EGE Ecole de Guerre Economique

Ouverture des données publiques par les administrations

Groupe de travail MRSIC 1

● Abdelhamid Abdi ● Bruno André ● Véronique Feingold ● Beatrice Ghorra ● Thomas Janier ● ● Philippa Launay ● Thomas Reiter ● Hassan Zelmat ●

Remerciements

Le groupe de travail tient à remercier chaleureusement toutes les personnes ayant pris part à la création et l'élaboration de ce travail collectif.

Nous remercions M. Christian Harbulot, directeur de l'Ecole de Guerre Économique, pour son temps, sa générosité et le partage de sa pensée pierre fondatrice du point de vue de ce travail.

Nous remercions M. Guy-Philippe Goldstein, consultant en cybersécurité et auteur de thrillers à succès, pour sa constante inspiration, son partage et sa passion continue autour du sujet de la donnée publique, de la souveraineté numérique et de la cybersécurité. Ses ouvrages furent un angle déterminant dans la rédaction de cet ouvrage.

Nous remercions M. Charles Huot, pour sa disponibilité, le partage de son point de vue et son aide à la compréhension des enjeux autour des données publiques.

Nous remercions tous nos camarades de promotion de la MRSIC 1 ainsi que nos encadrants et intervenants pour leur soutien et encouragements tout au long de ce cursus. Notre passion collective pour les thématiques abordées ainsi que les heures passées à débattre autour des différents sujets abordés ont été essentielles à l'élaboration de cet ouvrage.

Enfin, nous remercions nos familles respectives qui nous ont soutenues tout au long de ce cursus.

Table des matières

Avant propos

Nous évoluons dans une société où la donnée est omniprésente. Qu'elle soit analogique ou numérique, nous la générons, la manipulons, la consommons et la détruisons à longueur de journée. Ainsi les écrits, les manuscrits, les tableaux et les images même les plus fades sont scannés, numérisés et archivés dans des banques de données placées aux quatre coins du globe, pixel après pixel. Il ne faut en perdre aucune miette. C'est notre histoire, celle de notre évolution au fil des conquêtes, des guerres et des reconstructions que nous archivons. Les albums souvenirs, joliment conservés sont devenus de simples adresses partagées sur des plateformes de stockage d'image, visibles par le monde entier. Cette boulimie informationnelle est alimentée par nos besoins de partage, de prouver que l'on existe par ce que nous laissons derrière nous, nos histoires, nos vécus, nos appartenances, nos travaux, nos communautés, nos convictions. Les institutions publiques n'ont pas échappé à cette transformation numérique et prônent ainsi l'ouverture de la donnée qu'elles génèrent, manipulent et détiennent.

Le concept d'*open data* public a vu le jour aux États-Unis et en Angleterre. Le président américain Barack Obama développe en 2009 un portail data.gov créant ainsi le premier système de diffusion de données publiques. En 2012, la Grande-Bretagne se dote, elle, d'un *Open Data Institute* visant à soutenir le dynamisme économique. En France, le site *data.gouv.fr* explique comment notre gouvernement ouvre ses données publiques, hissant la France à la 3ème place du classement mondial "Open Data Index"[1], confirmant ainsi la tradition de transparence démocratique et de partage des informations détenues par la puissance publique.

Cette ouverture d'information englobe la mise à disposition d'un large panel de données créées et diffusées par les services de l'administration dans le cadre de leurs missions respectives. Seules les données revêtant un caractère secret ou confidentiel pouvant mettre en cause la sécurité nationale ou d'un individu sont exemptes de divulgation. Les administrations se doivent d'anonymiser ou de pseudonomiser ces informations afin de les rendre vagues et les divulguent selon un calendrier qu'elles maîtrisent. Ainsi, une bonne partie des données ne sont pas des données en temps réel. L'accès à ces données est ouvert à tout un chacun souhaitant consulter, télécharger ou traiter ces jeux de données à travers les sites Internet des administrations ou de la mission gouvernementale Etalab instaurée en 2011 pour un gouvernement ouvert.

Il existe une différence entre une donnée et une information. Une donnée est un fait brut qui n'est pas encore interprété. Elle devient une information lorsqu'on la contextualise en la

[1] L'administration change avec le numérique - http://www.modernisation.gouv.fr/en/node/86858

combinant avec une fonction de calcul, un algorithme par exemple. Ce croisement de données et leurs transformations en informations pourraient permettre l'identification d'individus, de lieux sensibles et induire des comportements qui pourraient porter atteinte à la sécurité informatique, la sécurité nationale et la sûreté des individus.

Par ailleurs, un des challenges posé par l'*open data* est de transformer les données ouvertes, disponibles en grand nombre, en information exploitable. Le fait de rendre l'information que l'on collecte pertinente pour le décideur est déjà un défi en soi, mais, avec l'*open data,* cela complexifie l'action. Les professionnels des sciences de l'information s'engagent dorénavant à apprendre à exploiter ces nouvelles informations et à prendre en compte, outre l'aspect qualitatif, l'aspect quantitatif, qui constitue l'essentiel de leur expertise.

Un autre défi lié à l'*open data* est la protection des informations stratégiques. L'*open data* c'est mettre à disposition ses données au public, or comment faire cela sans menacer un certain « secret des affaires » et ne pas dévoiler des données qui, traitées pertinemment, peuvent donner un atout stratégique à la concurrence ?

Enfin, des interrogations s'élèvent sur le terrain juridique de la protection de la vie privée et sa potentielle violation par l'ouverture des données.

Ainsi, l'*open data* crée une nouvelle source d'informations très abondante mais pose en même temps la question essentielle des risques induits par la gestion et l'exploitation de cette nouvelle mine d'or.

L'objectif de ce rapport est de mener une réflexion sur la transformation de la donnée publique en information. A travers différents scénarios que nous avons imaginés, nous allons montrer comment un individu, un groupe, ou un État pourraient utiliser ces données brutes à des fins contraires à leur mise à disposition première, pour nuire à un individu, à une collectivité ou à la sécurité nationale.
Nous aborderons ainsi quatre grands thèmes : L'éducation, pilier de notre République, l'énergie, secteur d'importance vitale, les entreprises, nerf de la guerre de notre économie et le secteur régalien de la justice.

Nous apporterons en conclusion des préconisations et des pistes de réflexion qui permettront d'enrichir le débat en vue de futurs chantiers d'amélioration.

Introduction

L'ouverture des données caractérise l'action d'une entité qui met à disposition de tiers externes certaines données qu'elle conservait auparavant en interne. Ainsi, des tiers peuvent utiliser ces données ouvertes (Open Data) librement. Ils sont en mesure de les manipuler, de les agréger avec d'autres données et peuvent aussi développer de nouvelles applications et de nouveaux usages. L'ouverture des données s'inscrit donc dans une chaîne d'au moins trois acteurs réunis au sein d'un écosystème.

1. Les émetteurs de données : les entités qui ouvrent leurs données en les rendant publiques.
2. Les « réutilisateurs » qui développent de nouveaux usages à partir de ces données ouvertes.
3. Les consommateurs qui bénéficient de ces nouveaux usages.

Les émetteurs de données publiques peuvent être de différentes natures :

- L'État avec la mission Etalab placée sous l'autorité du Premier Ministre et chargée de la création d'un portail unique des informations publiques data.gouv.fr.
- Une collectivité locale : par exemple Rennes avec son site www.data.rennesmetropole.fr.
- Une administration : l'Insee dont le portail www.insee.fr regorge de données statistiques.

Les réutilisateurs de données ouvertes sont aussi hétérogènes :

- Telle ou telle administration ou collectivité locale qui interface ses propres données avec d'autres données ouvertes à des fins de benchmarks par exemple.
- Une entreprise qui incorpore ces données ouvertes dans son système d'information ou développe une application.
- Des développeurs et fournisseurs de technologies : API, plateformes de développement…

Afin de délivrer tout leur potentiel, les données ouvertes doivent répondre à des caractéristiques précises définies depuis 2007 par un groupe de travail international indépendant : l'Open Government Data.

Ces huit principes ont été définis à l'origine pour les données ouvertes d'origine publique (État, administration, collectivités locales…) mais sont également généralement repris pour tous les types de données ouvertes quelles qu'en soient les sources, publiques ou privées.

Caractéristiques essentielles des données ouvertes selon l'Open Government Data Group.

Les données ouvertes doivent être :

1. Complètes : chaque jeu de données doit comporter toutes les données disponibles à l'exception des données sujettes à des limitations concernant la vie privée, la sécurité ou des privilèges d'accès.
2. Primaires : les données ouvertes sont des données brutes, prises directement à la source, aussi détaillées que possible et sans traitement ni modification.
3. Opportunes : les données doivent être rendues disponibles aussi vite que possible pour être le plus à jour possible.
4. Accessibles : les données doivent être disponibles pour le plus grand nombre.
5. Exploitables : c'est-à-dire prêtes à être traitées par des outils informatiques.
6. Non discriminatoires : accessibles sans inscription.
7. Non-propriétaires : disponibles dans des formats ouverts.
8. Libres de droits

Ces qualités sont en accord avec les recommandations publiées en 2009 par le W3C (World Wide Web Consortium) au sujet de l'ouverture des données : transparence, participation, collaboration, inclusion, interopérabilité, innovations, efficience, économies.

Les données ouvertes qui respectent ces caractéristiques constituent une matière brute librement accessible et prête à être traitée pour développer de nouveaux usages. Les données ouvertes peuvent être croisées avec d'autres jeux de données (ouvertes ou non) pour démultiplier le potentiel des données ouvertes. L'association de ces données produit de nouvelles informations et de nouvelles possibilités. Si l'ensemble des données combinées entre elles sont ouvertes, la combinaison peut elle-même devenir un jeu de données ouvertes et s'insérer dans une chaîne de valeur ajoutée. Ces « mash-up » démultiplient le potentiel des données ouvertes et créent eux-mêmes de nouveaux usages, tout comme Google Maps par exemple qui permet d'intégrer des données géographiques sur des cartes existantes pour les enrichir.

L'*open data* créé une nouvelle source d'informations phénoménale qui pose une question essentielle : quels sont les risques induits par la gestion et l'exploitation de cette nouvelle mine d'or ?

Un des challenges posés par l'*open data* est de transformer les données ouvertes, disponibles en grand nombre, en information exploitable. L'intelligence économique est de cette manière touchée dans son premier pilier, celui de la veille et de l'analyse. Le fait de rendre l'information que l'on collecte pertinente pour le décideur est déjà un défi de l'IE en soi, mais, avec l'*open data,* cela devient un vrai challenge. Les consultants en intelligence économique devront

apprendre à raffiner ces nouvelles informations et à prendre en compte l'aspect quantitatif en plus de l'aspect qualitatif sur lequel ils basent, pour le moment, l'essentiel de leur expertise.

L'éducation

Nos maternelles, exposées à la menace

Nous sommes dans l'hypothèse d'un individu malveillant qui cherche à attaquer une école maternelle avec une arme type kalachnikov ou un véhicule piégé dans le cas d'une attaque terroriste. Cette personne a choisi une école maternelle (cible molle, c'est-à-dire présentant plus de vulnérabilité), dans le but de renforcer l'impact de son action terroriste contre la Nation.

Dans cet objectif, cette personne malveillante va utiliser les données publiques mises à sa disposition :

Sur le site de l'éducation nationale, sont recensées : la taille des établissements, les origines des élèves (garçon ou filles) et leurs filières respectives (scientifique, littéraire…).

On trouve également les adresses de tous les établissements avec des informations sur chacun d'entre eux : les adresses e-mail, les numéros de téléphone et fax ainsi qu'un classement suivant le type d'établissement 'privé' 'public' 'école' "collège" " lycée". Un onglet 'carte' permet de géo localiser de manière précise l'école sur cette dernière.

Des outils comme Google map permettent de voir l'établissement en 3D en vue d'une attaque malveillante. En prenant l'exemple d'une école à Grenoble, nous pouvons facilement trouver le plan d'une école maternelle.
https://www.grenoble.fr/cms_viewFile.php?idtf=5062&path=Dossier-de-presse-plan-ecole.pdf Il existe également des sites d'écoles communautaires qui facilitent encore davantage le ciblage:
https://www.alloj.com/fr/ecole-juive.html

L'individu va pouvoir utiliser ces informations afin de localiser et choisir une école à cibler.
http://www.education.gouv.fr/pid24301/annuaire-de-l-education.html

Dans un deuxième temps après avoir sélectionné l'école, l'individu va chercher à savoir s'il peut aller dans cette école pour effectuer son acte, l'idéal dans sa logique étant d'y aller quand il y a le plus de monde. En retournant sur le site de l'Education Nationale, les informations sur les emplois du temps sont accessibles notamment dans le cas des écoles maternelles : En effet, un moteur de recherche donne les activités périscolaires des petits enfants et leur temps de présence dans l'école.
http://www.education.gouv.fr/pid29074/rythmes-scolaires.html.

Le détail des activités périscolaires est disponible sur les sites des mairies des villes.

L’attaquant n’a donc plus qu’à recouper les informations pour trouver les activités périscolaires, et donc choisir soit de faire son acte dans l'école, quand celle-ci est remplie, soit de cibler les enfants de l'école (par exemple lorsqu’ils sont en activité piscine ou autre).

Conclusion: en quelques clics, une personne malveillante a récolté de manière tout à fait anonyme des informations qui vont lui permettre d’effectuer un acte de type tuerie de masse sur des petits enfants, c’est-à-dire un acte avec un impact national, voire international.

> Piste de réflexion : la traçabilité des recherches relatives aux données des établissements scolaires consultables en accès libre pourrait être renforcée et faire l’objet d’un suivi automatisé qui permettra par exemple de mettre en relief par les services qualifiés, qu’un individu présentant un profil risqué (fiché S par exemple) a un comportement suspect, ou tout simplement de récolter des informations sur une personne opérant ce type de recherche.

Les parents sur écoute

Nous sommes dans l’hypothèse d’un individu ou une organisation malveillants qui souhaite agir contre les intérêts de l’établissement et sa communauté d’élève, et par extension contre les services de l’Etat.

On peut réutiliser les données publiques trouvées sur le site de l’éducation nationale qui comprennent notamment la liste des e-mails de chaque établissement ce qui permettrait à l’individu d’envoyer un lien par ce biais pour propager un virus informatique ou employer une attaque de type Phishing. En cliquant sur le lien, la personne croyant insérer ses identifiants d'une façon légitime se trouve piégée et donnerait ainsi accès à ces informations à l'attaquant. Ces accès permettraient à l’individu de naviguer dans les différentes applications, de bloquer le système ou d'amplifier l'attaque et ses conséquences.

Plusieurs vulnérabilités peuvent être exploitées dans le milieu scolaire, dont la très répandue application Pronote qui est un logiciel de gestion de notes, absences, retard, punitions, sanctions, dossiers scolaires, cahier de texte, menus de la cantine, vacances, agenda de l’établissement, etc… Aujourd’hui plus de 6 700 établissements utilisent cette plateforme pour communiquer et des millions d’utilisateurs consultent leur espace tous les jours.
(Source : index-education.com).

Une autre source d’inquiétude légitime est l’application Apple « eParents », application officielle choisie par le ministère de l’Education Nationale dont le but est d'accompagner les enfants tout au long de leur vie scolaire avec l’accès à plusieurs services du portail « Scolarités Services ». Comme l’indique Apple et son développeur, le ministère de l’Education Nationale, l’application « eParents » ne collecte aucune donnée personnelle relative à l’individu ou à ses enfants puisque les informations sont stockées localement sur le terminal de l’individu et ne sont pas

transmises au ministère ou à un tiers. Néanmoins, ces données stockées sur les téléphones Apple sont susceptibles d'être enregistrées par Apple dans le cadre des sauvegardes récurrentes sur ces équipements par exemple. La marque prépare depuis quelques années un usage transparent de ses services sur tous les équipements au sein d'une famille via son service iCloud. Ainsi pour faciliter la portabilité des données depuis un ordinateur de bureau de la marque ou un iPad, l'utilisateur est obligé d'activer la sauvegarde des données sur le iCloud, cloud privé d'Apple. La consultation des dossiers ou de fichiers pdf par exemple peut être démarrée à partir d'un iPad pour être terminée sur un ordinateur portable pour être imprimée.

On peut retrouver l'application pour les parents d'enfants scolarisés à l'adresse : https://itunes.apple.com/fr/app/eparents-mon-enfant-Ã%C2%A0-lÃ©cole-et-au-collÃ¨ge/id1142988556?mt=8

> Piste de réflexion : Aujourd'hui, les établissements scolaires sous tutelle des ministères de l'Education Nationale, de l'Agriculture et de l'Europe et des Affaires Étrangères ne sont pas inscrits parmi les sites qualifiés Point d'importance vitale (PIV) car aucun opérateur dit d'importance vitale (OIV) ne les a qualifiés de cette importance.
> Néanmoins, le bon fonctionnement des établissements scolaires sans être d'importance vitale, présente un intérêt à être qualifié Point de service essentiel (PSE) par les Opérateurs de Service Essentiel (OSE) et s'inscrire dans une logique vertueuse et nécessaire de mise aux normes de leur politique de sécurité des systèmes d'information.

L'exode

Nous sommes dans l'hypothèse où une organisation ou un État utilisent les données publiques dans l'objectif de détourner des potentielles et/ou des futures élites soit pour leur propre profit soit pour en faire des acteurs qui agiront contre les intérêts de la France.

Il est ainsi commun de nos jours de voir étalées sur les réseaux sociaux des informations, des images ou des vidéos mettant en scène des entreprises et leurs personnels heureux d'appartenir à leur "boîte", des industries de défense et leurs technologies voire les différents corps d'armées en entraînement ou sur les théâtres d'opérations tout sourire sur les clichés. C'est toute une opération de séduction mise en place dont le but est de mettre en avant le savoir-faire de chaque parti.

L'usage fait par les armées des outils du civil permettrait ainsi de rapprocher les audiences, de renforcer le lien Armée/Nation et créer un lien plus solide avec le grand public. Ainsi, en partageant le quotidien, les défis et la cohésion du groupe en opération, l'armée n'est plus un univers fermé et lointain que l'on voit uniquement le jour du 14 Juillet à la télévision. L'armée est désormais accessible à tous, tant pour recenser les patriotes que les non-patriotes. Aussi faut-il susciter des vocations pour renouveler ses rangs surtout après les attentats qu'a connus la

France récemment. Cet usage permet d'identifier à date des soldats en opération avec leur localisation approximative. Il suffit d'être attentif aux conditions de prise de vue des photos pour en extrapoler la localisation par exemple. Le partage des informations sportives via l'application Strava avait quant à elle révéler la localisation de bases secrètes sur certains théâtres sensibles.

Les réseaux sociaux sont une mine d'or informationnelle, les individus y sont bien bavards. Les réseaux sociaux dits généralistes comme Twitter et Facebook ou Instagram sont des livres ouverts sur les idées, les pensées, les affinités et les cercles de connaissance d'un ou de plusieurs individus. Mais le vrai ciblage s'opère au niveau des forums de discussions à thème, plus refermés et traitant de sujet de niche. Bien que ces données n'émanent pas de rapports ministériels, elles n'en restent pas moins publiques et exploitables à l'échelle.

Le jeu consiste alors de cibler des talents pour influer sur leurs comportements dans la direction souhaitée. L'illustration de ce cas serait d'imaginer une entreprise aéronautique étrangère ciblant des talents d'écoles d'ingénieurs ou de futurs cadets de programmes de pilote de ligne. Ces individus seraient suivi par leur pseudonyme sur Internet à travers toute la toile créant ainsi un profil complet de leurs humeurs, états d'âmes, forces et faiblesses et envies. L'industrie en chasse pourrait ainsi proposer des concours sur mesure permettant de mettre en valeur les talents des individus ciblés et de les fidéliser à travers des jeux ou des places VIP à des salons ou des expositions sélectifs et prestigieux voire à travers de rencontres avec des stars du milieux. Le but étant de fidéliser cette audience et de la rendre accro à la "marque" ou à l'entreprise. Il serait aussi utile de se rappeler de la chasse aux talents Français opérée par Google via leur programme Google Science Fair où un talent Français remporta le prix "Incubateur" à juste 13 ans pour son invention. C'est en créant se liant et en promettant un avenir assez scénarisé que les entreprises et industries attireront leurs futurs talents.

Nous trouvons déjà actuellement en accès publique des indicateurs qui permettent un ciblage des talents en France :

- Classement des lycées et des collèges

https://www.lexpress.fr/education/palmares-des-lycees-notre-methodologie_1100956.html
https://www.lexpress.fr/palmares/lycees/

- Lauréat des bourses

http://www.lfbogota.com/nos-eleves-recompenses-par-la-bourse-dexcellence-major/

- Lauréats de prix entrepreneuriat étudiant

https://data.enseignementsup-recherche.gouv.fr/explore/dataset/fr-esr-laureates-et-laureats-prix-pepite-tremplin-pour-lentrepreneuriat-etudiant/liste/?sort=-millesime_prix&location=2,18.44935,8.24636&basemap=jawg.streets

La guerre économique fait rage aujourd'hui et sa prochaine bataille, à l'aune de la révolution numérique, se tiendra dans le champ de l'information et plus particulièrement, probablement

dans la détermination, le développement et l'influence des talents des Nations. Les outils existent déjà, à l'image de l'application mondialement réputée *Linkedin,* qui permet à la firme américaine de faire du ciblage professionnel à l'échelle internationale, une vraie mine d'or pour les services de renseignements américains qui y découvrent très probablement le potentiel des individus et les réseaux qui les lient entre eux. Un autre exemple d'outil existe encore, après les technologies de reconnaissance faciale qui permettent d'identifier une personne : les technologies qui analysent les émotions pour savoir ce qu'une personne ressent. L'amas même anonymisé de ces informations renforcera la connaissance des groupes d'individus, puis une fois croisée avec d'autres données complètera la connaissance ultime des individus qui isolés seront rendus plus vulnérables à l'influence d'une entité qui les dépassera en nombre et pouvoir, c'est-à-dire en influence. C'est l'enjeu de l'atomisation de la Nation qui d'un côté répond à la nécessité de respect du droit des libertés individuelles mais qui par ailleurs, doit composer avec l'intérêt supérieur de la Nation. Dans le cas contraire,c'est la puissance de la Nation qui s'affaiblit.

Enfin, on relèvera un dernier exemple d'outil existant avec Facebook qui récolte des données émotionnelles sur ses utilisateurs en analysant leurs réactions à un article, une vidéo, une photo. Ayant étoffé le nombre de boutons leur permettant d'exprimer leurs sentiments (au-delà du «Like» original), les «J'aime», «Haha», «Wow», «Triste» et «En Colère» sont autant d'informations que le réseau social peut exploiter pour servir un contenu encore plus ciblé, et tout simplement développer une meilleure connaissance des individus pour mieux les influencer. L'affaire « Cambridge analytica a parfaitement illustré ce risque qui pèse sur la souveraineté nationale.

La simple combinaison des applications Linkedin et Facebook, permet aujourd'hui d'analyser tant la sphère professionnelle que privée d'un individu, d'un groupe d'individu, d'une entreprise, ou encore d'un service de l'Etat. Cet outil imaginé en exemple constitue un atout majeur pour un service de renseignement qui se déciderait à l'employer, hypothèse hautement probable à ce jour. De surcroît, il présente l'intérêt d'être économiquement très rentable pour la Nation américaine.

Pistes de réflexion

Face au risque de fuite des cerveaux et des influenceurs pour une cause contraire à nos intérêts, nous préconisons en prévention une meilleure traçabilité des personnes qui recherchent ce type d'informations et une protection accrue de nos futures élites avec un meilleur suivi, c'est-à-dire en veillant à faire perdurer l'appartenance à la culture française, à la Nation, à la communauté des grandes écoles par exemple.

Dans le même temps, afin de renforcer le soft power de notre Nation, il s'agira d'appliquer de manière offensive ces modes opératoires défensifs encore imparfaitement pratiqués, pour influencer les cerveaux étrangers, avec la volonté de cibler les futurs décideurs ou les futurs entrepreneurs et ce qui fait la richesse du pays

au niveau humain. En effet, demeurer sur une posture défensive c'est certes résister mais c'est aussi se priver d'une possibilité d'agir dont ne s'embarrassent pas d'autre nations. Pour cela, il appartiendra par exemple de déterminer des indicateurs de la performance des cibles, de leur talent latent ou avéré et ainsi aboutir à un *profiling* automatisé.

L’énergie

Vers la paralysie du pays demain ?

« L’énergie est un des secteurs les plus ciblés par les cyber attaques aujourd’hui » Frédéric Julhes, Directeur de la cybersécurité, Airbus

La loi Lemaire *oblige* les administrations à publier toutes les données produites par l'Etat. C’est le principe de l'ouverture des données publiques par défaut.

Elle va même un cran plus loin en favorisant l'ouverture des données d'acteurs semi-publics ou privés exerçant une mission d'intérêt général.

Objectif : libérer du potentiel économique, encourager l'innovation.

Puis, dans la lignée de la loi Lemaire, la loi portant sur la Nouvelle Organisation Territoriale de la République (NOTRe) introduit l’open data des collectivités territoriales.

A ce jour il existe 145 jeux de données publiés sur l’énergie, dont les plus populaires sont la production et la consommation d’énergie en France.

« (...) les partenariats public-privé ont du bon - le système GPS est un bon exemple : dans les années 1980, le gouvernement américain l'a ouvert aux civils, créant des opportunités extraordinaires. »
Noam Bardin, PDG de Waze, filiale de Google (24/10/16, Les Echos)

Or, dans l’énergie, l’identification des points générateurs et consommateurs, leur type de production, leurs habitudes, à un fin degré, si elle facilite une gestion du parc énergétique en interne et favorise l’émergence de nouveaux acteurs économiques, représente un avantage concurrentiel clé, ainsi qu’une menace militaire potentielle face à des pays qui n’ouvrent pas leurs données. C’est pourquoi nous parlerons ici d’”exposition” plutôt que de publication de données.

Dans l’énergie, il n’est pas nécessaire pour l’ennemi de prendre le contrôle ou d’actionner à distance l’outil industriel ou civil. Il suffit de paralyser ou simplement d’introduire un caillou dans le système pour générer des effets aux conséquences désastreuses, dont les dégâts peuvent coûter des sommes considérables et prendre des années à réparer. En effet ceux-ci peuvent être d’ordre de dérèglements économiques et financiers, mais également occasionner des pertes humaines et des dommages environnementaux massifs.

En 2016, 80% des entreprises du secteur disent avoir vu un nombre croissant d'attaques informatiques "réussies" sur leurs infrastructures[2].

Ainsi nous verrons dans cette partie quelques exemples de données librement "exposées", afin de développer une approche scénaristique des risques et menaces qui pèsent sur le pays en matière d'énergie. Nous terminerons cette partie par l'émission de quelques recommandations de prévention.

Etat des lieux de la donnée publique dans le secteur de l'énergie: ce qui est exposé

Un réglage fin de la publication des données: la loi NOTRe rend obligatoire de communiquer les données des collectivités de plus de 3 500 habitants, incluant les données de consommations énergétiques.

Nombre de communes ont été volontaristes en mettant en ligne leurs données.

Voici quelques exemples liés au secteur de l'énergie :

- Consommations énergétiques du patrimoine bâti, des ouvrages d'eau et des ouvrages d'assainissement, et synthèse des consommations énergétiques
- Consommation de chauffage des collèges
- Production photovoltaïque sur les bâtiments communaux et certificats d'économies d'énergie
- Chaufferies bois et réseaux de chaleur, installations géothermiques et unités de méthanisation subventionnés par l'ADEME
- Bâtiments de la communauté d'agglomération du Grand Poitiers, du département de l'Oise, de la métropole Montpellier Méditerranée et de la ville de Montpellier
- Bâtiments publics de la ville de Digne-les-Bains, de la ville de Marseille et de la communauté d'agglomération Arles Crau Camargue Montagnette

Des dizaines de jeux de données sont créés et libérés chaque jour dans une frénésie d'open data en vogue, pas toujours très contrôlée, et surtout qui, par le truchement des bases de données et leur manipulation, permettent de localiser dans l'espace et dans le temps des groupements toujours plus fins.

[2] *World Energy Perspectives, The Road To Resilience - World Energy Council*

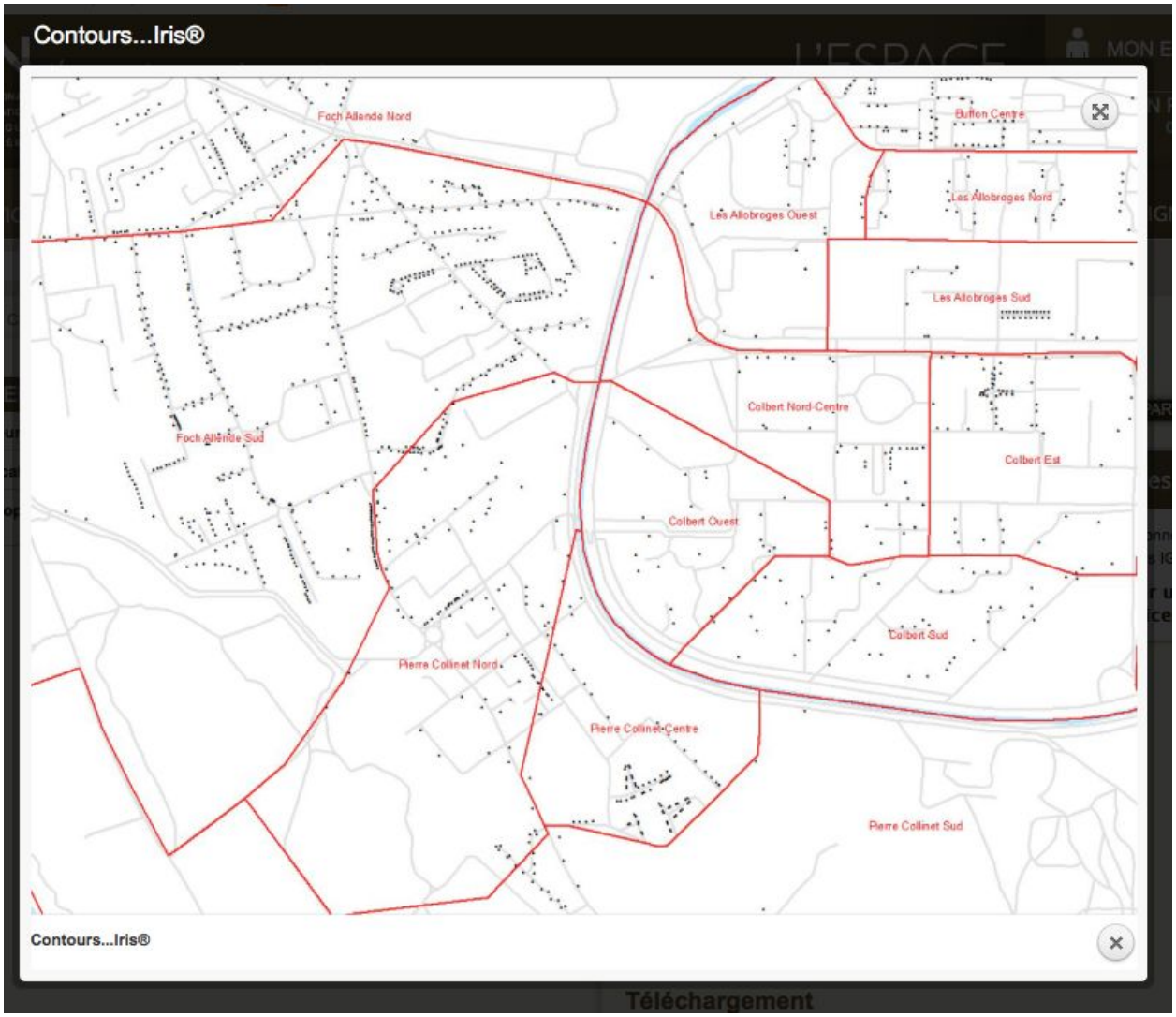
Nous prenons ici deux exemples de bases de données de croisement.

A. Croiser les données, toucher les populations: Ilôts Regroupés pour l'Information Statistique (IRIS)

L'IRIS est à ce jour un des niveaux les plus précis d'identification, qui permet de lier une base de consommation ou production avec un groupe d'individus donné.

Ce fin niveau d'identification permet une approche malveillante de ciblage "top down": on prend le jeu de données et on cible une personne / une entreprise (dirigeants ou anonymes ciblés aléatoirement ou en fonction de critères discriminants).

Qu'est-ce qu'IRIS?
Il s'agit de la maille la plus fine utilisées par l'INSEE pour diffuser des données issues du recensement de la population.



Origine

Afin de préparer la diffusion du recensement de la population de 1999, l'INSEE avait développé un découpage du territoire en mailles de taille homogène appelées IRIS 2000. Le sigle faisait référence à la taille visée de 2 000 habitants par maille élémentaire.

Depuis, l'IRIS constitue la brique de base en matière de diffusion de données infra-communales. Il doit respecter des critères géographiques et démographiques et avoir des contours identifiables sans ambigüité et stables dans le temps.

Les communes d'au moins 10 000 habitants et une forte proportion des communes de 5 000 à 10 000 habitants sont découpées en IRIS. Ce découpage constitue une partition de leur territoire. La France compte environ 16 100 IRIS dont 650 dans les DOM.

Par extension, afin de couvrir l'ensemble du territoire, on assimile à un IRIS chacune des communes non découpées en IRIS.

On distingue trois types d'IRIS :

- les IRIS d'habitat (code H) : leur population se situe en général entre 1 800 et 5 000 habitants. Ils sont homogènes quant au type d'habitat et leurs limites s'appuient sur les grandes coupures du tissu urbain (voies principales, voies ferrées, cours d'eau, ...) ;
- les IRIS d'activité (code A) : ils regroupent environ 1 000 salariés et comptent au moins deux fois plus d'emplois salariés que de population résidente ;
- les IRIS divers (code D) : il s'agit de grandes zones spécifiques peu habitées et ayant une superficie importante (parcs de loisirs, zones portuaires, forêts, ...).

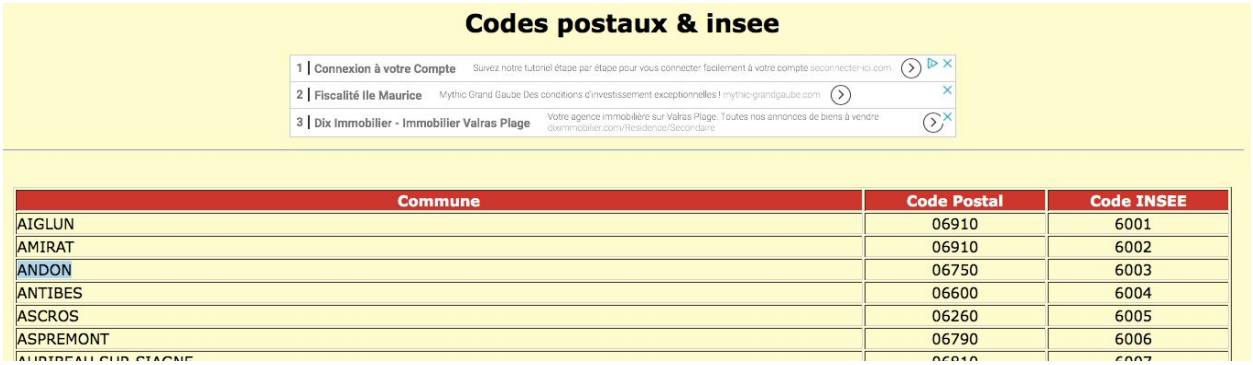
Pour les communes non découpées en IRIS, le type de l'IRIS est codé à Z.

Au 1er janvier 2008, 92 % des IRIS étaient des IRIS d'habitat et 5 % des IRIS d'activité.

Enedis est le premier distributeur d'électricité européen à se lancer dans la publication libre d'une large gamme de données publiques.

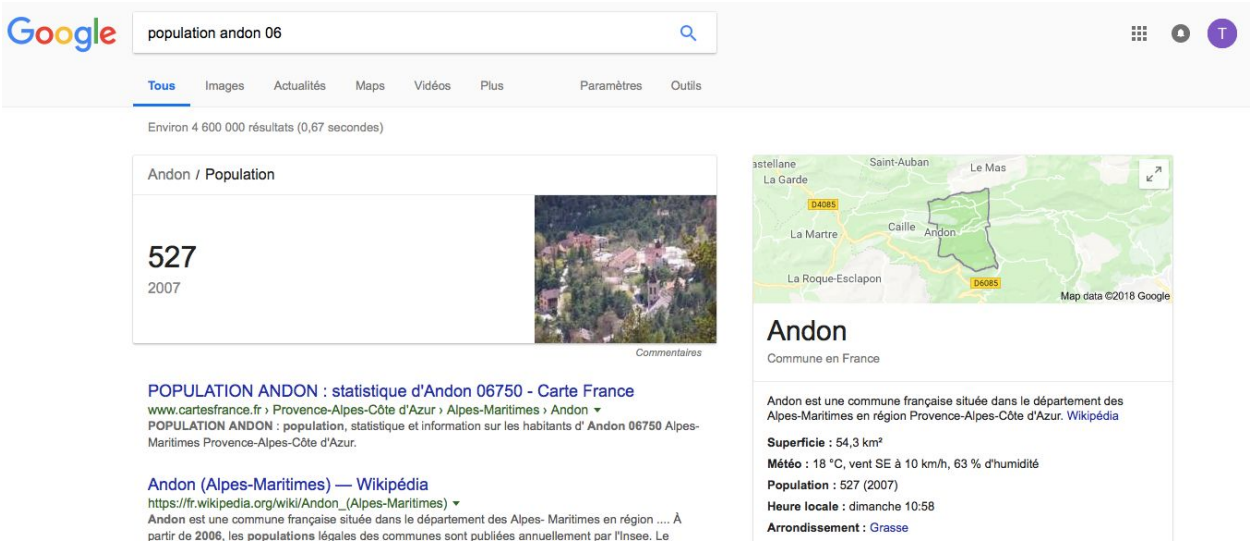
Dans le cadre de la transition énergétique, la société publie fréquemment de nouvelles données de consommation électrique annuelle jusqu'à l'échelle « IRIS ». Ces données, accessibles par secteur d'activité et sur une période de 5 ans, sont destinées aux acteurs de la transition énergétique, notamment les collectivités locales, d'évaluer plus précisément la consommation d'électricité sur leurs territoires.

Or, dans nos recherches, nous avons constaté plusieurs entités IRIS représentant un échantillon de moins de 1 000 habitants (voir l'exemple ci-dessous de Andon, Alpes-Maritimes, 527 habitants).



Codes postaux & insee

Commune	Code Postal	Code INSEE
AIGLUN	06910	6001
AMIRAT	06910	6002
ANDON	06750	6003
ANTIBES	06600	6004
ASCROS	06260	6005
ASPREMONT	06790	6006



Enedis tend à aussi devenir gestionnaire de systèmes électriques et gestionnaire de big data énergétique ; ses données de consommation électrique ont été publiées à l'échelle INSEE de l'IRIS (échelle du quartier) en open data sur 5, à la fois sur le site d'Enedis et celui d'Etalab, et il est prévu d'ouvrir aussi des données sur la production, anonymisées, mais venant aussi des compteurs communicants (qui doivent aussi permettre aux énergies renouvelables de mieux s'intégrer dans le réseau électrique).

(voir partie 2 Linky et vulnérabilités avérées du SCADA)

B. Croiser les données, toucher les populations: la Base des Adresses Nationale (BAN)

L'État, et La Poste ont lancé le 15 avril 2015 la mise à disposition d'une base de données de toutes les adresses de France – et ce en incluant la géolocalisation. Elle est réutilisable librement, à condition toutefois de participer à l'amélioration des informations.

Ces informations en libre accès à tous, peuvent être combinées à IRIS et aux données, très populaires, de production et de consommation d'électricité, par exemple et entre autres, pour permettre d'extrapoler et de cibler des adresses nominatives, pour peu que l'on se déplace pour mettre un nom sur une boîte aux lettres physique.

Une autre façon de cibler de manière précise un point de consommation énergétique par extrapolation et en open data est le cadastre, consultable en présentiel en mairie… Pour le moment.

Quelles informations délivre le cadastre ?

Le cadastre, c'est d'abord un plan cadastral qui décrit les 101 millions de parcelles recensées sur le territoire des 36 000 communes françaises, sous la forme de 600 000 feuilles (ou planches) cartographiques dont l'échelle varie en général du 1/1 000 (1 cm sur le plan correspond à 10 m de parcelle) au 1/2 000. C'est un bon document d'approche qui, d'ailleurs, sert de support à la commune pour l'élaboration de son plan local d'urbanisme (PLU).

A ce plan s'ajoute la matrice cadastrale, une base de données d'informations littérales issues de celles recueillies dans le fichier immobilier.

Le fichier immobilier, détenu par la conservation des hypothèques, constitue un registre dans lequel sont inscrits les actes notariés. Si le cadastre assure l'identification et la détermination physique des biens, le fichier immobilier, de son côté, en détermine l'état juridique.

Sur la feuille de plan figurent :

- Une représentation graphique du bâti « dur » (habitation) et « léger » (constructions ouvertes, comme un hangar ou un garage) ;
- Les limites avec les parcelles voisines et la voie publique ;
- L'orientation (indication du nord);
- Les chemins ruraux, voies communales et autres voies publiques ;
- Les cours d'eau ;
- La nature des cultures ;
- La mitoyenneté sur un mur ou une haie, des chemins matérialisés en pointillé sur le plan ;

- Quelques détails topographiques tels que rond point, bord de voirie, lieux de culte, structure sportive, puits...

Sur la matrice cadastrale figurent :

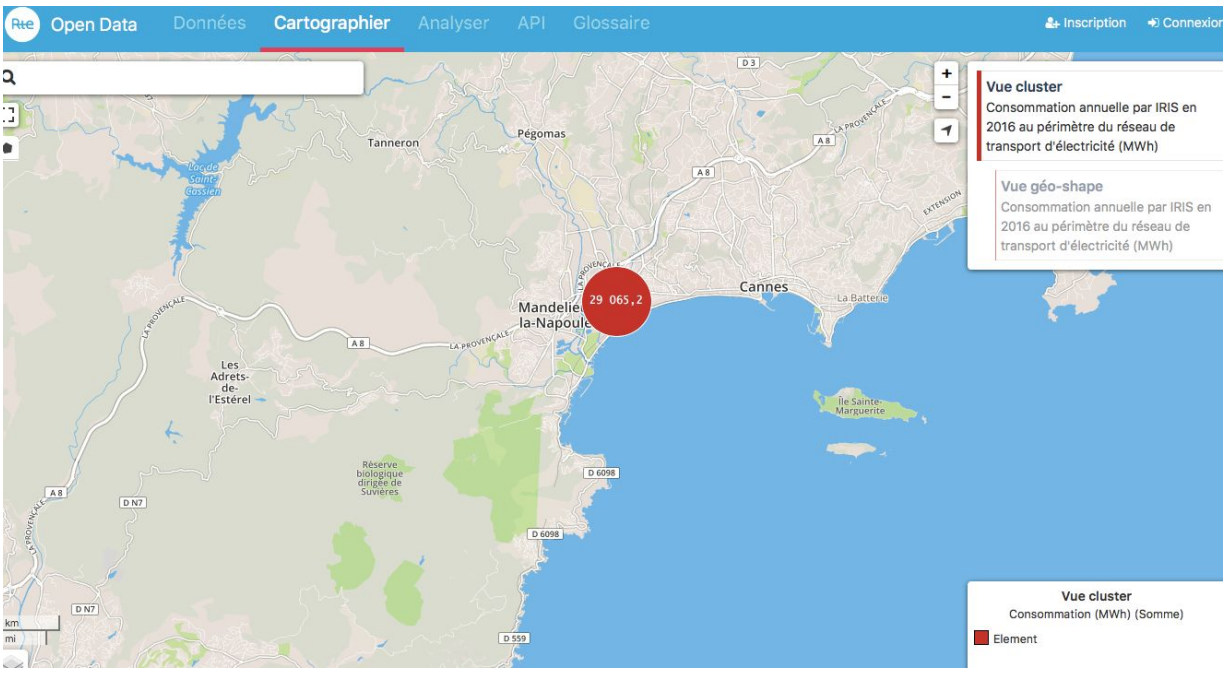
- L'état civil (nom et adresse) du propriétaire de la ou des parcelles, une même propriété pouvant être divisée en plusieurs parcelles ;
- Un éventuel régime d'usufruit ou d'indivision : dans ce cas, le nom de tous les propriétaires concernés est indiqué ;
- L'adresse de la parcelle (nécessairement un terrain d'un seul tenant) ;
- L'identifiant de la parcelle : c'est l'équivalent du numéro de Sécurité sociale pour les personnes. Cet identifiant comprend la désignation de la commune ou du lieu dit, selon que le terrain se situe en zone urbaine ou rurale, la section cadastrale (une à deux lettres) et le numéro proprement dit de la parcelle ; - la contenance cadastrale, indiquée en ares, centiares ou hectares. Attention : il ne s'agit pas d'une superficie exacte du bien. Car le cadastre est un document... fiscal.

Pour l’heure, les données libérées ont un décalage volontaire dans le temps, mais nous observons une tendance à la réduction du délai entre la publication et l’actualisation de la donnée.

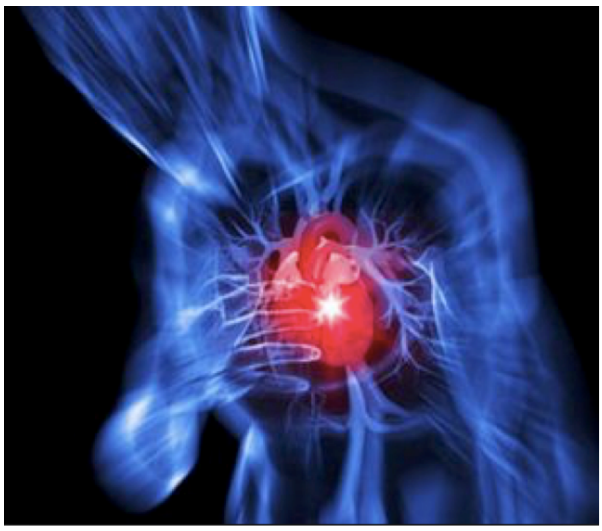
Nous mesurons dans une seconde partie les risques, menaces et conséquences potentielles liées à une telle ouverture des données de la population et de ses acteurs économiques.

Risques, menaces, conséquences

“le meilleur moyen de s’attaquer à un pays est de lui couper l’électricité. C’est l’équivalent, cyber, d’une attaque nucléaire : quand il n’y a plus d’électricité, tout s’arrête.
“



Menace (nf, -s): le fait qu'une personne ou une entité ait la possibilité (...) d'infliger des blessures, la mort ou des dommages matériels à une autre personne ou groupe de personnes. (Source: wikipedia)



A. Attaques industrielles: un rappel historique des attaques internationales récentes

15/08/12	Saudi Aramco, le groupe d'État qui gère toute la production pétrolière de l'Arabie Saoudite, a subi une attaque virale qui a endommagé environ 30 000 ordinateurs par infection de logiciels malveillants et détruit 85% du matériel sur les appareils de l'entreprise. Le virus, appelé «Shamoon», ne visait pas seulement Saudi Aramco en tant qu'entité; il a attaqué l'économie entière du pays.
23/12/15 Et Déc 2016	Des pirates informatiques ont pénétré dans les systèmes informatiques et SCADA de la société de distribution d'électricité ukrainienne Kyivoblenergo et ont débranché sept sous-stations de 110 kV et vingt-trois 35 kV, provoquant une panne de trois heures pour environ 80 000 clients. Cette attaque a été le premier événement cybernétique publiquement reconnu ayant un impact sur l'approvisionnement en électricité d'un pays.
07/2017	Les autorités américaines ont diffusé une alerte après la découverte d'attaques visant des entreprises du secteur énergétique.
08/2017	La presse irlandaise révélait que l'entreprise publique de distribution électrique avait été pénétrée par un groupe de pirates étatiques.
11/2017	Un responsable du renseignement britannique révèle que des pirates s'en seraient pris notamment au secteur de l'énergie du Royaume-uni.

Dans son rapport annuel, le Réseau de transport d'électricité (RTE) – chargé de l'acheminement du courant haute tension dans l'Hexagone – explique avoir déjoué en 2016 « 4 300 attaques » et « 200 virus » par mois.[3]

Le SCADA et Linky

Un système d'acquisition et de contrôle de données (SCADA) (anglais : Supervisory Control And Data Acquisition, sigle : SCADA) est un système de télégestion à grande échelle permettant de traiter en temps réel un grand nombre de télémesures et de contrôler à distance

[3] Source: *https://www.lemonde.fr/pixels/article/2017/12/08/le-reseau-electrique-francais-peut-il-etre-pirate_5226462_4408996.html#p9yi395qxkPpPfLG.99*

des installations techniques. C'est une technologie industrielle dans le domaine de l'instrumentation.[4]

A l'occasion de l'installation de 35 millions de compteurs connectés et collecteurs de données Linky, la société a demandé au gestionnaire du réseau de distribution, la société Enedis, de lui transmettre les données de ses clients correspondant à leur consommation journalière d'électricité ainsi que les données de consommation à la demi-heure, selon un communiqué de la Cnil, relayé par L'usine nouvelle.

Or les défaillances et vulnérabilités des systèmes Scada sont connues et bien documentées.

Les 10 principales défaillances des systèmes Scada selon l'expert en cybersécurité, Lexsi[5]:

1. Absence de développement sécurisé dû à des développements internes qui répondent aux besoins internes. Il en résulte des comptes d'accès stockés en clair ou encodés trivialement (le nom de la société comme mot de passe principal, par exemple).
2. L'absence de test de sécurité découle en toute logique de l'absence d'intégration de la sécurité informatique dans les projets.
3. Une mauvaise gestion des comptes où l'on retrouve l'usage d'identifiant par défaut (user/user…), des mots de passe trop faibles ou inexistant (vides, nom du client, mot du dictionnaire évident…) ou encore des utilisateurs disposant de privilèges administrateur sur l'OS.
4. L'interconnexion des systèmes de gestion avec les systèmes industriels pas assez sûre. Une perméabilité qui permet à un attaquant qui s'introduirait dans le système de gestion informatique de poursuivre sa route sur le réseau industriel.
5. L'absence d'antivirus sur les postes de travail et serveurs qui laisse tout loisir aux agents malveillants de se propager. Lexsi a ainsi constaté la présence du vers Conficker sur des postes de supervision industrielle dans 50% des cas.
6. Absence de veille en cybersécurité qui rend difficile la détection de signaux d'alerte et la remontée d'information. Au mieux, souligne Lexsi, les logs sont enregistrés par les firewall mais rarement analysés.
7. Des sessions Windows non verrouillées qui rendent l'accès permanent aux interfaces de contrôle (IHM) ou consoles de pilotage. Là encore, en cas d'attaque, la prise de commande distante est un jeu d'enfant pour l'assaillant.
8. Absence d'outils de surveillance des systèmes (sondes de détection/prévention d'intrusion) et pas de centralisation des journaux systèmes et de leur analyse.
9. Des protocoles courants (FTP, Telnet, VNC, SNMP…) utilisés sans chiffrement qui ouvre l'accès à la récupération de login/mot de passe, à des connexion illégitimes aux

[4] https://fr.wikipedia.org/wiki/Syst%C3%A8me_d%27acquisition_et_de_contr%C3%B4le_de_donn%C3%A9es_(SCADA)

[5] Silicon.fr

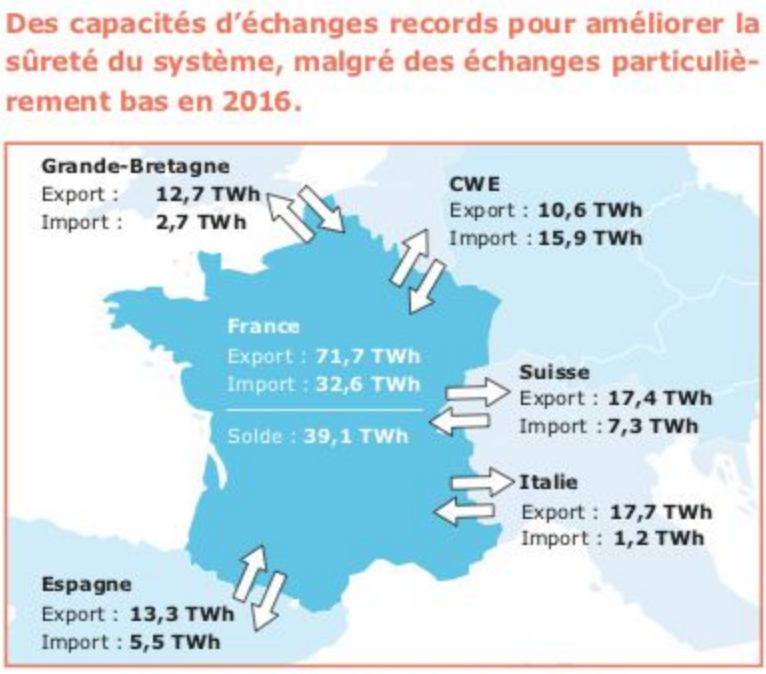
serveurs, à des attaques hors ligne, des dénis de service par modification des configurations réseau…

10. Des OS et firmware obsolètes et non mis à jour. Si Windows XP est encore très présent dans le monde industriel, Lexsi constate également encore la présence de Windows 2000 et même NT4. Une obsolescence qui permette des prise en main rapide sur le Scada avec pour conséquence des compromissions instantanée des équipements et le risque de rebondir sur d'autres périmètres.

La société IO Active, dans une étude sur la sécurisation des réseaux intelligents[6], relève dix risques :

1. outre les atteintes à la vie privées
2. l'usurpation d'identité,
3. la surveillance,
4. l'espionnage ou
5. la malveillance en font partie, mais aussi
6. le simple bug informatique et tous les problèmes inhérents à tout système informatique de réseau.

Bilan sûreté RTE 2016: les importateurs nets d'électricité, cibles privilégiées



Bilan des échanges contractuels en 2016

[6] *https://www.techniques-ingenieur.fr/actualite/articles/compteurs-intelligents-des-risques-de-piratage-informatique-16947/*:

Une étude menée par deux chercheurs de Cambridge, spécialistes de la sécurité informatique, Ross Anderson et Shailendra Fuloria, montre que tout réseau de communication et d'information peut faire l'objet d'intrusions plus ou moins malveillantes, et les compteurs intelligents envoient des données précises, mais peuvent aussi recevoir des ordres à distance : « Du point de vue de l'attaquant – gouvernement hostile, organisation terroriste ou de protection de l'environnement –, le meilleur moyen de s'attaquer à un pays est de lui couper l'électricité. C'est l'équivalent, cyber, d'une attaque nucléaire : quand il n'y a plus d'électricité, tout s'arrête. »

Valentin Brehier, ingénieur système chez SDEL Contrôle Commande, une entreprise du réseau OMEXOM[7].

« Il y a des pirates étatiques qui s'installent dans les réseaux électriques qui peuvent passer à l'action à tout moment », avertit pour sa part un ancien responsable de la cybersécurité d'un très grand groupe français qui désire rester anonyme[8].

L'objectif du pirate est de prendre le contrôle de postes à haute tension pour isoler des zones géographiques ciblées (département, région).

« Une région qui a une forte consommation, mais peu de production, doit être connectée à une région qui produit beaucoup, mais consomme peu. Si on parvient à les séparer, l'une surproduit, et la centrale va alors se désynchroniser, et celle qui surconsomme va faire s'écrouler le réseau. Si on arrive à ouvrir les bons disjoncteurs, on peut plonger un pays dans le noir » - Valentin Brehier.

Les experts pensent aujourd'hui que seuls des Etats disposent de moyens techniques et financiers nécessaires à une attaque de grande ampleur. Si une telle attaque survenait en France et conformément à sa doctrine en la matière, Paris pourrait répliquer, y compris militairement[9].

Coût potentiel

.

Un dysfonctionnement ou une défaillance opérationnelle de l'infrastructure énergétique aurait un impact en cascade sur d'autres infrastructures essentielles (transport, approvisionnement en

---

[7] (source:https://www.techniques-ingenieur.fr/actualite/articles/compteurs-intelligents-des-risques-de-piratage-informatique-16947/)

[8] *Source: https://www.lemonde.fr/pixels/article/2017/12/08/le-reseau-electrique-francais-peut-il-etre-pirate_5226462_4408996.html#p9yi395qxkPpPfLG.99*

[9] Source: https://www.lemonde.fr/pixels/article/2017/12/08/le-reseau-electrique-francais-peut-il-etre-pirate_5226462_4408996.html#BVDIywoI4GGLoKJw.99

eau) et dans l'économie (les usines pourraient avoir fermé leurs portes pour économiser l'énergie).

Par exemple, une étude estimant que des attaques malveillantes simultanées sur 50 générateurs dans le nord-est des États-Unis suggère qu'elles pourraient réduire la puissance énergétique à 93 millions de personnes, entraînant des dommages économiques d'au moins 243 milliards $ US.

À titre de comparaison, le tremblement de terre et le tsunami de 2011 au Japon ont coûté 300 milliards de dollars US, alors que le coût de l'ouragan Sandy en 2012 aux Etats-Unis a été de près de 100 milliards de dollars.

Deux études en Anglais de Swiss Re en guise d'avertissement

Swiss Re pour Swiss Reinsurance Company, est une société d'assurance et de réassurance fondée à Zurich en 1863.

Swiss Re / World Energy: Impacts of cyber risks in the energy sector[10]

Impacts	Examples/Illustrations
Market disruption	Hacking into company data on reserves could impact derivatives and future market for oil and gas, and may cause industry-wide problems. Accessing company information on coal reserves could include information related to commodity pricing.
Physical infrastructure damage	Attacks to dams and levees could result in massive property damage and compromise water supply. Gaining control of a wind turbine could change the wind vane speed, damaging the equipment.
National security	Attacks on systems of national interest and critical infrastructure could have significant impacts on a country's economy, international competitiveness, public safety, or national defence and security.
Human harm	An attack on nuclear plant equipment could lead to a core meltdown and dispersal of radioactivity. An infiltration of the electric grid that results in black-outs can cut off access to running water, refrigeration or other services dependent on electricity.
Network effects	Breaching a control system at a generating facility could serve as an access point for another facility that has a larger impact, taking large portions of the grid offline.13 An attack could impact operations of solar panels and cut energy to a given area.
Financial loss, liabilities	Attacks can lead to financial losses including the cost to replace broken equipment and upgrade systems affected by an attack, regulatory fines, loss of business opportunity, and loss of intellectual capital as

[10] Swiss Re, 2016, Worldenergy.org

	well as – in a secondary stage – liability of power producers towards manufactures in case of continued business interruption and delays in manufacturing.

Swiss Re / World Energy: Cyber Risk core scenarios[11]

Risk scenario	Blackout of energy supply following property damage	Fire/explosion
Scenario description	The infiltration of malware into an industrial control system via USB stick /internet provides the attackers with the ability to remotely control the infected processes. A breakdown in a power plant (or power grid) due to malicious remote controlling by hackers leads to a large scale, long-lasting power outage, severe equipment damage, and affects infrastructure and services.	The infiltration of malware into an industrial control system via USB stick /internet provides the attackers with the ability to remotely control the infected processes. A pressure increase due to malicious remote controlling by the hackers leads to an explosion/fire and destroys part of or the whole plant.
Industries affected	Electrical grid/power generation companies and all dependent systems All industries dependent on energy supply	All industries, in particular critical infrastructures such as oil / gas, chemical / pharmaceutical, power generation, pipelines, storage
Consequences	Fire/explosion, machinery breakdown, business interruption, contingent business interruption, bodily injury, third party property damage, environmental damage, loss of profits/bankruptcy leading to shareholder claims	Destruction of plant, business interruption following fire/explosion, contingent business interruption, bodily injury, third party property damage, environmental damage, loss of profits/bankruptcy leading to shareholder claims
Potential contingent business interruption (CBI) claims	Outage of several power plants can be compensated (low CBI) Outage of a large number of plants may result in large CBI, e.g., supply bottleneck to repair power plants, partial shortage of electrical power over certain period	Outage of several refineries /plants can be compensated (no/low CBI) Outage of large number of plants may result in large CBI e.g., interruption of plastic production
Potential insurance claims	Property/engineering, workers' compensation and employers' liability, general/product liability (incl. pollution liability), directors and officers	Property/engineering, workers' compensation and employers' liability, general/product liability (incl. pollution liability), directors and officers

[11] Swiss Re, 2016, Worldenergy.org

Pistes de réflexion

L'enjeu de la maintenance

Mais le vrai problème des réseaux industriels, c'est leur faible niveau de maintenance. Les protocoles industriels sont en retard sur ceux de l'informatique : les notions d'authentification et de chiffrement sont moins fortes dans l'industrie pour des raisons de performance.

« *Le gros défi de ces infrastructures, c'est de les moderniser alors qu'elles ne sont pas conçues pour ça. Dans l'industrie, un système validé et qui répond aux besoins fonctionnels, on n'y touche plus. Si toutefois il est connecté à une passerelle Internet, il n'est plus protégé* », décrit l'ingénieur système de SDEL Contrôle Commande. Il faut donc créer une couche de sécurisation autour de ces anciens systèmes.

Le secteur de l'électricité offre par ailleurs de plus en plus de prises aux attaquants. D'abord parce qu'il doit relever le même défi que le reste de l'industrie : des systèmes vieux de plusieurs décennies soudainement connectés à Internet, des pratiques de mises à jour pas toujours optimales…[12]

Plus fondamentalement, le réseau électrique se complexifie et devient hétérogène avec l'arrivée des réseaux intelligents, compteurs communicants, producteurs d'énergie renouvelables – qui n'ont pas toujours les mêmes contraintes ni les mêmes expériences en matière de sécurité informatique.

Cette multiplication des points d'entrée dans le réseau ajoute autant de portes potentielles pour des pirates. *« On passe de systèmes énergétiques fermés où il y avait très peu de données rendues disponibles et très peu d'organes manœuvrables à distance à des systèmes plus ouverts où il y a de plus en plus d'équipements télé-opérés et donc un potentiel accru d'attaques »,* confirme-t-on à la Commission de la régulation de l'énergie (CRE), le régulateur français du secteur.

Nous savons aujourd'hui que plusieurs dizaines de groupes de pirates s'intéressent au secteur. *« Le secteur énergétique est beaucoup plus visé par les cyberattaques qu'il ne l'était il y a dix ans, notamment parce que les connaissances des attaquants en matière de systèmes industriels ont fortement progressé »,* explique Gabrielle Desarnaud, consultante chez Capgemini et auteure pour l'Institut français des relations internationales d'un rapport sur le

[12] https://www.lemonde.fr/pixels/article/2017/12/08/le-reseau-electrique-francais-peut-il-etre-pirate_5226462_4408996.html

sujet. Désormais, de plus en plus d'assaillants disposent, outre d'un savoir-faire en matière d'espionnage, de capacités de sabotage.

***Les préconisations de WorldEnergy.com: t*outes les parties prenantes clés doivent jouer un rôle actif dans la gestion des cyberrisques:**

Par exemple, les secteurs de l'assurance et financier: adapter la couverture pour faire face à l'évolution continue du cyber risque. Le secteur doit travailler avec l'industrie de l'énergie pour améliorer la connaissance des produits d'assurance cybernétique, développer davantage le marché de la cyber assurance et, de ce fait, soutenir l'industrie énergétique dans la détermination et la collecte des données critiques sur les cyberrisques.

Le secteur doit rester informé des développements technologiques en constante évolution, car ceux-ci informeront les risques assurés. Ils doivent surveiller les cyberrisques couverts par les produits d'assurance existants et, au besoin, les adapter, par exemple au moyen d'une tarification ou d'une limitation, et se concentrer sur la gestion des risques d'accumulation émergents et changeants. Enfin, le secteur de l'assurance et le secteur financier doivent réagir à l'évolution de la cyber-régulation.

Secteur de l'énergie: considérer les cyberrisques comme un risque commercial majeur, évaluer et comprendre efficacement les cyberrisques spécifiques à l'entreprise et développer des stratégies solides de résilience technique et humaine. Les entreprises doivent travailler pour sensibiliser les autres parties prenantes de l'énergie à l'impact des cyberattaques; cela garantira que la communauté de l'énergie au sens large soit incluse dans les mesures de résilience.

Les gouvernements, également, doivent être mobilisés; ils doivent soutenir des réponses fortes des entreprises aux cyberrisques en stimulant l'introduction de normes ou en imposant des réglementations spécifiques. Cependant, les exigences réglementaires et de déclaration ne devraient pas devenir trop complexes pour ce risque dynamique. Les gouvernements doivent soutenir le partage d'informations entre les pays, les secteurs et au sein de l'industrie, et ils doivent améliorer la coopération internationale sur les cadres de cybersécurité.

Les entreprises technologiques au service du secteur de l'énergie doivent intégrer les caractéristiques et les considérations de sécurité lors du développement des technologies et travailler avec le secteur de l'énergie pour utiliser les dernières technologies afin de surveiller la nature des cyberattaques.

Les associations industrielles sont là pour soutenir et stimuler le partage d'informations et l'adoption des meilleures pratiques, conduire des évaluations par les pairs et aider les entreprises et le secteur à développer une culture de cyber-conscience robuste et active.

Le système judiciaire

De la fiction à la réalité

La bonne nouvelle tombe à 16h ! Les données sont enfin accessibles sur le portail « Open data ». Cela fait deux semaines que des informations fuitent sur la mise à disposition prochaine d'un service de téléchargement gratuit. Nous sommes sur le pont à préparer la réception des données depuis près de deux mois. Notre premier prototype est opérationnel et déjà testé sur un échantillon réduit des même données. L'équipe attend cette nouvelle comme un enfant qui attend le matin de Noël.

Et ce lundi après-midi la "loi n° 2016-1321 du 7 octobre 2016 pour une République numérique" est bien appliquée . Les services de l'Etat s'exécutent et publient massivement tous les jugements de premier ressort, d'appel ou de cassation. Avec surprise nous constatons que l'historique sur 5 ans est également disponible. Nous avons accès à plusieurs dizaines de milliers de décisions de justice !

Depuis 2016, et dès l'instant où le gouvernement a annoncé sa volonté de publier gratuitement les données des organisations publiques, dont les décisions de justice, a germé dans mon esprit l'idée d'exploiter ces informations pour développer de nouveaux services et créer ma Start-up. Lorsque j'ai annoncé en 2016 à mon manager que la disponibilité de ces informations serait une mine d'or, sa réaction a été de douter : « *je ne peux pas imaginer que des données judiciaires puissent être publiées ouvertement et gratuitement. Il y a la loi sur la protection des données personnelles qui bloquera l'application de cette loi, ce qui reste n'a pas de valeur* », disait-il.

Il avait raison et tort à la fois. Il se plaçait du point de vue de Monsieur et Madame tout le monde. On ne pouvait pas en effet exposer les démêlés judiciaires de ces derniers publiquement. C'était évidemment prévu par la loi : « Cette mise à disposition du public est **précédée d'une analyse du risque de ré-identification** des personnes ». En somme, une anonymisation des personnes devait précéder la publication. Dans son esprit, on ne pouvait exploiter les données personnelles pour délivrer des services aux entreprises.

Mais en se plaçant de ce point de vue, il n'envisageait pas l'exploitation des autres données contenues dans les décisions de justices. Comme les décisions en elles-mêmes, les noms des avocats, les noms des greffiers, les noms des magistrats et autres noms d'entreprises impliqués dans toutes sorte de procédures judiciaires. Et en me plaçant dans une autre perspective, celle de délivrer des services aux publics et non aux entreprises, j'imagine rapidement à cette époque quelques services.

Des dizaines d'idées traversaient mon esprit pendant mes longues journées ennuyeuses à un poste de développeur junior . Je cherchais à esquisser des applications web ou mobile à développer rapidement sur la base de ces données. Le projet initial qui me paraissait le plus utile, et le plus évident, était un projet de développement de plateforme de comparaison d'avocats et pourquoi pas de magistrats.

Le premier, « AvocatAdvisor » serait mis à disposition du grand public pour lui permettre de choisir un avocat selon des critères objectifs et non plus selon une vague réputation ou un site Web promotionnel. Plus concrètement avec les nouveaux algorithmes d'analyse de texte je peux créer des profils d'avocats selon les thèmes traités par leurs cabinets, réaliser des statistiques, et créer un scoring de pertinence selon le type d'affaires. Qui sait aujourd'hui combien d'affaires le cabinet de Maître X à Paris peut-il prétendre avoir défendues et sur quels thèmes ? Peronne ! Et c'est bien là la valeur d'un service de comparaison : présenter des indicateurs objectifs et quantitatifs pour aider au choix.

Le deuxième, « MagistratAdvisor » serait mis à disposition des avocats pour leur permettre de comparer les magistrats selon des critères objectifs et quantitatifs. Sur la même base technologique et algorithmique je vais pouvoir déployer plusieurs services !

L'excitation et la fièvre de la ruée vers l'or retombées, je décidais en 2016 de commencer à explorer concrètement le sujet et à développer un prototype. Après tout, dans cette décennie de transformation digitale, il n'y a pas de raison que les métiers d'avocat et de magistrat soient épargnés par la déferlante vague du Big Data. Je n'allais pas être seul sur le coup, mais j'avais l'ambition et la volonté affichées de créer ma start-up. Et comme mon manager ne voyait pas le potentiel, j'allais négocier un congé sans solde de 3 mois pour aller jusqu'au bout de mon idée.

Dès l'été 2017, j'avais téléchargé gratuitement deux milles décisions de justice sur le site Legifrance. Cet échantillon est simplement la jurisprudence disponible en ligne. Elle concerne plusieurs Cours de justice en France et représente une palette plutôt large de type de procédures. Je retrouve bien les noms des magistrats, des avocats ainsi que ceux des entreprises. Mon seul problème à cette époque était que chaque avocat ou magistrat n'était représenté que peu de fois dans le corpus. Je ne pouvais donc pas faire de statistiques représentatives, mais commencer toutefois mon prototype sur cette base dans l'attente de la publication massive annoncée.

En trois mois, j'avais créé une solution d'analyse de texte capable de répondre à 70% de mes objectifs. Et simplement en utilisant les algorithmes disponibles, mis à disposition par des communautés de « Data Scientist ». Mes connaissances techniques m'ont permis de créer des applications web rapidement pour mettre en valeur mes deux comparateurs. C'était suffisant à ce stade pour tenter d'obtenir du financement et lancer ma start-up.

Fier de ma création j'étais persuadé de pouvoir convaincre : du haut de mes 25 ans je présentais ma solution au nom évocateur « AvocatAdvisor ». Un nom qui faisait sourire mes

interlocuteur mais ses derniers semblaient impressionnés par les capacités des nouveaux algorithmes à base de réseaux de neurones. Les résultats étaient partiellement au rendez-vous. Mais une initiative personnelle sans fond propre n'était pas considérée comme suffisamment sérieuse.

Un an après, l'annonce de la disponibilité des données qui me manquaient était la nouvelle qui me permettait de relancer mon projet. Dorénavant, nous sommes deux associés développant une application de profilage. Nous avons renoncé au nom ridicule de « AvocatAdvisor », et avons décidé de travailler exclusivement en sous traitance d'un groupe international dont un des services est un annuaire d'entreprises. Ils intègrent le résultat de nos traitement dans leurs base de données pour augmenter la pertinence des notes qu'ils attribuent aux entreprise. Nous n'avons pas de vision exact de leur utilisation des données, mais un des points clés de notre nouveau cahier des charges est de leurs permettre d'exploiter les informations relatives aux entreprises impliquées dans des procédures judiciaires.

Exemple réel

Pour illustrer concrètement la fiction ci-dessus, à titre d'exemple, nous allons analyser une décision de justice à l'aide de ce type d'algorithme, décision à télécharger sur le site légifrance, Numéro 07/00194.

Sans avoir à lire le contenu de cette décision, observons les résultats extraits automatiquement. Pour faciliter la lecture nous utilisons une représentation visuelle compréhensible dans le cadre de ce document.

1. *Synthèse* : Sans lire le document, on constate qu'une **société** "**CONFORMA**" est

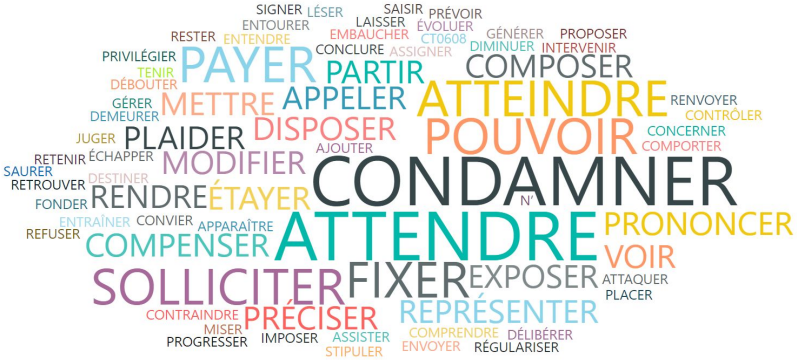
concernée, que **DAVID** est un principal sujet et que cela concerne la **Rémunération** d'une personne.

2. *Les personnes, lieux, ou organisations :*

On distingue facilement les parties prenantes de la décision de justice: **David**, un **employeur**, un **salarié**, la **ville** d'Ajaccio et leurs **représentants**. Également les noms et prénoms des magistrats, avocats et greffiers.

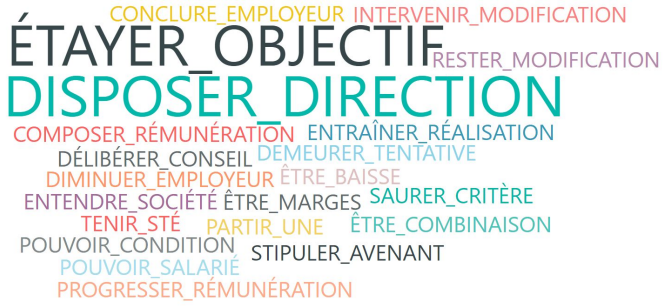
3. *Les principales actions*

L'algorithme souligne rapidement les infinitifs des verbes utilisés : l'infinitif pour faciliter les comparaisons sur une base commune en faisant abstraction des personnes et des temps, qui sont stockés sous forme de Métadonnées.



4. *Les sujets et compléments d'objet direct*

Avec une analyse à 2 niveaux on peut préciser 'Qui subit l'action accomplie par le sujet ?



5. *La lecture humaine et lecture automatique*

En examinant la décision de justice en question, on comprend que la société Conforama a été condamnée à payer une compensation de salaire à son employé David X. On retrouve cette idée dans les différents éléments de langage. Même si la lisibilité n'est pas facilité pour l'homme, les algorithmes se chargent d'ingérer ce format brute et sont capables de le transformer en analyse. A l'échelle de millions de documents par an !

Cet exemple simple démontre qu'il est possible de structurer un texte et de l'indexer dans une base de données. En y ajoutant quelques algorithmes intelligents et des statistiques, il sera possible d'aller plus loin en terme de fouille et de classification...

Pistes de réflexions

L'esprit de la loi est de mettre à disposition du public à titre gratuit les décisions de justice du premier ressort, d'appel ou de cassation. En respectant la vie privée des personnes et s'imposant une analyse du risque de réidentification des personnes.

Oui mais cette démarche pose un véritable problème pour les magistrats, les avocats et les entreprises qui ne sont pas concernés explicitement par la procédure d'anonymisation.

Dans l'état actuel de l'application de la loi, se posera un choix politique d'inclure ou exclure les professionnels du système judiciaire dans le processus d'anonymisation. L'application de cette loi d'une République Numérique devra également s'opposer à l'interprétation de la loi GDPR 2018.

Références

- Article 20 de la LOI n° 2016-1321 :

  L'article L. 10 du code de justice administrative est complété par quatre alinéas ainsi rédigés :

  « Ces jugements sont mis à la disposition du **public** à titre **gratuit** dans le **respect de la vie privée des personnes** concernées.

  « Cette mise à disposition du public est **précédée d'une analyse du risque de ré-identification** des personnes.

  « Les articles L. 321-1 à L. 326-1 du code des relations entre le public et l'administration sont également applicables à la **réutilisation des informations publique**s figurant dans ces jugements.

  « **Un décret** en Conseil d'Etat fixe, pour les jugements de premier ressort, d'appel ou de cassation, les conditions d'application du présent article. »

- Article 21

  Le chapitre unique du titre Ier du livre Ier du code de l'organisation judiciaire est complété par un article L. 111-13 ainsi rédigé :

« Art. L. 111-13. - Sans préjudice des dispositions particulières qui régissent l'accès aux décisions de justice et leur publicité, les décisions rendues par les juridictions judiciaires sont mises à la disposition du public à titre gratuit dans le respect de la vie privée des personnes concernées.
« Cette mise à disposition du public est précédée d'une analyse du risque de ré-identification des personnes.
« Les articles L. 321-1 à L. 326-1 du code des relations entre le public et l'administration sont également applicables à la réutilisation des informations publiques figurant dans ces décisions.
« Un décret en Conseil d'Etat fixe, pour les décisions de premier ressort, d'appel ou de cassation, les conditions d'application du présent article. »

Les entreprises

Un cadrage précis mais un contexte complexe

L'ouverture des données est une source d'innovation et de croissance qui permet et encourage le développement de nouveaux usages à partir de données mises à disposition d'un écosystème lui-même ouvert.

Pourtant, derrière l'apparente simplicité de l'intention, il découle de cet écosystème une extrême complexité qu'il convient de mettre en lumière. C'est en appréhendant cette complexité que l'on peut déduire les risques qui y sont liés.

Le coût élevé des données pivots, données de référence.
Les "données pivots", ou données de référence, sont des données considérées comme identifiantes, par l'administration ou par l'usager, pour nommer ou identifier des produits, des entités économiques, des territoires ou des acteurs (personnes physiques et morales). L'utilisation de ces données ne peut être mise en place que s'il n'existe aucun obstacle financier ou réglementaire et que leur qualité est irréprochable. Les coûts d'extraction et de fabrication de ces données essentielles étant élevés, l'État se concentre donc sur ce type de données pour créer un véritable effet de levier dans la réutilisation des données publiques en France. Ces données, dont le GFII a publié une liste[13], doivent être à 100 % en open data, c'est-à-dire lisibles par machine, livrables par API, exhaustives avec une mise à jour rapide et, point capital, gratuites.

La difficulté d'identifier le périmètre des données présentant un potentiel économique
Au-delà des données de référence dont il est demandé une complète ouverture, il existe de nombreux jeux de données détenus par l'État et ses établissements publics, les collectivités territoriales ou des organismes chargés de missions de service public. Ces données présentent souvent un potentiel de croissance qui reste ignoré en raison de l'absence totale d'un cadre de diffusion couvrant les aspects économiques, techniques et juridiques potentiellement constitutifs d'une véritable économie de la donnée.

Des craintes et des incertitudes qui ralentissent le processus d'ouverture des données
Malgré l'élaboration de textes, notamment la transposition de la directive de 2013 par la loi 2015-1779, la loi Valter ou la loi "pour une République numérique", malgré la création d'un service du Premier ministre, de nombreuses démarches de communication de collectivités locales, l'ouverture des données soulève des réserves. Malgré les avancées, les attentes ne sont manifestement pas si nombreuses et les craintes liées aux données personnelles des

[13] http://www.gfii.fr/uploads/docs/GFII_Donneespivots.pdf

citoyens se dressent face à une promesse hypothétique de croissance. Les textes et les débats n'ont pas suscité de démarche favorable à plus d'ouverture des données du côté des producteurs publics, sans doute en raison de moyens insuffisants, mais aussi en raison d'une crainte de plus d'insécurité lorsqu'il s'agit du secteur privé[14].

Une valeur économique encore peu probante.
La mesure de la valeur de l'open data peut s'établir selon deux axes selon une étude Bluenove-BVA[15].

1. Soit c'est la **valeur du marché des informations publiques** qui est mesurée (par exemple dans le cadre de l'étude MEPSIR de 2006[16]) alors que nombre d'entre elles ne sont pas ouvertes.
2. Soit ce sont les **gains d'opportunité** qui sont comptabilisés (Mc Kinsey Global Institute).

Suivant cette deuxième logique, les rapports se sont succédés et concluent tous à des montants colossaux. McKinsey en 2013, situait entre 3 220 et 5 290 milliards de dollars la valeur annuelle de l'open data.

Cependant, les gouvernements ont été nombreux à mettre en place des politiques d'ouverture des données publiques pour des raisons politiques (une gouvernance plus ouverte), économiques (permettre le développement de nouvelles activités qui permettront des bénéfices économiques pour l'état et la société) ou pour impulser la modernisation de l'action publique. Cependant, on ne constate que peu d'exemples probants de valeur économique générée par des données publiques ouvertes.

Quelles sont les startups qui doivent tout à l'Open Data ? De même, les entreprises ne s'engagent pas de façon très nette dans l'ouverture des données qu'elles détiennent, génèrent ou manipulent, à l'exception de quelques acteurs du secteur des transports et de la mobilité. Plus encore, on a parfois du mal à comprendre si l'open crée de la valeur pour les données… ou alors en détruit (par la logique de gratuité notamment).

La difficile mesure d'impact économique et de la valeur de l'Open Data
Certaines réutilisations des données publiques sont particulièrement évidentes quand il s'agit d'applications mobiles grand public qui le revendiquent. En revanche, la majorité des réutilisations de données publiques ouvertes n'est pas communiquée à l'extérieur de l'organisation qui les utilise. On mesure donc difficilement ce qui ne nous est pas donné à voir.

[14] http://utp.fr/system/files/Actu/Position/20160310_Position_Loi_Lemaire.pdf

[15] Etude Bluenove-BVA *Open data, quels enjeux et opportunités pour l'entreprise ?* Une initiative de bluenove, en partenariat avec SNCF, Le Groupe La Poste, Suez Environnement, Groupe Poult sur la base de l'étude bluenove – BVA, publiée en *novembre 2011*

[16] MEPSIR Measuring European Public Sector Information Resources, Final Report of Study on Exploitation of public sector information – benchmarking of EU framework condition, Juin 2006

Cette opacité peut devenir problématique dès lors que l'articulation du modèle « open data » n'est même pas identifié.

Toutefois, des démarchent existent pour mieux identifier et révéler les utilisations, notamment la plateforme ouverte des données publiques data.gouv.fr qui permet aux utilisateurs de mettre en ligne les réutilisations qu'ils ont faites des données mises à disposition.

Les risques liés à l'Open data

Même si l'ouverture des données semble faire consensus et permettre la transparence de l'action publique et le développement du marché de l'information, nous venons de mettre en lumière les principaux obstacles à son développement et à son modèle de croissance. Voyons maintenant quels sont les principaux risques liés à la démarche d'ouverture des données.

1. **Le risque lié à la transparence**

L'Open data peut induire un changement des rapports de force entre les citoyens et l'Etat quant aux décisions à prendre, et notamment quant à l'évaluation et la légitimité des politiques de service public et de leur efficacité en fonction des données chiffrées disponibles. Face aux données, le citoyen peut être tenté de discuter du bien fondé d'une décision. Par exemple, un groupement de citoyens pourrait remettre en question la nécessité d'ouvrir un commerce ou un bureau de poste en fonction des données démographiques disponibles. Le gouvernement pourrait donc être confronté à une perte de pouvoir sur les données et sur la prise de décision en général.

Plus encore, les ré-utilisateurs peuvent détourner les informations fournies afin de produire des sites intrusifs ou discriminants et source d'inquiétudes pour les citoyens. L'exemple des Etats-Unis, où les réutilisateurs se sont rués sur les données de criminalité pour produire des cartographies par ville ou recensant l'origine ethnique des habitants d'un quartier, est problématique.

Enfin, l'ouverture des données pourrait avoir pour conséquence sur le long terme la privatisation des services publics. L'accès à ces données et leur réutilisation par des entreprises privées pourraient laisser penser que l'Etat n'a plus le monopole des services publics, le privé pouvant alors assumer certaines fonctions, et notamment la production de données publiques pertinentes agencées et remaniées en leur faveur. L'exigence de clarté et de transparence ne pèse dès lors que sur le secteur public. Pourtant, la communication de leurs données par les entreprises privées pourrait aussi s'avérer aussi fort utile. C'est le cas des opérateurs de télécommunications qui disposent de données anonymes précises sur la fréquentation des quartiers de la ville, sur le trafic sur les grands axes routiers, ainsi que des fournisseurs d'énergie ou des entreprises de collecte des déchets. Dès lors, il existe un risque de

déséquilibre entre l'administration dont le travail est totalement dévoilé et les entreprises privées qui continueront à garder la maîtrise de leurs données.

2. **Le risque lié à la vie privée**

Certaines données ne sont pas concernées par ce mouvement d'ouverture : les informations détenues par des organismes culturels ou d'enseignements et de recherche, les données sur lesquelles des tiers détiennent des droits de propriété intellectuelle, et naturellement les documents contenant des données personnelles.

Toutefois, il est parfois délicat de faire la part entre données publiques et données personnelles. Le débat suscité par l'adoption de la LOPPSI, loi d'orientation et de programmation pour la performance de la sécurité intérieure, en témoigne. En effet, un de ses articles a suscité de vives réactions car il menaçait structurellement l'Open data en France en proposant de soumettre les réutilisateurs de données publiques sous licence à un contrôle de comportement (de moralité). Plusieurs amendements ont été déposés. Un amendement de la commission des lois a été adopté afin de restreindre ces contrôles de moralité à la seule réutilisation des données d'immatriculation des véhicules.

En effet, l'Open data ne subissait ici que le dommage collatéral de la légalisation d'une pratique : l'utilisation et la vente par l'État **des données personnelles** relatives aux cartes grises à des fins de marketing. Une confusion entre données personnelles et publiques est créée. Il existe donc un risque pour que l'Open data concerne parfois des données pouvant être qualifiées de personnelles puisque même pour le législateur, on constate des points de confusion.

3. **Le risque lié à la monétisation des données**

Dans le cadre de sa mission de mise en valeur du patrimoine numérique public, l'Agence du patrimoine immatériel de l'Etat (APIE) a proposé deux modèles de licences de réutilisation des données publiques basées sur le paiement de redevances, sauf dans l'hypothèse d'une réutilisation non commerciale. Cette orientation consistant à marchander les données publiques, alors qu'elles sont le fruit de l'exercice du service public financé par tous n'est pas sans risque. En effet, il ne faudrait pas qu'un tel système empêche les petits acteurs d'obtenir une licence trop chère.

Ceci créerait **un déséquilibre entre les grandes entreprises pouvant librement s'offrir des données publiques exploitables et les petites ne pouvant y accéder**. La question de la tarification est centrale, et la réutilisation des données par un maximum d'acteurs en dépendra. Le risque étant que l'exploitation de ces données ne soit réservée qu'à des lobbies qui, grâce à leur puissance et leurs moyens financiers, en fassent un usage à leur avantage au détriment des instituts publics de production de données géographiques ou statistiques qui faute de budget devront diminuer leurs activités. Afin d'éviter ce risque, certains préconisent l'utilisation de modèles de licences libres.

4. **Le risque lié à la fiabilité et à la complexité des données**

En ouvrant au public ses données, il pourrait être mise en évidence un certain manque de fiabilité qui pourrait discréditer l'action publique et remettre en cause le travail des administrations.
D'autre part, les données pourraient aussi être difficiles à comprendre car trop complexes, si bien que les dépenses mises en œuvre par l'Etat pour assurer leur mise à disposition du public ne serviraient pas au citoyen qui ne peut les interpréter mais uniquement à des spécialistes, souvent financés par des entreprises puissantes. Ici encore, un déséquilibre dans l'accès aux données pourrait naître. A terme, on pourrait redouter que ces entreprises falsifient les données avant de les communiquer au public afin de les tourner à leur avantage.

5. **Le risque lié à une utilisation massive et abusive des données**

Un risque majeur est de voir des entreprises étrangères profiter à moindre coût de ces données pour en faire un usage commercial, parfois abusif en ce qu'il serait à la limite du pillage ou de l'espionnage. Les données publiques françaises seraient alors exploitées par des groupes étrangers sans que cela ne rapporte de bénéfices à l'Etat puisque le propre de ces entreprises est de ne pas payer d'impôts en France. L'exemple constamment cité est celui de Google qui est souvent accusé de « pillage » en raison de ces litiges l'opposant aux titulaires de marques ou aux éditeurs. Ce risque d'exploitation massive et intrusive des données est bien réel.

Il est capital que tous les acteurs (réutilisateurs et pouvoirs publics) prennent conscience de ces risques pour les limiter et faire de l'Open data un mouvement juste et équilibré.

Quand l'entreprise met en scène l'Open Data

Voici trois scénarios fictifs et un cas d'usage réel mettant en scène l'Open Data. Par cette incarnation, il est plus immédiat de mettre en lumière les bénéfices, les inconvénients et les risques de certains usages, notamment en termes de discrimination ou de contrôle de l'ordre public.

Scénario 1

Une entreprise fabriquant des outils de navigation fonde une partie de son business model sur la revente des données de ses clients et de ses utilisateurs à d'autres acteurs économiques.
La vente de GPS ne correspond qu'à 10% de ses revenus quand le business de la data qu'il opère est en très forte croissance.
Parmi ses clients, l'entreprise compte la police qui utilise les données d'utilisateurs pour positionner ses radars sur les lieux où les dépassements sont les plus fréquents.

Scénario 2
Dans le cadre de la valorisation de son catalogue de données publiques et pour en augmenter l'utilité auprès de ses citoyens, la ville de Lille a initié un concours d'applications. Pendant un mois, chaque citoyen ou entreprise locale est invité à s'exprimer et à présenter une idée d'application « pour améliorer la ville de Lille ». Lors de la 1ère édition du concours, 57 applications ont été générées pour une valeur totale estimée à 2 millions d'euros. Les vainqueurs du concours ont bénéficié de subventions de la ville pour développer les applications. Parmi elles, on a noté une application permettant de savoir en temps réel s'il y a une file d'attente dans les restaurants ou encore de savoir quel restaurant a reçu une amende sanitaire. Toutes ces applications utilisent les données ouvertes rendues publiques par la ville de Lille.

Scénario 3
La SNCF incube une startup chargée de développer des services à forte valeur ajoutée pour améliorer le quotidien de ses usagers. Par exemple, une application permet, dans le cadre d'un projet immobilier, de savoir quels sont les terrains constructibles proches de gares TER. Cela permet de mieux orienter son choix en amont de son investissement immobilier.
Une autre application permet de situer tous les points de services (boulangeries, supermarchés, banques…) entre ma gare d'arrivée et mon domicile.
Tous ces services applicatifs ont un objectif commun : développer des services connectés à la mobilité (en TER) de chacun.

Un cas d'usage réel
Une entreprise a développé pour la CUB - Communauté Urbain de Bordeaux - un service de télécontrôle qui permet de lutter contre les inondations et aide au pilotage en temps réel du réseau d'assainissement. Des capteurs remontent les informations liées au fonctionnement des réseaux ou au niveau de la Garonne par exemple, croisées avec des données météorologiques et des cartes pluviométriques. Le tout alimente des systèmes de calculs complexes et de visualisation sur un mur d'images. Ces prévisions permettent d'anticiper les risques pour les habitants et de gérer la sécurité des agents. Plusieurs de ces données sources sont candidates à être réutilisées dans le programme Open Data de la CUB.

Les bonnes pratiques pour réduire le risque

Les bonnes pratiques couvrent à la fois des aspects stratégiques, économiques, commerciaux, marketing, juridiques ou techniques. Certaines consistent avant tout à se poser les bonnes questions avant d'avancer.

Ouvrir les données, c'est créer de nouveaux produits, développer de nouveaux services qui doivent être clairement définis. Il est donc indispensable de répondre aux questions suivantes pour réduire les risques évoqués plus haut.

- Quelles données publier ?
- Sont-elles disponibles ?
- Dans quel(s) format(s) ?
- A quelle fréquence ?
- Comment garantir la pérennité de l'approvisionnement et de la fourniture de ces données?
- Qui sont les réutilisateurs potentiels ?
- Quelle(s) licence(s) d'utilisation faut-il accorder ?
- Comment les contacter et évaluer avec eux :
    - Leur intérêt pour les données
    - De quels outils ont-ils besoin pour accéder aux données
    - Le potentiel de création de valeur
    - L'utilisation qu'ils pourraient en faire
    - Les applications qui pourraient en découler

En synthèse, l'Open Data en entreprise répond à des enjeux commerciaux liés à des retours sur investissement et donc à des arbitrages budgétaires.

Le manque de connaissance de la démarche d'ouverture des données et la méconnaissance des enjeux risquent de limiter les investissements en la matière alors que seulement 34% soit 1/3 des responsables interrogés[17] déclarent que leur entreprise n'a pas besoin d'une démarche d'ouverture de données.

Concernant la sphère des risques, il paraît clair que les usages et exploitations doivent être définis et cadrés bien en amont. Comme le dit Françoise Tournassoud Responsable Information Voyageurs et Relation Clients à la Direction des Services Transilien : « les données temps réel sont un sujet sensible, car en lien avec l'exploitation. Réorienter les usagers vers un mode de transport, donner des informations conjoncturelles est en lien fort avec les prises en charge par les exploitants ; les usages et utilisations doivent être réfléchis en amont. »[18]

Enfin, il est important de considérer les effets systémiques de l'ouverture des données en particulier lorsqu'il s'agit de données en temps réel.

- Quels impacts l'utilisation de ces données pourrait avoir sur l'activité de l'entreprise?
- Quelles rétroactions peut-on anticiper ? Avec quels effets ?

Autant de questions complexes qu'il faut avoir résolues avant de s'engager dans une voie risquée.

Face à ces opportunités, les entreprises doivent se préoccuper de plusieurs points clés pour réduire les risques.

[17] Etude Bluenove-BVA *Open data, quels enjeux et opportunités pour l'entreprise ?* Une initiative de bluenove, en partenariat avec SNCF, Le Groupe La Poste, Suez Environnement, Groupe Poult sur la base de l'étude bluenove – BVA, publiée en *novembre 2011*

[18] Ibid

1. L'agrégation maîtrisée des données

L'un des facteurs d'intérêt de la réutilisation de données publiques est l'agrégation de plusieurs jeux de données pour générer une information nouvelle voire inédite. Mais cette agrégation pose le problème de la dispersion des données : sources et formats multiples, fréquences des mises à jour, évolution des jeux de données, traitements nécessaires à l'intégration. Autant de difficultés à résoudre avant de prétendre offrir un service stable et de qualité.

2. Le respect du cadre juridique

Il est indispensable de vérifier au préalable le cadre juridique dans lequel se place le propriétaire des données et quels sont les droits d'utilisation qu'il accorde à des tiers. A titre d'exemple, les informations de trafic routier en temps réel publiées par la Direction Interdépartementale des Routes Île de France (DIRIF) disponibles sur le site sytadin.fr sont couvertes par le droit d'auteur « Creative Commons » et peuvent être réutilisées « à condition de citer la référence à la DIRIF, n'effectuer aucune modification, adaptation ou travail dérivé et n'en faire aucune utilisation commerciale ».

3. Le cadrage des données privées et sensibles

L'enregistrement, la conservation et l'utilisation de données nominatives sont strictement encadrés par la loi Informatique et Liberté[19] sous la responsabilité de la CNIL. En particulier, la loi précise que « Les informations ne peuvent pas être réutilisées de manière incompatible avec la finalité pour laquelle elles ont été collectées. » Il faut être particulièrement vigilant à ce propos dans la mesure où l'agrégation de données, dont des données nominatives, pourraient amener à utiliser ces informations à des fins très éloignées des objectifs de départ.

4. L'investissement raisonné

Certains des moyens évoqués peuvent être dérivés des moyens informatiques existants dans l'entreprise. Mais de nouveaux investissements et de nouvelles dépenses sont à envisager pour mettre en œuvre une stratégie d'ouverture des données. Alors que la majorité des entreprises n'attendent pas de retour tangibles et de ROI mesurable et alors que les budgets informatiques des entreprises sont sous pression, il est important de budgétiser ces dépenses pour en vérifier la faisabilité et la pérennité.

Focus numérique

Pour la mise en place d'une stratégie d'ouverture de données efficace et où le risque est maîtrisé, il est essentiel de mettre en œuvre une chaîne de production informatique sans faille. Celle-ci doit recueillir les données, les formater, les nettoyer et les mettre à disposition au sein d'infrastructure capable de tenir ces données à disposition des réutilisateurs.

5. Le processus de recueil des données

[19] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée en 2004

Il comprend le recensement et l'identification des sources, le choix des données à ouvrir, la vérification de la stabilité des sources d'approvisionnement, l'audit de la qualité des données, l'inventaire des tâches de transformation des données avant leur ouverture (nettoyage, formatage, exclusion, vérification…). Sécuriser les données : Les données ouvertes doivent être sécurisées dans de multiples domaines. Il s'agit en effet de garantir que les données ouvertes sont bien de bout en bout celles que l'entreprise a voulu ouvrir. Il s'agit donc de sécuriser la ou les sources et de garantir l'intégrité des données tout au long de leur traitement jusqu'à la mise à disposition afin que les utilisateurs disposent bien des bonnes informations et non de données falsifiées.

6. **L'infrastructure de mise à disposition des données**

Il ne s'agit pas seulement de sécuriser le système d'information au sein de l'entreprise mais d'établir des process de sécurité numérique qui dépassent les limites de l'entreprise. Selon le type de données, leurs formats et leurs fréquences de mise à jour, il sera nécessaire de déployer une infrastructure ouverte. Cette infrastructure doit être en mesure de supporter une éventuelle montée en charge importante, capable de tenir les informations à jour (en particulier dans le cas de données temps réel) et d'offrir un accès stable (constant) aux données soit sous la forme de fichiers consultables / téléchargeables soit sous la forme d'API pour permettre aux applications externes d'accéder aux informations dont elles ont besoin.

Le cadre légal

Malgré toutes les démarches d'ouverture des données publiques en France, la question économique reste entière. Pourquoi est-il difficile de déterminer l'impact économique de l'Open Data ?

Les services de l'Etat et des collectivités locales ont progressivement favorisé la réutilisation de leurs données publiques. La **loi du 17 juillet 1978** qui prévoit la liberté d'accès aux documents administratifs régulé par la Commission d'accès aux documents administratifs (CADA) a été modifiée en 2005 (transposition de la directive européenne du 17 novembre 2003) afin de prévoir un **droit de réutilisation** de ces données publiques. La loi précise que, pour se faire, les administrations sont libres de fixer des redevances qui doivent être décrites dans une licence fournie par l'établissement public aux réutilisateurs. En général, la réutilisation des données pour un usage non commercial est libre de toute redevance.

Le système mis en place permet aux citoyens, aux acteurs économiques, aux chercheurs, ou à la presse de se saisir des données publiques. Dans une société où le web mobile se développe à grande vitesse, et où les technologies permettent un traitement rapide d'un grand nombre d'informations, la réutilisation des données publiques est une promesse de croissance.

La loi pour une République numérique a été promulguée le 7 octobre 2016. Elle prépare le pays aux enjeux de la transition numérique et de l'économie de demain. Elle promeut l'innovation et le développement de l'économie numérique, une société numérique ouverte, fiable et protectrice des droits des citoyens. Elle vise également à garantir l'accès de tous, dans tous les territoires, aux opportunités liées au numérique.

Ce qu'apporte la loi
La consultation publique et le travail parlementaire réalisés sur le projet de loi, ont conforté l'ambition initiale du Gouvernement.

- **Libérer l'innovation** en faisant circuler les informations et les savoirs, pour armer la France face aux enjeux globaux de l'économie de la donnée.
  - **Créer un cadre de confiance** clair, garant de droits des utilisateurs et protecteur des données personnelles.
  - **Construire une République numérique ouverte et inclusive**, pour que les opportunités liées à la transition numérique profitent au plus grand nombre.

La loi numérique crée l'obligation pour les organisations publiques de publier sur internet leurs bases de données, sous réserve d'anonymisation et de protection de la propriété intellectuelle et du secret industriel et commercial. Ces données pourront ainsi être exploitées et réutilisées facilement par chacun, particulier comme entreprise. Certains acteurs privés (entreprises titulaires des marchés publics, bénéficiaires de subventions publiques…) seront également tenus de communiquer des données d'intérêt général, qui pourront concerner l'exploitation des services publics de l'énergie ou de l'eau, les transactions immobilières, ou encore la gestion et le recyclage des déchets.

Pistes de réflexion

L'ouverture des données est une des tendances qui va le plus impacter le fonctionnement des organisations publiques et des entreprises dans les prochaines années. Le mouvement est mondial et s'inscrit dans une dynamique plus vaste d'ouverture des échanges.

Si l'on en croit les visions prospectives convergentes de certains cabinets de conseil, les données ouvertes prennent progressivement une place prépondérante et seront bientôt considérées comme une ressource essentielle. Pour autant, les entreprises avancent prudemment sur un terrain complexe où il reste encore beaucoup à explorer. Au-delà des obligations légales et réglementaires s'appliquant à toutes les entreprises ou aux administrations publiques, l'ouverture volontaire des données reste une stratégie d'innovation. Il est néanmoins fondamental que la démarche soit encadrée pour réduire les logiques de prise

de risques qui peuvent nuire à la promesse de croissance économique aussi bien qu'à l'intention démocratique initiale.

Conclusion

L'ouverture des données publiques par nos administrations reste un vecteur essentiel de transparence et un relai informationnel entre l'Etat et les citoyens. Cette mise à disposition des jeux de données, permet de dynamiser le marché de l'innovation en impliquant le monde civil, amateurs comme professionnels à trouver des solutions à des problématiques connues ou encore méconnues. "L'innovation ne viendra pas des entreprises"[20], ni des institutions d'ailleurs. Leurs complexités respectives les en empêchent. Il ne faut pas oublier cependant que la valeur d'une innovation réside en partie dans son partage et sa mise en contexte. Les risques inhérents à ces ouvertures ne doivent cependant pas nous freiner dans le partage mais nous pousser plutôt à trouver de nouveaux moyens pour encadrer ces ouvertures et les contrôler. C'est en transformant l'ouverture de données en atouts et en opportunités et en sensibilisant les utilisateurs, développeurs ou consommateurs que nous réussiront à maitriser la chaine de communication de bout en bout.

L'ouverture des données publique dans l'éducation apporte principalement une aide à la décision, pour les parents, les élèves mais également pour les établissements en leur permettant d'améliorer l'accès, la logistique ou le classement. Malgré cela, de nombreuses questions se posent quant au risque majeur d'attaques physiques ou de détournement d'élites et c'est dans ce contexte qu'il parait obligatoire d'établir une traçabilité des requêtes sur l'ensemble des données publiques de l'éducation.

La France n'a pas encore été touchée par une attaque d'ampleur, paralysante, dans le secteur de l'énergie, peut-être de par sa force de dissuasion. On sait pourtant que les réseaux industriels de production et de distribution d'énergie sont exposés, et vulnérables. La prise de conscience naissante du secteur en matière de nouvelles technologies, ainsi que la mobilisation des secteurs connexes que sont l'assurance, la puissance publique et les nouvelles technologies de l'information sont les clés d'une plus grande maîtrise des risques liés à la libération des données.

En regardant le point de vue du monde juridique, l'ouverture des données est une perspective positive pour les citoyens. Elle leur permet de se réapproprier l'état effectif de l'application des lois dans leur pays induisant ainsi une amélioration, une cohérence et une meilleure transparence autour de la jurisprudence. Mais des questions se posent quant à la protection des identités de ceux qui appliquent ces lois. Comment concilier transparence de l'information et anonymisation des noms des magistrats, des greffiers et des avocats ?

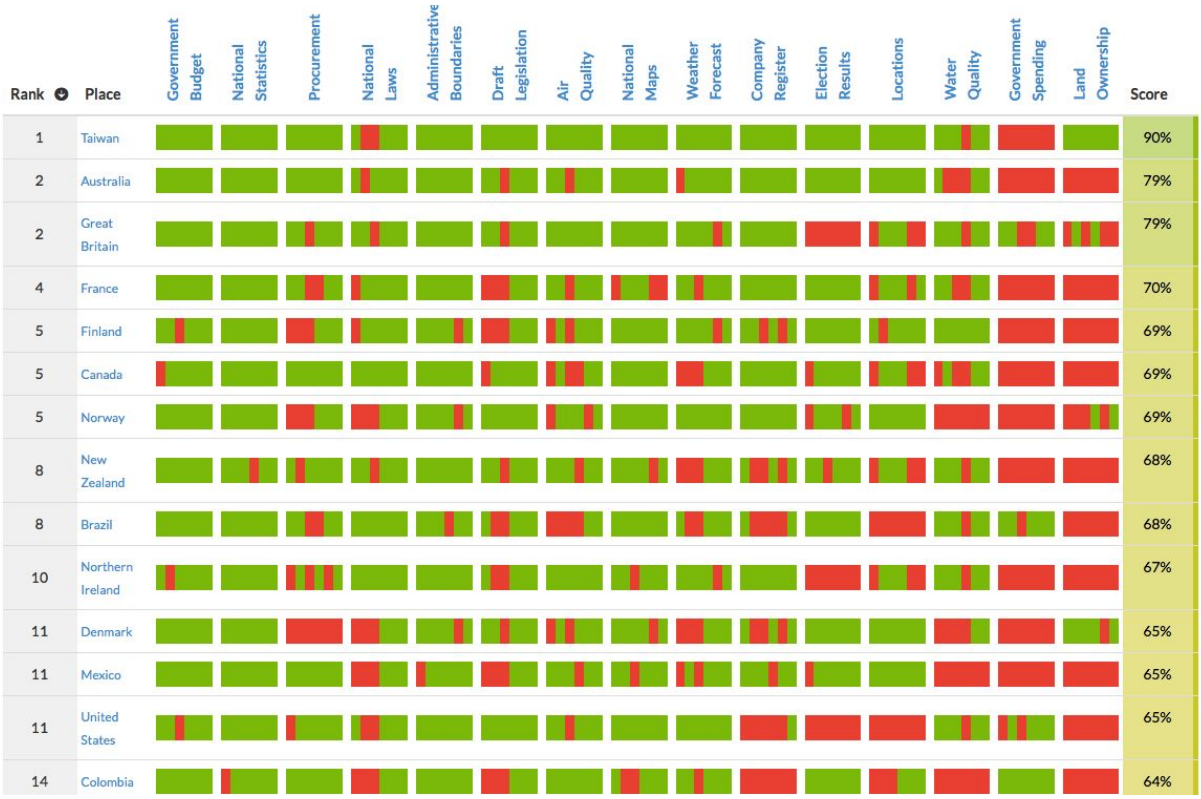
Pour l'entreprise, le principal bénéfice de l'ouverture des données réside dans le fait qu'elle représente une source d'innovation et de croissance. Cette croissance doit permettre de développer de nouveaux services et usages pour mener à bien un projet de puissance

[20] Citation de Guy-Philippe Goldstein.

économique numérique pour la France. Cependant, l'extrême complexité du paysage et la difficulté à développer un cadre réglementaire exhaustif suscitent une prudence de mise de la part des entreprises.
C'est dans l'encadrement de la démarche globale et dans la courbe d'expérience des entreprises pilotes que l'on pourra défricher le territoire encore pour partie inexploré du modèle vertueux. La mise en lumière des risques tels que le risque lié à la transparence, le risque lié aux données personnelles, ou encore le risque de marchandisation abusive permet de contribuer efficacement à une démarche maîtrisée et profitable de croissance économique.

ANNEXES

Données publiques: la France 4e puissance mondiale en nombre de jeux de données publiées



Rank	Place	Government Budget	National Statistics	Procurement	National Laws	Administrative Boundaries	Draft Legislation	Air Quality	National Maps	Weather Forecast	Company Register	Election Results	Locations	Water Quality	Government Spending	Land Ownership	Score
1	Taiwan																90%
2	Australia																79%
2	Great Britain																79%
4	France																70%
5	Finland																69%
5	Canada																69%
5	Norway																69%
8	New Zealand																68%
8	Brazil																68%
10	Northern Ireland																67%
11	Denmark																65%
11	Mexico																65%
11	United States																65%
14	Colombia																64%

Source: https://index.okfn.org/

Références Entreprise

Loi n° 78-753 du 17 juillet 1978

Loi pour une République numérique dite loi Lemaire

Loi du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public dite loi Valter

Blog de la mission Etalab http://etalab.gouv.fr/

MEPSIR 2006, Measuring European Public Sector Information Resources, Final Report of Study on Exploitation of public sector information – benchmarking of EU framework condition, Juin 2006

Site Datanomics / données ouvertes.info, Simon Chignard

Etude Bluenove-BVA *Open data, quels enjeux et opportunités pour l'entreprise ?* Une initiative de bluenove, en partenariat avec SNCF, Le Groupe La Poste, Suez Environnement, Groupe Poult sur la base de l'étude bluenove – BVA, publiée en *novembre 2011*

Open data: Unlocking innovation and performance with liquid information, McKinsey Global Institute-McKinsey Center for Government-McKinsey Business Technology Office, Octobre 2013

Open Government Data: A Stage Model, Evangelos Kalampokis, Efthimios Tambouris and Konstantinos Tarabanis, Informatics and Telematics Institute, Centre for Research and Technology Hellas, Greece-University of Macedonia, Thessaloniki, Greece

Guide Open Data pour les communes, Glossaire de la donnée publique, Juin 2016