Qu'est-ce que le Big Data?

Quel impact a-t-il sur ma vie ?

Depuis un certain temps, le terme "Big Data" est sur toutes les lèvres. Qu'est-ce qui se cache derrière ce terme ? Dans ce dossier, BEE SECURE expliquera ce que sont ces "grosses données", comment ces montagnes de données sont utilisées et quel est leur impact sur notre quotidien.

"Big Data" signifie littéralement "grosses données". Il s'agit de quantités de données qui sont trop grandes, trop complexes, trop éphémères et ne pas assez structurées pour l'exploitation à l'aide d'un traitement manuel des données (p.ex. lecture d'une valeur d'un tableau Excel).
Le "big" dans Big Data se rapporte à trois dimensions :

- Volume (volume de données) : De combien de données s'agit-il ?
- Velocity (vitesse) : Avec quelle vitesse les données sont-elles générées, déplacées et à nouveau modifiées ?
- Variety (variété) : Quelles sont les différentes sources des données ? Quels différents types de donnée existent-ils ?

Le terme francophone serait "mégadonnées", mais celui-ci n'est que très rarement utilisé dans les médias, raison pour laquelle nous utiliserons également le terme anglais Big Data.

Le terme désigne non seulement les quantités de données en soi, mais également les technologies nécessaires pour les collecter et traiter : par exemple, pour les visualiser ou les préparer d'une autre façon pour qu'elles soient compréhensibles pour les êtres humains.

L'avancée technologique a permis de stocker des quantités de données toujours plus importantes. Dans les années 90, des disquettes d'une capacité de stockage de seulement 1,4 MB et des disques durs avec quelques centaines de MB étaient la norme, tandis qu'aujourd'hui des clés USB avec plusieurs gigabits sont distribuées comme gadgets publicitaires et des disques durs avec plusieurs téraoctets sont intégrés dans des ordinateurs standards. Pour faire court : ce n'est plus un problème aujourd'hui de stocker une grande quantité de données - grâce à des "clouds storage", il ne vous faut même plus de disques durs physiques chez vous, il est possible de louer des espaces en ligne partout dans le monde. En même temps, la capacité du processeur a également augmenté, de sorte qu'il est possible de traiter des quantités de données toujours plus importantes.

D'où viennent les données du Big Data ?

De manière générale, les données considérées comme "Big Data" peuvent venir de tous les domaines de vie imaginables pour lesquels des données sont nécessaires. De manière générale, nous ne devons pas oublier en tant qu'utilisateurs que toutes les données qui peuvent être collectées seront en effet collectées. La liste suivante n'est donc en aucun cas exhaustive, mais elle offre un aperçu des sources de données du Big Data :

Visites de sites Web : Par défaut, la date, l'heure, l'adresse IP, la position approximative de la connexion Internet, le système d'exploitation utilisé (Windows, OS X, Linux, version précise incluse), le navigateur utilisé, les plugins installés, la taille de la fenêtre du navigateur, la résolution de l'écran, les sites Web visités précédemment (si la visite s'est faite en cliquant sur un lien) et les liens cliqués sont "loggés" (à savoir saisis numériquement). Pour des fins d'études de marché, quelques autres données plus précises sont loggées, par exemple, jusqu'où une page de site Web a été défilée (p.ex. pour savoir si un article a été lu dans son intégralité), où le curseur s'est déplacé, sur quoi a été cliqué, etc. Parfois, ces informations sont enregistrées dans des cookies, afin de pouvoir reconnaître les utilisateurs lors de leur prochaine visite. Le site Web clickclickclick.click permet de façon ludique de découvrir tout ce qui est suivi – cliquez sur les liens, montez le volume et laissez-vous surprendre par tout ce que le site Web sait sur vous !

Communication électronique : e-mails, SMS, WhatsApp, Messenger Facebook, etc. (sont au moins suivis : "Qui, quand et avec qui ?")

Les données des réseaux sociaux : non seulement les publications, mais également les renseignements sur ses sentiments, ses relations, les paramètres et les préférences en font partie

Paiements : l'utilisation des cartes de débit (bancomat) et de crédit, et les cartes de fidélité

Sport et activités physiques : enregistrement des données d'activité grâce au smartphone, aux bracelets de fitness ou autres "wearables"

Voyages : vols et voyages en train, par la reconnaissance de la plaque d'immatriculation il est également techniquement possible de saisir les voyages en voiture

Automobile : la "voiture connectée" engendre une grande quantité de données

Smart Home : l'"Internet of Things" ainsi que les "Smart Meters" (compteurs électriques électroniques) collectent également des données

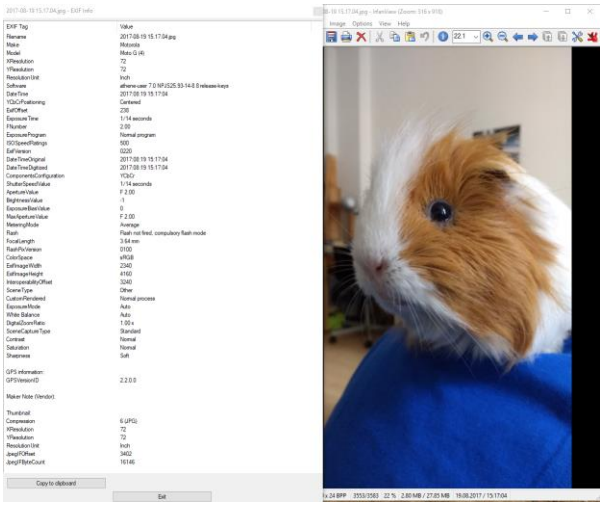
Données biométriques : empreintes digitales, Iris-Scan, reconnaissance faciale et autres techniques sont de plus en plus souvent utilisées par les smartphones, ordinateurs portables et d'autres appareils pour déverrouiller "facilement" l'appareil. Les gouvernements collectent également de telles données (p.ex. pour les passeports). Les données biométriques sont très sensibles. Contrairement aux mots de passe, elles ne peuvent pas être modifiées en cas de vol.

Surveillance : caméras (de la police ou personnes privées), reconnaissance faciale, "conservation des données" de la communication électronique

Données de contenu et métadonnées

Pour toutes ces données, on distingue généralement entre deux types de données : **données de contenu et métadonnées**. La différence peut être facilement expliquée à l'aide d'un e-mail : Les **données de contenu** sont dans ce cas le texte de l'e-mail (et les éventuelles pièces jointes et photos). Les **métadonnées** sont au moins l'expéditeur, le destinataire et l'objet. Souvent, de nombreuses autres données sont envoyées dans ce que l'on appelle le "Mail-Header" : date et heure d'envoi, programme de messagerie utilisé, langue utilisée. Souvent, chaque e-mail a son numéro d'identification unique. S'il s'agit d'une réponse à un e-mail, l'ID de celui-ci est également envoyé.

Les photos contiennent elles aussi des métadonnées, ce que la plupart des utilisateurs ignorent. Dans chaque photo JPG, les données EXIF, à savoir de nombreuses informations sur la photo et l'appareil avec lequel elle a été prise, sont cachées. Ainsi, dans une photo prise avec un smartphone, on y trouve le modèle, le fabricant et la version précise du système d'exploitation. Des informations techniques liées à la photo, telles que la date et l'heure, l'orientation de la caméra, la focale, l'exposition, l'ouverture, le flash et la valeur ISO, sont également indiquées. Une vignette (thumbnail) peut également être enregistrée - parfois, celle-ci reste conservée en cas de modification de la photo. Aussi, des coordonnées géographiques peuvent être enregistrées, ce qui permet une localisation précise. La plupart de ces informations sont également enregistrées par un appareil de photo standard. Notre photo d'exemple montre (outre un mignon cochon d'Inde) les métadonnées enregistrées à l'aide d'un smartphone standard.



Exemple d'une photo réalisée à l'aide d'un smartphone – métadonnées comprises.

Les métadonnées d'une photo peuvent être lues à l'aide d'un programme de traitement d'images (p.ex. le logiciel gratuit IrfanView) ou à l'aide du navigateur. Il est également possible de les supprimer, comme l'explique cette notice.

Dans le cas de ce que l'on appelle la *conservation des données*, souvent seules les métadonnées des communications sont enregistrées. Cela semble anodin, mais en réalité ces métadonnées permettent à elles seules de créer des profils très précis. Dans une expérience, Malte Spitz, homme politique allemand du Parti Écologiste, a attaqué en justice Deutsche Telekom afin de la forcer à remettre six mois de données sur son téléphone, qu'elle avait rendu disponibles à un journal allemand. La visualisation en résultant démontra un historique détaillé des mouvements de Spitz. En effet : lorsque des données prétendument anodines et insignifiantes sont réunies, il est possible de créer une image précise d'une personnalité. Pour chaque activité en ligne - et en raison des appareils connectés également en-dehors des navigateurs - des métadonnées sont créées qui peuvent être collectées. Et nous devons supposer que dès que des données sont engendrées, il y aura forcément quelqu'un qui les collecte, classe et analyse.

Applications et vente de données

Bien entendu, non seulement les gouvernements collectent des (méta)données, mais également les entreprises. Les géants d'Internet, tels que Amazon, Google, Facebook, etc. créent des profils pour leurs utilisateurs selon leurs propres informations pour mieux vendre leurs produits - le but est de proposer aux utilisateurs des offres adaptées. Le géant du e-commerce essaie de prévoir nos envies d'achat avant même que nous y songeons, le but du moteur de recherche et du réseau social est de nous proposer des publicités qui correspondent exactement à notre personnalité. Les données collectées permettent également parfois de décider comment nous traiter en tant que clients et si nous sommes solvables.

Outre les données que les entreprises ont collectées elles-mêmes, il est également possible d'acheter des données. Sur des plateformes, telles que big.exchange, des données d'utilisateurs sont souvent proposées à la vente - et les éditeurs de sites Web ont la possibilité de vendre les données de leurs utilisateurs. Unroll.me, un service qui permet aux utilisateurs de se désabonner d'une newsletter, s'est fait prendre en avril 2017 en train de vendre des données au prestataire de taxi "Uber" - notamment d'utilisateurs qui ne souhaitaient plus lire la newsletter de "Lyft", le concurrent de Uber. La plupart des ces deals de données se font en secret et l'utilisateur ne se rend compte de rien - même s'ils sont souvent couverts par les CGV des services.

Nous devons donc garder à l'esprit : **Si nous pouvons utiliser un service "gratuitement", nous devons tout de même payer - non pas avec de l'argent, mais avec nos données**.

Si vous souhaitez voir comment la collecte et la vente de données fonctionnent, essayez le jeu en ligne "Data Dealer". On y montre de manière ludique avec quelles méthodes les collecteurs de données arrivent à mettre la main sur nos données et qui pourrait s'intéresser à quelles données.

Le Big Data ne s'applique pas seulement à la publicité et à la vente : des secteurs comme le trafic, la médecine, l'éducation, les sciences et le journalisme peuvent profiter du traitement d'énormes montagnes de données pour rendre notre vie plus agréable. Outre les dangers pour notre sphère privée, le Big Data offre également d'énormes chances : des foyers d'infection peuvent être déterminés, des trajets optimisés, de nouvelles interrelations trouvées et des scandales dévoilés.

Comme toute technologie, le Big Data n'est ni bon, ni mauvais et il n'est pas neutre. Il s'agit d'un outil - et comme tout outil, le Big Data peut être utilisé à différentes fins. Malheureusement, pour nous, en tant qu'utilisateurs, ce n'est pas toujours clair quelles de nos données sont utilisées à quelles fins - souvent, nous ne savons même pas où nous laissons quelles données.

Garder le contrôle !

Outre la vente des données mentionnée plus haut, il existe également le risque du vol des données : les hacks, les leaks et les failles de sécurité nous concernent tous. Aucun service ne peut offrir une sécurité à 100 %. Ainsi, il vaut mieux ne pas laisser trop de traces sur Internet.

Comment pouvons-nous en tant qu'utilisateurs garder le contrôle ? BEE SECURE conseille :

Prendre les bonnes décisions

informez-vous avant de vous inscrire à une application, un réseau social, un jeu ou à un autre service. Des pages, telles que tldrlegal.com, peuvent vous être utiles.

L'art d'en dire moins

il est aujourd'hui quasiment impossible de ne rien dévoiler de soi. Cela ne veut pas dire que chaque service doit tout savoir sur vous ! Réfléchissez bien quelles informations vous saisissez. Il est souvent préférable de laisser des blancs. Avant de publier des photos, il est possible de supprimer les métadonnées.

Garder un œil sur les paramètres

gardez un œil sur les paramètres de votre navigateur et de vos réseaux sociaux afin de dévoiler le moins de données possible.

- Pour vérifier de manière générale les paramètres de votre navigateur : Is your Browser safe from tracking?
- Mozilla Firefox : Paramètres pour la vie privée, l'historique et "ne pas me pister"
- Google Chrome : Livret blanc sur la confidentialité
- Apple Safari: Paramètres de confidentialité
- Microsoft Edge : Confidentialité pour Edge
- Facebook : Privacy Settings & Tools
- Twitter : Conseils sur la confidentialité
- Instagram : Quels paramètres à prendre en compte ?
- Snapchat : Les risques de Snapchat

Auto-défense numérique

Outre les configurations, il est également possible d'adopter une démarche proactive et de pratiquer ce que l'on appelle "l'auto-défense numérique". Vous trouverez tous les conseils sur digital-selfdefense.com ou auprès de la Electronic Frontier Foundation. Il existe également quelques plugins qui aident à ne pas trop dévoiler de choses sur soi :

- Privacy Badger
- Ghostery
- disconnect.me

Le navigateur TOR permet de surfer anonymement via le réseau TOR - le navigateur est configuré de sorte qu'il laisse le moins de traces possible.

Pour encore plus de sécurité, il est possible d'utiliser la distribution Linux "Tails". Le système peut être booté à l'aide d'une clé USB et promet de ne pas laisser de traces.

Plus d'informations

BEE SECURE publie régulièrement d'autres informations sur le Big Data. Vous pouvez nous suivre sur Facebook, Twitter et Instagram ainsi que les hashtags #BIGDATA #fettdonneeen #BEESECURE.

En cas de questions sur le BIG DATA ou de manière générale sur l'utilisation d'Internet, vous pouvez à tout moment contacter la BEE SECURE HELPLINE au 8002-1234.

Si vous avez besoin de conseils juridiques ou si vous souhaitez faire une réclamation concernant un usage abusif de données, n'hésitez pas à contacter la CNPD-Commission nationale pour la protection des données.

Sources :

- Gabler Wirtschaftslexikon
- Wikipedia
- Timeline der Entwicklung von Speichermedien
- Suivis des sites Web :
- Guardian: Tracking the Trackers
- clickclickclick.click
- Vente de données :
  big.exchange
- New York Times: Unroll.me Service Faces Backlash Over a Widespread Practice: Selling User Data
- Data Dealer

Pour toute question au sujet de l'arnaque en ligne ou sur l'utilisation d'Internet en général, contactez la BEE SECURE Helpline:

Herausgeber: BEE SECURE · B.P. 707 · L-2017 Luxemburg
Tel.: (+(+352) 247-86427 · Fax.: Fax: (+ 352) 46 41 86
bee-secure@snj.luwww.bee-secure.lu

THE GOVERNMENT OF THE GRAND DUCHY OF LUXEMBOURG
Co-financed by the European Union
Connecting Europe Facility