolages<sup>3</sup> ou des vols de voitures<sup>4</sup>. Ce sont aussi des formules informatiques qui génèrent l'attribution des places dans les universités, grâce à l'application ParcoursSup<sup>5</sup>.

Des algorithmes, qui apprennent des situations précédentes pour prévoir ce qui pourrait se passer, existent dans différents domaines : identification de personnes sur des photos, sur Facebook par exemple<sup>6</sup>, ou détection de cellules malades sur des images d'IRM<sup>7</sup>.

De nombreuses découvertes ont été possibles grâce à ces immenses bases de données et à leur manipulation par des algorithmes. Ceux-ci sont souvent présentés comme une solution idéale et objective. Il faut néanmoins rester vigilant. Déjà certains algorithmes apprennent et décident à partir de situations existantes qui peuvent être biaisées. Ainsi, les algorithmes de Google proposent-ils des offres d'emploi. Mais une étude a montré que les postes les plus prestigieux étaient proposés aux hommes plutôt qu'aux femmes. Une situation logique, puisque ce sont actuellement les hommes qui occupent les fonctions d'encadrement, mais une situation de discrimination anormale, si l'on considère que tout le monde devrait avoir un accès égal à ces emplois. Les formules mathématiques font aussi des erreurs – on parle de faux positifs<sup>8</sup> – souvent sans conséquences graves. Mais si un algorithme de détection des visages est utilisé par la police et se trompe, il faut être sûr qu'il ne soit pas le seul à décider de l'arrestation !

Alexandre Léchenet, journaliste indépendant, spécialisé dans l'analyse des données

1. <https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/>

2. <https://www.numarama.com/pop-culture/433922-netflix-trompe-t-il-ses-utilisateurs-noirs-avec-ses-vignettes-de-films-personnalisees.html>

3. [https://www.lesechos.fr/30/06/2018/lesechos.fr/0301881430259\\_quand-la-gendarmerie-utilise-les-algorithmes-pour-prevoir-les-cambriolages.htm](https://www.lesechos.fr/30/06/2018/lesechos.fr/0301881430259_quand-la-gendarmerie-utilise-les-algorithmes-pour-prevoir-les-cambriolages.htm)

4. <https://agd.data.gouv.fr/2018/01/12/predire-les-vols-de-voitures/>

5. [https://www.sciencesetavenir.fr/high-tech/informatique/bac-2018-l-algorithme-de-parcoursup-explique-par-les-deux-chercheurs-qui-l-ont-concu\\_124407](https://www.sciencesetavenir.fr/high-tech/informatique/bac-2018-l-algorithme-de-parcoursup-explique-par-les-deux-chercheurs-qui-l-ont-concu_124407)

6. <https://gizmodo.com/facebooks-new-face-recognition-features-what-we-do-an-1823359911>

7. <https://presse.inserm.fr/un-programme-informatique-capable-de-detecter-et-didentifier-automatiquement-des-lesions-cerebrales/31634/>

8. <https://www.theguardian.com/uk-news/2014/dec/02/youre-the-bomb-are-you-at-risk-from-anti-terrorism-algorithms-automated-tracking-innocent-people>

# LES TRACES, L'OMBRE DE L'IDENTITÉ NUMÉRIQUE

L'identité numérique se construit à partir des traces volontaires ou involontaires, subies ou choisies, que laissent les internautes lors de leurs activités sur Internet (recherche et publication), particulièrement quand ils utilisent les réseaux sociaux et les objets connectés.

## Objectifs

- Comprendre les notions d'identité et de trace numériques (ÉMI cycle 4)
- Se familiariser avec les notions d'espace privé et d'espace public (ÉMI cycle 4)

## Ressources

- Michel Arnaud et Louise Merzeau, *Hermès - Traçabilité et réseaux*, CNRS, 2009.
- Dominique Cardon, *Le design de la visibilité*, FYP Éditions, 2009.
- Antonio Casili, *Être présent en ligne : culture et structure des réseaux sociaux d'Internet*, Idées économiques et sociales, 2012.

## DÉROULEMENT

### Vue numérique et vie privée

Utiliser le jeu « Médiasphère » comme outil de médiation, et notamment les questions de couleur jaune « Ma vie numérique », pour mettre en place un débat autour des notions de vie privée, vie publique, vie professionnelle et de l'influence qu'ont les traces numériques dans l'interconnexion de ces sphères.

Faire comprendre que les sphères privée, publique et professionnelle s'interconnectent à l'heure d'Internet et des réseaux sociaux et que ces interconnexions sont dues aux traces que nous laissons sur les réseaux.

### Découvrir la notion de traces numériques

#### Traces de navigation

Sur le navigateur Firefox, installer l'extension « Lightbeam », puis demander aux élèves d'effectuer une série de recherches sur différents sites. Une fois les recherches terminées, ouvrir l'extension « Lightbeam » pour visualiser sur un graphique les interactions de la navigation.

Le graphique met en évidence que ces interactions se répartissent en deux catégories : les sites visités par les élèves (représentés par des cercles) et les sites tiers connectés aux sites visités via des cookies (représentés par des triangles). Faire comprendre que ces traces relèvent de deux modes de navigation : volontaire (sites visités) et involontaire (sites tiers).

#### Traces de publication

Demander aux élèves, à l'aide du moteur de recherche Webmii, de réaliser une requête sur une personnalité. Une fois la page des résultats affichée, naviguez sur les différents profils proposés (Facebook, Twitter, etc.)

Faire identifier par les élèves les trois types de traces : traces liées au profil (créées lors de la création du profil), traces de publication (textes, images, vidéos, partages...) et traces de publication par des tiers (likes, commentaires, partages...). Faire comprendre que ces traces sont de deux types : volontaires (publication/profil) et subies ou héritées (publication de tiers).

### Traces liées aux objets connectés

Demander aux élèves d'utiliser la simulation interactive « Vivez une journée de données ».

Mettre en évidence que nos objets connectés du quotidien (smartphone, carte bleue, badge de télépéage, carte de transport, montre connectée...) collectent des traces de différents types (localisation, santé, surveillance). Faire comprendre que les objets connectés permettent de retracer le quotidien de tout le monde.

### Liens entre traces et identité numérique

Regarder la vidéo « Dave le voyant », sous-titrée en français sur YouTube.

Demander aux élèves de réaliser le portrait des quidams ayant participé à l'expérience de Dave le voyant. Classer les traces selon les sphères auxquelles elles appartiennent (privée, publique, professionnelle). Engager une réflexion critique : quelles exploitations peuvent être faites des traces numériques collectées ? Faire comprendre que les traces collectées d'un individu peuvent permettre de réaliser son portrait fidèle.

## PROLONGEMENTS

Réaliser un album de fin d'études (« Yearbook »), réel ou virtuel, dans lequel apparaîtra le portrait numérique de chaque élève d'une classe. Les portraits seront constitués uniquement à partir des traces volontaires ou non, subies ou non.

Ce travail, mené sur la durée, permet aux élèves de réinvestir les notions apprises sur les traces numériques, à savoir les différentes sphères de la vie numérique, les différents types de traces et le lien entre traces et identité numérique.

Sylvain Joseph, formateur,  
Pôle Labo, CLEMI

# LE BIG DATA, MINERAI DU XXI<sup>e</sup> SIÈCLE

Usages des réseaux sociaux, requêtes sur des moteurs de recherche, achats en ligne, mails, écoute de musique... toutes ces activités sur Internet génèrent chaque jour des milliards de données et forment ce que l'on appelle le big data. Analysées et exploitées, sans cesse actualisées, ces données représentent potentiellement d'immenses sources de revenus.

## Objectifs

- Prendre conscience de la production de données lors de nos usages du numérique.
- Savoir maîtriser ses données personnelles.
- Analyser le lien entre gratuité des services internet et utilisation des données personnelles.

## Ressources

- [www.caminteresse.fr/economie-societe/qui-exploite-nos-donnees-personnelles-1197308/](http://www.caminteresse.fr/economie-societe/qui-exploite-nos-donnees-personnelles-1197308/)
- [www.challenges.fr/high-tech/vos-donnees-personnelles-sur-internet-peuvent-valoir-de-1-or\\_57690](http://www.challenges.fr/high-tech/vos-donnees-personnelles-sur-internet-peuvent-valoir-de-1-or_57690)
- [www.cnil.fr/fr/maitriser-mes-donnees](http://www.cnil.fr/fr/maitriser-mes-donnees)
- [www.mesdatasetmoi-observatoire.fr/](http://www.mesdatasetmoi-observatoire.fr/)
- <https://www.webmarketing-com.com/2013/05/24/21068-les-business-models-de-google-et-facebook>

## DÉROULEMENT

### Le modèle économique des GAFAM

Les élèves travaillent en 2 groupes. Un groupe prend en charge le modèle économique de Google, l'autre celui de Facebook. Les élèves recensent les services que ces derniers mettent à disposition des utilisateurs et réfléchissent aux raisons de leur succès (gratuité, simplicité et fédération d'identité). Ensuite, à partir d'une recherche d'information, ils décryptent le fonctionnement des régies publicitaires affiliées à ces plateformes.

Faire rechercher aux élèves les conditions générales d'utilisation (CGU) des réseaux sociaux et des applications qu'ils utilisent. Puis leur faire créer une carte mentale présentant tous les types d'informations personnelles auxquelles ils ont accès, au-delà de ce que chacun échange via photos, vidéos et messages textes.

Quels sont les liens existant entre ces services fournis et le modèle économique de ces géants de l'Internet ? *Ils collectent les données et métadonnées des utilisateurs à des fins de ciblage publicitaire.* Quelles sont les limites de la gratuité sur Internet ?

### Valorisation financière des données personnelles

Demander aux élèves d'analyser les publicités qu'ils reçoivent personnellement sur les réseaux sociaux ou via leurs navigateurs (nature du produit, nom des marques). Analyser les résultats en comparant les ressemblances et les différences pour appréhender le champ des publicités ciblées.

De quelle nature sont les publicités ciblées, personnalisées ? Les publicités sont-elles en relation avec le genre, les loisirs, le sport pratiqué, les centres d'intérêts, les goûts commerciaux (achats sur Internet ou likes de produits...). Sont-elles liées à la géolocalisation ?

### Valeur de ces données

À l'aide des ressources à disposition ou d'autres qu'auront recherchées les élèves, recenser les informations chiffrées sur les recettes liées aux données personnelles (prix des ventes, chiffres d'affaires, bénéfices...).

Construire un schéma logique qui présente le circuit des données personnelles et les acteurs qui s'enrichissent grâce à celles-ci et faire prendre conscience que ce ne sont pas les données brutes mais des données retravaillées, croisées et analysées qui créent de la valeur.

### Le nouveau règlement européen RGPD

Demander aux élèves de trouver ce qui a changé pour les citoyens européens, depuis mai 2018, en matière de données personnelles. Est-il désormais possible de s'enrichir en vendant soi-même ses propres données ?

### Comment limiter le phénomène ?

Faire découvrir aux élèves la diversité des moteurs de recherche et notamment leur faire recenser ceux qui ne collectent pas de données personnelles, soit des moteurs de recherche alternatifs (Duck duck go, Qwant, Framabee...).

Apprendre aux élèves à effacer l'historique de navigation et les cookies pour réduire les traces, via les paramètres avancés des moteurs de recherche.

## PROLONGEMENT

**Débat :** « Si c'est gratuit c'est toi le produit » ; est-il possible de sortir de cette réalité aujourd'hui ? Que sommes-nous prêts à payer ?

<https://www.20minutes.fr/high-tech/2266675-20180505-pourquoi-version-payante-facebook-possible-futur-lointain>

Nathalie Barbéry, coordonnatrice académique CLEMI, académie Dijon

# QUAND LES DONNÉES PERSONNELLES S'ÉCHAPPENT : L'AFFAIRE « CAMBRIDGE ANALYTICA »

En mars 2018, le quotidien américain *The New York Times* et le britannique *The Guardian* ont révélé comment une entreprise fondée à Londres, et inconnue du grand public jusque-là, a recueilli les données de dizaines de millions d'Américains sans leur consentement début 2014. Quelques mois plus tard, ce sont près de 50 millions de comptes utilisateurs de Facebook qui ont été piratés. Des affaires qui relancent la question de la protection des données personnelles.

## Ressources

- Marie Adam-Normand (CLEMI Besançon) et Guillaume Bonzoms (DANE Besançon), « Connais-moi, échappe-toi », un jeu d'évasion autour des données personnelles. En ligne : <https://dane.ac-besancon.fr/connais-moi-echappe-toi-evasion/>
- Data Gueule, « Big data : données, données, donnez-moi ! » #DATAGUEULE 15. <https://youtu.be/5otaBKsz7k4>
- Thomas Huchon, « Comment Trump a manipulé l'Amérique ? », Arte, 2018. <https://www.arte.tv/fr/videos/082806-000-A/comment-trump-a-manipule-l-amerique/>

87 millions de personnes, abonnées à Facebook, ont vu leurs données personnelles siphonnées par une société tierce, Cambridge Analytica, dans le but d'aider Donald Trump à remporter l'élection présidentielle américaine. Très concrètement, la société londonienne Cambridge Analytica a développé un logiciel pour la campagne de Donald Trump, permettant de cibler plusieurs millions d'électeurs en fonction de leur profil.

## UN SIMPLE TEST DE PERSONNALITÉ A PERMIS DE SIPHONNER DES MILLIONS DE DONNÉES

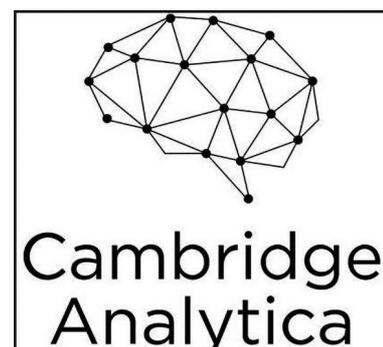
Pour récolter ces données, l'entreprise a fait appel à un professeur en psychologie de l'université de Cambridge, Aleksandr Kogan. Pour ses travaux de recherche, celui-ci a en effet développé une application : un simple test de personnalité pour lequel les personnes étaient payées un peu moins de 4 euros. Une condition était requise pour passer le test et recevoir l'argent : se connecter à l'application avec son compte Facebook.



Se faisant, la personne donnait accès (sans le savoir, à moins de regarder dans le détail les autorisations données au moment de cliquer) à un grand nombre de données personnelles, comme sa date de naissance ou l'ensemble de ses « likes » sur Facebook. En combinant les informations du test de personnalité avec les données Facebook, Kogan pouvait déterminer un « profil psychologique » de la personne, qu'il envoyait à Cambridge Analytica.

## DES DONNÉES QUI ONT SERVI À DONALD TRUMP

L'application scannait également les profils des amis de la personne visée pour établir les profils de ces derniers. Ainsi, Kogan et Cambridge Analytica ont récolté directement les informations de 270 000 personnes et 87 millions de leurs amis, majoritairement américains.



Ces données ont ensuite servi à l'élaboration de campagnes de communication ultra ciblées en fonction des profils établis. « Nous avons exploité Facebook pour récolter le profil de millions de personnes. Et nous avons construit des modèles qui utilisaient ce qu'on savait d'eux pour cibler leurs démons intimes. C'était la base autour de laquelle l'entreprise était créée », a ainsi expliqué un ancien salarié lanceur d'alerte au Guardian.

Et si Cambridge Analytica a prélevé ces données personnelles pour le compte de Donald Trump, ce n'est pas un hasard : le milliardaire Robert Mercer, propriétaire de la société londonienne, est le principal donateur individuel de la campagne du Républicain. Et l'un des fondateurs de CA est Steve Bannon, conseiller de campagne puis à la Maison Blanche de Trump.

### UNE AMENDE DE 500 000 EUROS INFLIGÉE À FACEBOOK AU ROYAUME-UNI

Les révélations des journaux anglo-saxons sur cette fuite de données ont poussé Facebook à réagir. Le réseau social autorisait en effet la collecte de données à l'époque des faits. Face au scandale Cambridge Analytica, le PDG de Facebook Mark Zuckerberg a ainsi déclaré : « Il y a une rupture de confiance entre Facebook et ceux qui partagent leurs données avec nous et qui s'attendent à ce que nous les protégeons. Nous devons réparer cela ». Il a toutefois expliqué que depuis 2014, la « faille » qui a permis d'aspirer autant de données est en partie colmatée : le réseau social avait alors pris la décision d'interdire la collecte des données des amis d'un utilisateur par une application. Sans avoir pour autant demandé des comptes aux entreprises (comme Cambridge Analytica) qui avaient pu le faire.

Pour ne pas avoir protégé les données personnelles de ses utilisateurs, Facebook a été condamné, en octobre 2018, à 500 000 euros d'amende au Royaume-Uni.

### 30 MILLIONS DE COMPTES FACEBOOK ONT ÉTÉ PIRATÉS EN SEPTEMBRE 2018

Alors que l'affaire Cambridge Analytica avait écorné l'image de Facebook, la société de Mark Zuckerberg a connu une nouvelle fuite de données en annonçant, le 28 septembre 2018, que près de 30 millions de comptes Facebook avaient été piratés par des auteurs inconnus à ce jour. Et parmi ces 30 millions de comptes, environ 3 millions concernent des Européens. À la suite de la révélation de cette faille de sécurité, l'Irlande, siège de Facebook en Europe, a ouvert une enquête pour le compte de l'UE pour savoir si le réseau social a « mis en place les mesures techniques et organisationnelles appropriées pour assurer la sécurité des données personnelles ». C'est une première en Europe, depuis le vote du règlement européen sur la protection des données (RGPD), censé mieux protéger les données personnelles.

Vincent Coquaz, journaliste formateur CLEMI

#### Suggestions pour la classe

Visionner la vidéo de Data Gueule « Big data : données, données, donnez-moi ! » sur l'exploitation des données personnelles. On y découvre que les données personnelles ont aussi été exploitées, lors de la campagne présidentielle de Barack Obama, pour récolter des fonds et adapter les arguments au profil des électeurs. Quel parallèle peut-on faire avec l'affaire Cambridge Analytica ?

Organiser pour les élèves « Connais-moi, échappe-toi », un jeu d'évasion pédagogique autour des données personnelles. L'objectif est de s'échapper d'un lieu grâce à des recherches sur l'identité et des informations personnelles notamment sur les réseaux sociaux du kidnappeur. Réalisé par le CLEMI et la DANE de Besançon, les éléments du jeu et son scénario sont librement téléchargeables.