

Je ne suis pas une somme de données personnelles, je suis un homme libre.

[Hubert Stefani](#) est Chief Innovation Officer – Associé chez [NovagenConseil](#) et [SWATICS](#)

1/ L'aspiration à la souveraineté numérique n'est pas revendiquée aujourd'hui que par des nations. Elle l'est aussi par des régions, des municipalités, des entreprises, et même des personnes ! Vous paraît-elle de nature à favoriser un morcèlement de l'autorité de l'Etat ?

Avant que nous ne transposions au numérique notre souhait d'agir au cœur de territoires sécurisés et sous contrôle, nous avons construit notre exposition par les usages. Nous avons créé (que l'on soit individu ou entreprise) des identités qui nous ont permis de communiquer, agir, consommer sans nous préoccuper véritablement du cadre légal que cela impliquait, ni de l'utilisation des données que nous concédions. Désormais, pour les différents maillages évoqués, nous mesurons à quel point cette première phase d'adoption a pu constituer au fil du temps une vulnérabilité que chacun voudrait voire atténuée. Si ce n'est un morcellement, il y a au moins un flottement sur ce sujet. L'état prend enfin des mesures qui sont en cohérence avec la sensibilité du sujet : prenons l'exemple de l'hébergement des données de santé, qui recoupe à la fois les données des personnes, le besoin des acteurs du monde médical en information sécurisée et exploitable, et doit symboliser notre capacité à traiter en toute confiance ces données... C'est en ce moment même que se décide de reconstruire le HDS selon une approche véritablement souveraine, ce qui n'était pas le cas de la première mouture.

Voilà un exemple qui pourrait nous donner confiance dans notre capacité à avoir une approche cohérente, aussi bien dans les intentions (protection des données personnelles, indépendance des infrastructures sensibles) que dans les faits. L'Etat doit favoriser les conditions d'émergence d'acteurs de taille critique pour que les différentes aspirations à la souveraineté numérique puissent se réaliser.

2/ Les données sont des attributs modernes de la personne. Comment revenir au « sujet » (joke) et faire primer la personne sur la somme de données qui le caractérisent ? Corollaire de la question : croyez-vous dans l'incarnation et / ou dans le calcul ?

Nos utilisations du numérique (je réalise des commandes, j'ai un CV sur une plateforme avec mon parcours, j'ai des activités sportives mesurées...) définissent notre portrait 'robot'. Les outils qui les exploitent s'appuient sur cette vision pour nous proposer des services (souvent fort appréciables). Or ces services nous incarnent dans des groupes de similarité (des segments ou clusters) et ce de manière de plus en plus subtile.

J'ajouterais que nous laissons bien plus d'informations que nous ne l'imaginons, et que les acteurs (GAFAM notamment) qui multiplient ces points de contacts ont une connaissance vertigineuse de nos vies : un rendez-vous chez un médecin spécialisé, des itinéraires utilisés régulièrement, des 'Like' sur des vidéos. La combinaison donne un pouvoir descriptif et prescriptif (tu devrais ..., nous te suggérons...) vertigineux et ceci d'autant plus que ce ne sont pas des règles déterministes qui sont en vigueur (si... alors...), mais des approches par 'apprentissage machine' dont on ne sait pas toujours quelles logiques algorithmiques ont été utilisées pour arriver à leurs conclusions.

Si l'on a en France des règles qui empêchent les compagnies d'assurance de s'appuyer sur des données médicales pour

établir certains contrats, on est en droit de douter de ce que certains GAFAM mettront en place alors qu'ils développent ces nouveaux marchés (Cf Amazon).

Pour faire primer la personne sur les données, le RGPD est un cadre intéressant, qui monte en puissance. Des approches complémentaires sont évoquées, qu'il faudra encore mûrir pour assainir l'utilisation des données telle qu'elle est faite actuellement : Explicabilité et Audit des algorithmes, 'Nutri-score' des données utilisées.

3/ Que vous inspire la récente déclaration du ministre finlandais des transports et communications, Timo Harraka, selon qui « « la confidentialité est un nom de code pour la méfiance. Pour profiter des avantages de la libre circulation des données, il faut plus de confiance et moins d'appréhension. »

Est-ce de l'inconscience ? du cynisme ? Pour les raisons exposées ci-dessus, on se doit d'avoir une approche bien plus défensive de l'usage des données. D'autant que lors de changements réglementaires, on pourrait imaginer des assouplissements qui libèrent l'usage de données confidentielles, au détriment de la vie privée. Et que dire de ce qu'implique un piratage de données (dont on ne peut pas exclure la probabilité de survenance – le présent est suffisamment riche de ce genre d'événements).

... Mais tentons de lui donner le bénéfice du doute et interprétons cette phrase positivement : Il y a des opportunités à faire circuler et converger des données : nous l'avons vu lors du covid, il y a eu des initiatives privées pour dynamiser les services d'informations, de prise de rendez-vous. Des services à haute valeur ajoutée se font jour quand on libère les données.

Mais on ne peut en aucun cas décréter la confiance si l'on ne prend pas des mesures de protection. Pseudonymisation, Anonymisation, Echantillonnage, Chiffrement seront nécessaires pour allier confiance et innovation.

4/ L'Ecole des Mines dont vous êtes issu a été fondée le 19 mars 1783 sur ordonnance du roi Louis XVI, dans le but de former des « directeurs intelligents » pour les mines du royaume de France. Vous voyez-vous comme un 'directeur intelligent pour les mines de données de la République » ?

Dans mon parcours il ne m'est pas vraiment arrivé d'œuvrer concrètement dans des mines. Néanmoins, dans nos projets de plateforme Data, nous employons volontiers le vocabulaire de l'industrie de l'énergie et des extractions : extraction, Raffinage, Bronze, Gold, Pipeline De Données, Gisements. Cela montre bien que nous sommes en présence d'une matière qui recèle des éléments – des gisements d'informations – qu'il s'agira de valoriser par de multiples processus techniques complexes.

Pour ce qui est de « l'intelligence » que l'on peut insuffler dans cette extraction, elle se décline selon plusieurs acceptions, parmi lesquelles: évaluation en amont des opportunités (orienter ses efforts sur des projets qui ont une valeur avérée), frugalité (ajuster des moyens adaptés à sa mission), créativité (il en faut pour assembler les activités et algorithmes selon une approche originale, différenciante) mais aussi éthique (anticiper les manipulations de données sensibles qui nécessitent des traitements spécifiques ou même qui exigent d'écarter certaines approches).

On peut pousser un peu plus loin le jeu des différences entre minerai et données : les cycles projets ne s'expriment pas en années mais en quelques semaines seulement, pour disposer d'un premier cycle complet de valorisation des données. Nous construisons et itérons sur des architectures logicielles complexes avec de faibles investissements, ce qui porte le rythme d'innovation à un niveau inégalé. Ou encore : la valeur des données s'accroît d'autant plus qu'on s'en sert.

5/ Il nous est demandé de croire que la data va

« optimiser » le monde (le grand mot) et libérer l'être humain. On a surtout l'impression que ça n'est là que le prolongement du vaste mouvement de marchandisation universelle dans l'univers immatériel du numérique. Qu'en pensez-vous ?

On pourrait invoquer ici 'Les Temps Modernes' de Chaplin : est-ce que l'on améliore les rouages de l'économie dans le but de se libérer et accéder à un 'progrès' ou n'est-on pas finalement un de ces rouages dont la machine a besoin pour satisfaire son existence propre. Quand on laisse des données de géolocalisation, ou que l'on est appelé à laisser son avis(c'est à dire de quantifier une émotion ou une réflexion, ou encore l'appréciation d'un repas), ne joue-t-on pas le jeu des plateformes qui se nourrissent de ces signaux pour s'augmenter perpétuellement.

J'aime personnellement distinguer ce qui est de l'ordre de l'entreprise et de la sphère privée. En entreprise, quand on fait parler les données intelligemment pour optimiser ses processus, on parle de Business Intelligence, il s'agit là du coeur de notre activité chez Novagen Conseil. On éclaire les décisions de nos clients sur la base d'indicateurs qui peuvent mobiliser des données multiples, difficiles à traiter, et auxquelles on offre une vitrine qui leur donne matière à décider et orienter leurs actions. C'est très efficace, le terrain de jeu est passionnant, mais de là à le décliner à la sphère privée... Je n'ai pas envie d'être le Chaplin qui tourne encore des boulons en rentrant chez moi, mais j'ai bien conscience que j'ai du mal à marquer véritablement cette frontière... Et quel impact, demain, aura la montée en puissance du metavers dans cette séparation des sphères privées et publiques ?

6/ L'intelligence artificielle qui débarque en Europe a-t-elle bien été entraînée à partir des tombereaux de données des Européens braconnés par les GAFAM ?

Nous avons été des contributeurs actifs et très innocents de ces intelligences artificielles, en acceptant sans discernement des usages qui au final ne nous récompensent pas à hauteur de la valeur que nous avons apportée. Nous pouvons nous en émouvoir, mais commençons par ne pas adopter sans discernement des solutions de captations des données.

Pour expliquer cette situation, on doit dire qu'il y a des Intelligences Artificielles très performantes qui éclosent hors des GAFAM, mais il est vrai que ceux-ci ont une avance conséquente dans les résultats qu'ils obtiennent, dans la capacité de les distribuer mais surtout dans leur puissance de collecte des données. Ils ont construit un cercle vertueux (ou vicieux selon que l'on se place de leur point de vue ou de celui des utilisateurs) où les clients de leurs services (Déplacement, Recherche de site, Réseaux sociaux, Mails) sont des contributeurs actifs (rédiger des Posts, des communications, évaluer la qualité d'un contenu, s'exprimer, partager ses coordonnées GPS). Les services ainsi constitués rendent des services très appréciables. Un des problèmes majeurs est que la valeur qui en est dégagée se concentre dramatiquement entre une poignée d'acteurs qui concentrent le monopole du cycle de gestion de ces données... et donc des bénéfices privés qui en découlent.

Il n'y a que très peu d'initiatives (ou alors balbutiantes) qui rétribuent les usagers, ne serait-ce qu'un peu. On pourrait citer le navigateur Brave qui avait initié le paiement par token (crypto monnaie) à ses utilisateurs qui concèdent le partage des données ; d'autres acteurs cherchent prochainement à rétribuer les contributeurs en tirant profit d'architectures blockchain : le sujet est à suivre.

Je voudrais illustrer mes propos par un exemple. Microsoft a créé un assistant de développement de code informatique appelé Copilot, dont on s'est rendu compte qu'il avait avalé tout le contenu de Github (plateforme largement 'gratuite' appartenant à Microsoft). Le problème est qu'il proposait des morceaux de

code qui étaient protégés par des licences open-sources qui en limitaient l'utilisation. Nous ne manquons pas d'exemples où un cas similaire se décline sur de nombreuses sources de données pour lesquelles la 'zone grise' des usages a largement été exploitée (Images, Sons, Textes).

7/ Comment imaginez-vous une possible place de marché pour l'échange de cette « commodity » moderne qu'est la donnée ?

Des solutions existent et il y a de véritables opportunités de marchés, encore peu exploitées, qui sont confirmées en observant comment ces secteurs fonctionnent actuellement. Par exemple, dans le monde pharmaceutique avec lequel j'ai travaillé, des éditeurs se sont développés en proposant des solutions de gestion aux acteurs, consolidant ainsi une somme de données qu'ils savent très bien valoriser (flux de distribution, prescriptions, tendances...). Voici une situation qui commence à être contestée par une approche place de marché qui redonne aux pharmacies ou groupements une possibilité de valorisation plus directe de leur 'trésor numérique'. Des places de marchés commencent à émerger. On doit pouvoir créer une échelle de règles simples à complexes pour préserver la sécurité et les usages.

L'anonymisation et la pseudonymisation permettent de partager la données sans la purger de tout signal intéressant : on peut effectuer des statistiques, des analyses sans accéder aux données nominatives, « l'apprentissage fédéré » pourrait aussi rendre des services : cette technologie permet aux algorithmes d'apprendre des données sans les faire sortir de leur hébergement sécurisé (une application en est l'imagerie médicale dont on exploiterait les données de chaque hôpital sans avoir à réunir les clichés en un seul endroit).

Au-delà de la technologie, il faut encore travailler sur les usages, créer des écosystèmes de taille critique pour rentabiliser l'effort de partage. Les acteurs français du

partage des données que je connais proposent des études d'opportunités aux entreprises. Identifié depuis quelques années , ce marché du « partage de la data » commence à décoller.

8/ Vous expliquez-vous ce mariage anachronique qui réunit dans une même expression deux termes issus d'époques si lointaines l'une de l'autre : « la souveraineté des données » ?

Les données sont immatérielles, 'liquides' ; nous les copions, déplaçons ; nous les manipulons par flots d'information. Il est effectivement difficile d'imaginer des frontières posées sur une matière si malléable, ubiquitaire ; d'autant plus qu'on les stocke dans les nuages [Cloud], c'est à dire des infrastructures distantes, mutualisées, proposant des services qui accentuent continuellement l'abstraction des couches matérielles (nous parlons d'architectures sans serveurs – serverless).

Pour nous assurer que la propriété et l'utilisation de nos propres données sont bien respectées, nous devons prendre des précautions que l'on ajustera selon nos différentes activités numériques.

Cela nécessite :

- D'être très précautionneux avec ses données critiques : choisir des fournisseurs dont on ne soupçonne pas qu'il peuvent s'affranchir de la protection des données (le cloud act, par exemple),
- De se mettre en ordre de marche pour se créer son patrimoine Data (et ne pas constituer celui de ses fournisseurs). Sur le sujet du Web Marketing qui m'occupe, Google occupe une place hégémonique sur la publicité, le référencement, la recherche ET l'analyse de la performance avec Analytics, tout ceci en ne respectant pas le RGPD. Il est dans ce cas vital d'adopter une solution qui donne à l'entreprise une autonomie et une objectivité dans les indicateurs de pilotage de sa performance marketing... Et grâce à laquelle nous n'aurons pas à payer pour

extraire nos propres données.

– D'adopter des solutions très performantes et souples, et moins 'souveraines' pour nos usages périphériques.

9/ Voyez-vous quelque chose de puissant, mais surtout de « noble » dans le web marketing qu'il pourrait être utile de porter à notre connaissance ?

Partons du constat qu'il est essentiel de faire connaître son identité, sa marque. Les leviers sont aujourd'hui nombreux et diversifiés, sans oublier qu'ils évoluent à un rythme effréné. Ceci rend la tâche d'autant plus complexe, au point d'y perdre beaucoup de temps, de moyen, et d'argent. La fuite en avant vers du 'toujours plus de présence, toujours plus d'investissements' peut être fortement contreproductive. Voilà pour le constat négatif que l'on peut en faire ; auquel on peut ajouter que les solutions proposées par les plateformes elles-mêmes, conduisent à des doutes légitimes sur l'objectivité des préconisations : leur laisser le soin de suggérer ce qui est bon en investissement (mots-clés, publicités), c'est se livrer à une dépendance aveugle.

Il est donc nécessaire de reprendre le contrôle pour entretenir sa notoriété avec des investissements efficaces, éclairés par des informations objectives et complètes, mais pas nominative ; c'est précisément l'objet de notre solution SWATICS. Une analyse de son écosystème Web Marketing sans dépendance à des tiers, avec des données qui alimentent le patrimoine data de nos clients.

Je reste par ailleurs convaincu que l'on peut développer une marque sans avoir accès à un parcours nominatif ('graal' que recherchent de nombreux responsables marketing – et on me l'a demandé à de multiples reprises lors de conception de stratégies Data – pourtant pas nécessaire pour disposer de l'essentiel de la valeur ajoutée de la Data et non conforme à la réglementation Européenne.

10/ Dans la grande course à la gouvernance et au « calcul », qu'est-ce qui est selon vous de nature à distinguer foncièrement les nations européennes de leurs homologues dans le monde ?

Force est de constater que dans les dernières étapes de cette course, nous n'avons pas été les plus rapides. Nous n'avons pas fait éclore les champions numériques avec une taille critique qui tire tout l'écosystème et nous nous sommes retrouvés face à une seule alternative : les fournisseurs Américains ou les fournisseurs Chinois. Ces derniers ont grandi au cœur d'un marché immense et protégé, tandis que les Etats-Unis ont fait preuve de pragmatisme en ne posant pas d'entrave au développement de géants (absence de lois anti-trust, pourtant monnaies courantes jusqu'à l'avènement du WEB), ce qui leur a donné une présence et une avance technologique qui a conduit à leur domination du moment.

Après un attentisme et des errements quant au numérique dans son ensemble, on peut noter que les derniers DMA et DSA offrent à L'Europe des moyens d'imposer des règles qui donnent des conditions de développement de solutions européennes, en particulier pour celles qui sont habilitées à traiter des données sensibles, sans risque d'application de lois extra territoriales.

L'Europe aspire à une gouvernance de la donnée qui soit équilibrée, protectrice et qui n'étouffe pas l'Innovation. On peut en apprécier la justesse avec la nouvelle version de L'hébergement des données de santé pour lequel ne devraient postuler uniquement des acteurs européens.

Il reste à noter que pour le calcul quantique, il y a une approche volontariste appréciable : voilà une technologie qui sera, par essence, proposée par un faible nombre d'acteurs de taille critique, et avec une puissance de calcul qui va redessiner le paysage du traitement des données. Il est vital de ne pas rater cette innovation de rupture, et il semble que nous ayons pris le bon train en marche !

Question subsidiaire : à combien estimez-vous la somme de vos données personnelles ?

2€ ? 10€ ? Quelques milliers de dollars si j'en juge les prix d'acquisition de twitter ou de LinkedIn dans lesquels je publie ? Pour botter en touche, je pourrais paraphraser Patrick MCGoohan dans 'Le prisonnier' en arguant que « Je ne suis pas une somme de données personnelles, je suis un homme libre », et que j'espère exister au-delà des principes purement quantitatifs. Pour conclure, il y a un principe auquel je ne dérogerai pas en tant que citoyen et footballeur : jamais je ne porterai des protèges tibia connectés ; mon style de jeu n'est pas quantifiable, et ne le sera jamais!.