SUJET NATIONAL POUR L'ENSEMBLE DES CENTRES DE GESTION ORGANISATEURS

Technicien principal territorial de 1ère classe
Examen professionnel d'avancement de grade

SESSION 2013

EPREUVE

Rédaction d'un rapport technique portant sur la spécialité au titre de laquelle le candidat concourt. Ce rapport est assorti de propositions opérationnelles.

Durée : 3 heures
Coefficient : 1

SPECIALITE : Ingénierie, Informatique et Systèmes d'information

A LIRE ATTENTIVEMENT AVANT DE TRAITER LE SUJET

Ce document comprend un sujet de 1 page, un dossier de 27 pages.

- **Vous ne devez faire apparaître aucun signe distinctif dans votre copie, ni votre nom ou un nom fictif, ni signature ou paraphe, ni numéro de convocation.**
- **Aucune référence (nom de collectivité, nom de personne, ...) autre que celle figurant le cas échéant sur le sujet ou dans le dossier ne doit apparaître dans votre copie.**
- **Seul l'usage d'un stylo soit noir soit bleu est autorisé (bille, plume ou feutre). L'utilisation d'une autre couleur, pour écrire ou souligner, sera considérée comme un signe distinctif, de même que l'utilisation d'un surligneur.**

Le non-respect des règles ci-dessus peut entraîner l'annulation de la copie par le jury.

Les feuilles de brouillon ne seront en aucun cas prises en compte.

Vous êtes technicien principal territorial de 1ère classe, dans la commune de Techniville.

Les élus de votre collectivité souhaiteraient mettre différentes informations contenues dans les bases de données de la commune à la disposition des habitants.

Dans un premier temps, la Directrice générale des services vous demande de rédiger à son attention, exclusivement à partir des documents ci-joints, un rapport technique sur l'ouverture des données publiques.

10 points

Dans un deuxième temps, elle vous demande d'établir un plan d'action visant à rendre effective la libération des données internes de votre commune.

10 points

Pour traiter cette seconde partie, vous mobiliserez également vos connaissances

Document 1 : « Big data # open data » - pro.01net.com - 26 septembre 2012 - 1 page

Document 2 : « Opendata & Etalab : La guerre française des licences s'achève ! » - www.regardscitoyens.org - 19 octobre 2011 - 2 pages

Document 3 : « Données publiques : une nouvelle économie des services publics en débat » - la gazette.fr - 21 décembre 2010 - 7 pages

Document 4 : « L'open data : c'est quoi ? » - www.data.rennes-metropole.fr - 2012 - 2 pages

Document 5 : « Open data : la transparence démocratique demeure virtuelle » - la gazette.fr - 15 mai 2012 - 4 pages

Document 6 : « Qu'est-ce que le Big Data ?» - www.journaldunet.com - 9 novembre 2011 - 2 pages

Document 7 : « Big data, un nouveau défi pour les entreprises », extrait du dossier "Le décisionnel à l'heure du big data" - www.lemondeinformatique.fr - 2 avril 2012 - 2 pages

Document 8 : « Vous avez dit Big Data ? » - www.alain-bensoussan.com - 2012 - 2 pages

Document 9 : « Big Data : décisionnel de M. Jourdain ou réelle innovation ? » - www.zdnet.fr - 23 juin 2012 - 3 pages

Document 10 : « Jusqu'où s'étend la responsabilité du DSI ? » - www.directeurinformatique.com - 13 avril 2012 - 2 pages

Documents reproduits avec l'autorisation du C.F.C.

Ce document comprend un sujet de 1 page, un dossier de 27 pages.

Certains documents peuvent comporter des renvois à des notes ou à des documents volontairement non fournis car non indispensables à la compréhension du sujet.

01net ENTREPRISES

01net Entreprises > BlogPro

Big data ≠ open data

Vincent Berdot, grand reporter
01 Informatique

Ils font tous les deux le buzz. Ils ont tous les deux la même sonorité. On est donc tenté de les associer. Et, pourtant, ce rapprochement entre big data et open data est plus que tiré par les cheveux. Il se fonde sur une idée communément acquise, et régulièrement reprise à la va-vite dans les conférences, selon laquelle l'ouverture des données (open data) génère de gros volumes d'information (big data) à analyser, à croiser et à modéliser. Un raisonnement qui n'a rien d'évident, au niveau tout d'abord du périmètre d'intervention. Les entreprises sont ainsi de plus en plus associées au phénomène de l'open data.

Au nom de l'open innovation, celles-ci seront incitées à ouvrir leurs données aux clients et partenaires. Sauf exception, cette volonté de collaboration et de transparence reste à prouver. A ce jour, et sûrement pour longtemps, l'open data reste le seul fait des administrations, d'ores et déjà engagées dans la libération des données publiques. Les jeux d'information relâchés sont cependant peu volumineux, car, pour la plupart, circonscrits à une collectivité ou une région. Ils font l'objet de simples consultations ou d'applications locales.

Mis à part les rares sociétés spécialisées dans le référencement, la réconciliation et la corrélation de données publiques, l'open data n'appelle aucun traitement relevant du big data. En revanche, l'inverse peut être vrai. Avec la géolocalisation ou les réseaux sociaux, le spectre des données explose. Et les entreprises cherchent à capturer un maximum d'information. Seulement, comme l'a expliqué le président du GS1 lors de son université d'été, pour que ces données restent de qualité, et ne fassent pas l'objet d'une contestation, elles doivent être contrôlées à la source, directement par les citoyens.

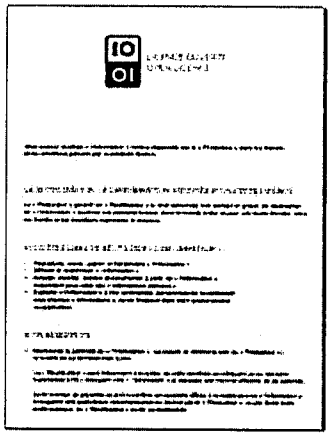
Serait-ce donc le big data qui induit l'open data, et non le contraire ? Dans l'affirmative, cela exigerait un devoir de transparence des entreprises dans leur collecte et leur analyse des profils de consommateur. Une réalité bien moins vendeuse...

DOCUMENT 2

RegardsCitoyens.org

Diffusion et Partage de l'Information Politique

OpenData & Étalab : La guerre française des licences s'achève !



Avec un peu de retard puisque la circulaire demandait sa publication avant le 24 août, la licence d'Étalab pour les données publiques a finalement été rendue publique hier. C'est un signal fort pour l'Open Data français : la « Licence Ouverte » de l'État (PDF) répond globalement aux attentes et demandes de la communauté des réutilisateurs. La France publiera donc prochainement ses données sous une licence de données assurant la compatibilité avec les licences libres existantes, et par conséquent avec un maximum de projets comme Wikipedia ou OpenStreetMap.

Une licence d'État compatible avec toutes les licences libres

Le choix opéré par le gouvernement français est celui d'une licence de type « BY », assurant une très large liberté puisqu'elle ne requiert de la part des réutilisateurs que la seule mention de la source des données (par exemple sous la forme d'un lien vers les données publiées). Le cadre de réutilisation des données prémunit les futurs usagers de tout risque lié au code de propriété intellectuelle, les droits inhérents à la réutilisation des données étant par ailleurs explicitement indiqués. La licence proposée assure ainsi un maximum de compatibilité avec les licences libres existantes en la matière, qu'il s'agisse des Open Data Commons de l'Open Knowledge Foundation comme l'ODbL, des Creative Commons BY ou BY-SA, ou encore de l'Open Government Licence du gouvernement anglais.

OpenData en France, vers un duopole de licences

De nombreuses collectivités comme Paris, Nantes, la Gironde ou bientôt le Grand Toulouse, ont fait quant à elles le choix d'ajouter une clause « SA » de redistribution des données enrichies (modèle « share-alike » de l'ODbL). Contrairement au schéma « BY » plus libéral choisi par Étalab, ce modèle cherche à encourager l'innovation par la redistribution contributive. En ce sens, ce type de schéma est de nature à favoriser l'émergence de petits acteurs, en contraignant les grands oligopoles existants à sortir d'une logique d'accaparation des données pour adopter une logique de partage. Il s'agit néanmoins de deux choix politico-économiques qui peuvent parfaitement cohabiter, assurant tous deux un véritable cadre Open Data aux réutilisateurs.

La « guerre française des licences » devrait donc s'achever. Il paraît probable et raisonnnable que les acteurs ayant opté pour la LIP, les licences de l'APIE ou encore leur propre licence, abandonnent pour certains le fantasme de la « non-altération », et basculent progressivement tous vers cette nouvelle « Licence Ouverte » ou vers l'ODbL, uniformisant ainsi petit-à-petit l'écosystème des données publiques françaises vers un duo de licences libres sain pour le développement de l'Open Data français.

Les formats, le prochain enjeu juridique pour Étalab

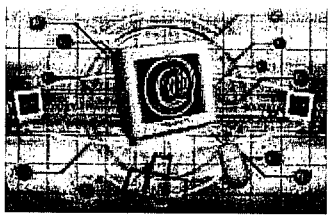
Le premier enjeu juridique semble donc aujourd'hui dépassé. Étalab et les administrations doivent désormais se concentrer sur l'autre enjeu crucial de l'Open Data : les formats. Afin de permettre à un maximum de réutilisateurs de se saisir des données publiques, il est essentiel que les développeurs puissent les manipuler numériquement aussi librement que possible, ce que ne permettent pas les formats propriétaires. En effet, l'utilisation de tels formats fait encourir un important risque juridique aux réutilisateurs : le « propriétaire » des spécifications peut en effet interdire à tout moment l'ouverture de documents enfermés dans « son » format. Ainsi, Microsoft, propriétaire du XLS et du ODATA, a fait subir de très grosses pressions juridiques aux développeurs d'OpenOffice.org pour les empêcher de continuer à créer des outils permettant notamment la lecture de ces standards non-libres.

Pour maximiser les réutilisations et sécuriser les réutilisateurs, il est donc essentiel que les données publiques soient mises à disposition dans les formats ouverts et structurés clairement définis et reconnus : CSV, ODS, XML, JSON, RDF, KML, SQL, ... C'est le sens qu'a donné le Gouvernement à la mission Étalab. Espérons qu'elle saura tenir également cet objectif...

Publié dans Le blog, Open Data, *www.regardscitoyens.org*

la gazette.fr

Données publiques : une nouvelle économie des services publics en débat



Par R. Mazon

Les quelques collectivités et intercommunalités qui se sont lancées dans l'ouverture de leurs données ont choisi des approches pragmatiques. Elles ne libèrent que des données « simples », qui ne soulèvent pas de question juridique ou technique. Mais rejoindre le mouvement de l'open data implique en réalité de faire des choix, notamment quant au statut libre ou payant des données mises à disposition. Et de réfléchir au rôle de la collectivité pour favoriser l'innovation, en acceptant d'abandonner le « monopole » de la production des services.

[...]

C'est une ordonnance du 6 juin 2005, transposition d'une directive européenne du 17 novembre 2003, qui modifie la loi du 17 juillet 1978 en créant un droit de réutilisation des données publiques. Il s'agit de permettre aux citoyens, aux acteurs économiques, aux chercheurs, de se saisir de données publiques pour produire... ce qu'ils jugent utile : un site sur l'activité de chaque député, une cartographie interactive des lieux accessibles aux personnes handicapées, la disponibilité en temps réel des vélos en libre-service à Rennes, ou à Bordeaux, le recensement de la biodiversité à Montpellier, ou les faits de criminalité par quartier à New York. Et tout ce qui n'est pas encore imaginé.

Avec, d'un côté le développement des technologies du web et du web mobile et, de l'autre, la masse considérable de données produites ou collectées par les administrations, le potentiel, économique et social, est jugé énorme. La Commission européenne, qui milite activement pour le développement d'une économie du numérique et une société de la connaissance estime le marché des données publiques à plus de 27 milliards d'euros. La Fondation internet nouvelle génération (FING), acteur important de l'Open data public, met aussi en avant le potentiel d'innovation sociale et démocratique.

Quelques villes ou territoires pionniers se sont engagés dans ce mouvement de l'Open data, mais il reste encore méconnu, et sous-exploité par les collectivités locales. Pour une administration,

ouvrir ses données pour réutilisation peut soulever des questions d'ordre juridique, et bousculer les pratiques ou des principes.

1 – Qu'est-ce qu'une donnée publique réutilisable ?

Juridiquement, le droit de réutilisation découle en partie du droit d'accès aux documents administratifs, régulé par la Cada : les documents ou informations librement accessibles ou publiés par l'administration (hors service public industriel et commercial) sont réutilisables. Cette règle comporte trois exceptions : les documents contenant des données personnelles, les informations détenues par des organismes culturels ou d'enseignements et de recherche, et les données sur lesquelles des tiers détiennent des droits de propriété intellectuelle.

Concrètement, *il faut distinguer deux sortes d'information publiques, précise Charles Népote, directeur du programme Réutilisation des données publiques de la FING : les documents, type PDF, qui sont peu structurés et donc peu réutilisables pour créer de la valeur, et les données qui, elles sont très structurées. Les données, ce sont des mesures, des statistiques, des coordonnées géographiques, des horaires, des longueurs, des quantités....* Leur exploitation par des réutilisateurs permet le croisement, l'agrégation, et la création de nouveaux services ou de nouvelles connaissances.

2 – Identifier les données publiques disponibles

Pour favoriser la réutilisation, la loi « Cada » de 1978 impose aux administrations la tenue d'un répertoire de leurs données publiques, porté à la connaissance des usagers et réutilisateurs. Une obligation que très peu de collectivités locales ont mise en œuvre, à l'inverse des services de l'Etat, poussés par l'action de l'Agence du patrimoine immatériel de l'Etat (Apie).

Les collectivités pionnières font preuve de pragmatisme. *En théorie, nous avons une centaine de logiciels métiers, qui produisent de la donnée, même si tous ne sont pas complètement renseignés. Cela donne une masse considérable de données* sur laquelle s'appuyer, relève Philippe de Tilbourg, collaborateur de cabinet en charge du numérique à la communauté urbaine de Bordeaux.
De fait, les collectivités publient sur leurs sites internet, donc dans un format communicable et réutilisable, quantité de données relatives aux services aux administrés. L'ensemble de ces données peut constituer la première brique d'une logique d'open data.

A Montpellier, Jean-Marie Bourgogne, qui est chargé d'animer le projet Montpellier numérique précise ainsi que *nous prévoyons de publier toutes les données disponibles qui ne posent aucune question relative aux données personnelles, ou à une « sensibilité.* Cet ex de la FING ajoute : *Nous discutons pour cela avec les services qui nous disent ce qui peut être sensible, à leurs yeux, par exemple les données sur les réseaux haute tension.*

La communauté de Brest Océane est confrontée à une autre question, celle de la propriété des données. En mars 2010, elle a adopté – première du genre, avant Rennes – une délibération déclarant l'ouverture de ses données géographiques, issues de son SIG. *Mais, les données du SIG viennent de différentes sources : les services internes, les délégataires... Mais pour ce qui concerne le réseau électrique ou de télécommunications, dont les données nous sont fournies par France Télécom ou EDF, la propriété est plus difficile à déterminer. Aujourd'hui, je n'ouvrirais pas les informations sur le réseau EDF, alors même que ce sont probablement des données publiques*, prévient François Vigouroux, responsable du SIG.

Les données qui ne posent aucun problème de propriété sont légions, et permettent de fournir, déjà, aux réutilisateurs une matière première très riche. Pour les autres, les collectivités vont devoir identifier les droits que des tiers peuvent détenir, ou négocier avec eux leurs conditions de mise à disposition.

Droit de propriété des agents

Un autre cas pourra se poser : celui du droit de propriété intellectuelle des agents publics, reconnu par le code de la propriété intellectuelle. *Le droit d'auteur appartient à l'agent, mais avec un mécanisme de cession légale des droits non commerciaux à son administration*, relativise Thomas Saint-Aubin, chargé d'enseignement en droit de la propriété intellectuelle à Paris I Panthéon-Sorbonne. *Pour les droits commerciaux, il faut que l'administration lève son droit de préférence pour en disposer. En principe, l'administration devrait ensuite prévoir un mécanisme d'intéressement de l'agent aux produits tirés de l'exploitation de l'œuvre, mais le décret n'est pas encore paru.* Pour Anne Josso, secrétaire générale adjointe de la Cada (lire l'interview) *c'est un point qui fait peur, mais à titre personnel, je crois que les agents publics n'ont des droits que dans des cas particuliers, et que cela ne concerne qu'une faible part des informations réutilisables détenues par l'administration.*

3 – Pas de contrôle sur la réutilisation

Dès lors qu'une administration libère des données, ou qu'elle est tenue de faire droit à une demande, elle ne peut exercer aucun contrôle sur ce qu'en feront les réutilisateurs, sous réserve toutefois que ces derniers se conforment aux principes généraux du droit de réutilisation. Les réutilisateurs n'ont pas à demander d'autorisation si les informations sont librement communicables, et ils ont pour seules contraintes légales de ne pas les altérer, ou en dénaturer le sens, et de mentionner leurs sources et leur dernière date de mise à jour.

Cette liberté accordée aux réutilisateurs ne va pas sans troubler, dans les administrations ou chez les élus. *Souvent, les agents publics estiment que les demandes des réutilisateurs ne sont pas légitimes, soit qu'ils demandent des données qui ne les concernent pas, comme des pièces de marchés publics, soit qu'ils créent des services qui, selon les fonctionnaires, relèvent des*

missions de service public, relate Anne Josso. *Il peut exister parfois un sentiment de dépossession dans les administrations, mais en réalité il n'en est rien*, confirme Danielle Bourlange, directrice générale adjointe de l'APIE. *Il n'y a rien d'anormal à ce que des opérateurs économiques créent de la valeur avec des données produites par les administrations. Il n'y a donc pas de dépossession mais au contraire, une valorisation. D'où l'intérêt des initiatives, comme celle de la ville de Rennes, qui montrent le potentiel de création de valeur économique et sociale.*

Perte de pouvoir

Pour les élus, c'est la perte de pouvoir sur les informations qui est en jeu. Le « risque » : que les réutilisateurs fassent un « mauvais usage de ces données », mettant en cause la légitimité de leur politique. L'exemple des Etats-Unis, où les réutilisateurs se sont rués sur les données de criminalité pour produire des cartographies par ville, ou quartier, est souvent cité. Mais d'autres exemples sont aussi avancés, comme la localisation des embouteillages, ou de la pollution, qui mettraient en lumière les failles, réelles ou supposées, d'une politique locale, surtout si ces informations étaient exploitées par l'opposition.

Mais ceux qui se sont frottés à l'ouverture des données relativisent ces risques, au premier rang desquels, Michel Briand, vice-président de Brest Métropole Océane, en charge de l'aménagement numérique du territoire : *J'anime des sites participatifs depuis 2003. En permettant l'écriture à tous, y compris les opposants. Certes, nous avons pris un risque, mais en fait, nous n'avons jamais rencontré de vrais problèmes. Je crois que les peurs sont dans nos têtes. En général, une dynamique d'ouverture entraîne une participation positive. Si on veut faire de la bonne politique, il faut donner de l'information. Et puis, qu'avons-nous à cacher ? Il faut vraiment se poser cette question.*

4 – Tarification, redevance, et retour sur investissement

La loi Cada de 1978 précise que les administrations sont libres de fixer des redevances, qui tiennent compte notamment des coûts de mise à disposition des données. Mais une marge de manœuvre est laissée à l'administration pour la tarification, qui ne doit cependant pas entraver l'accès aux réutilisateurs. Cette redevance doit être décrite dans une licence (lire l'encadré), fournie par l'établissement public aux réutilisateurs. En général, la réutilisation des données pour un usage non commercial sera libre de toute redevance. Il en va tout autrement pour les usages commerciaux.

Le potentiel économique attribué au marché de la réutilisation des données aiguise l'appétit des opérateurs économiques, de certaines administrations, et suscite un vif débat dans la sphère publique.

Le site de la fondation pour la recherche sur les administrations publiques et les politiques publiques (Frap), de tendance libérale, écrit que les données publiques représenteraient « un marché de 820 millions d'euros mais qui pourrait atteindre 1,3 milliard d'euros », ce qui « représente une source supplémentaire de revenus que les ministères désireraient ardemment capter dans un contexte de contrainte budgétaire particulièrement forte. Or désormais, ils en ont l'opportunité et le moyen ».

Les redevances peuvent inclure, outre le coût de mise à disposition et de traitement des informations en vue de leur réutilisation (par exemple l'anonymisation), une rémunération raisonnable des investissements, dont une part au titre de la propriété intellectuelle. Dans certains cas, les redevances peuvent même prévoir un intéressement au chiffre d'affaires... C'est ainsi le cas d'un arrêté du 4 août 2010, qui prévoit un intéressement de 20 % du chiffre d'affaires pour la réutilisation commerciale de la base des données comptables des collectivités du ministère des Finances.

Cette logique est défendue par Danielle Bourlange, et l'APIE : *A partir du moment où il y a une réutilisation commerciale de ces informations publiques, il est envisageable qu'une partie de la valeur créée revienne à l'administration à travers une redevance équitable. Ceci est de nature à enclencher un cercle vertueux, car dans un contexte budgétaire contraint, ces ressources supplémentaires pour les administrations permettent d'améliorer la qualité des informations produites, et la qualité des conditions de mise à disposition.* Mais, ajoute-t-elle, *nous préconisons une approche pragmatique et réaliste en matière de réutilisation. Ce n'est ni une logique du tout gratuit, ni une approche du tout payant.*

Accès libre et gratuit

Les tenants du « priorité au gratuit » mettent en avant plusieurs arguments :

- D'une part, pourquoi les administrations percevraient-elles une sorte de retour sur investissement alors que les données qu'elles produisent le sont au titre de leurs missions de service public, financées par l'impôt, et donc déjà payées ?
- D'autre part, comment assurer une réutilisation équitable entre opérateurs si les redevances empêchent certains d'entre eux de consentir cet investissement ?
- Ensuite, dans le cas des collectivités, instituer une redevance suppose la mise en place d'un dispositif de gestion qui peut s'avérer coûteux, surtout s'il s'agit de faire payer des petites sommes à un grand nombre d'acteurs.

Enfin, plus philosophiquement, les partisans du gratuit parient sur la dynamique créée par le large et libre accès aux données, inspirés par le mouvement de l'open source en matière de logiciels : plus d'innovateurs potentiels, plus de nouveaux services inventés, et un retour économique indirect, mais réel par les activités créées.

Un certain pragmatisme est aussi à l'œuvre dans les collectivités. Ainsi, Michel Briand, pour la métropole de Brest, fait le choix de la fourniture gratuite : *Notre position actuelle, c'est qu'on*

donne les données. On y a intérêt, pour les opérateurs GPS par exemple : plus les cartes et les informations nous concernant sont à jour, mieux c'est. Car notre territoire n'est pas assez attractif pour que ces opérateurs tiennent les informations à jour. Mais je pense aussi que rechercher la rentabilité est une erreur fondamentale. Les Anglais libèrent sans entrave, et se rémunèrent par l'activité créée.
Roch Giraud, responsable internet et nouvelles technologies à la direction de la communication de Marseille Provence 2013, l'association qui prépare l'année européenne de la culture, est sur la même ligne : *Je pense qu'il y a des risques réels à penser l'open data comme un service payant. Cela ne peut qu'être limité. Le gratuit, c'est ce qui crée de l'innovation, donc de l'emploi. Les collectivités locales sont au service des citoyens, leur activité ne consiste pas à gagner de l'argent. Ce qui ne veut pas dire qu'on ne peut pas monétiser, à certains moments. Il faut donc partir de la mission qu'on se donne, et favoriser les services qui autorisent sa réalisation. Si on monétise pour les entreprises privées, on monétise, en fin de parcours, pour les utilisateurs.*

A ce stade du développement de l'open data dans les collectivités, l'impératif est surtout de faire décoller les projets, et donc d'éviter les barrières.

Différents types de licences

Juridiquement, l'adoption d'une licence type n'est obligatoire que si l'administration instaure une redevance. Cependant, les administrations, comme les réutilisateurs, peuvent préférer encadrer la réutilisation même non commerciale, par sécurité juridique – c'est d'ailleurs ce que recommande la Cada. Ou pour en faire un outil de communication. C'est le cas de la Licence « Information publique librement réutilisable » (LIP), élaborée par le bureau de la stratégie éditoriale du ministère de la Justice, dirigée par Thomas Saint-Aubin, chargé d'enseignement en droit de la propriété intellectuelle à Paris I Panthéon-Sorbonne. A ce jour, plus de 900 jeux de données sont déjà diffusés sous cette licence sur le portail www.data-publica.com.
L'Agence du patrimoine immatériel de l'Etat (APIE), propose pour sa part 3 types de licences, que peuvent adopter, et adapter toutes les administrations :

- conditions générales pour la réutilisation gratuite des informations publiques (dont l'APIE recommande la mise en ligne sur leurs sites internet par les organismes publics),
- Licence pour une livraison unique d'informations publiques,
- Licence pour une livraison successive d'informations publiques régulièrement mises à jour.

Les licences Creative Commons

Les licences basées sur les licences Creative Commons et ses dérivées sont aussi fréquemment utilisées. Mais, issues du droit anglo-saxon, elles ne sont pas parfaitement adaptées au droit français.

Danielle Bourlange, directrice générale adjointe de l'APIE précise que *les licences libres de type Creative Commons, par exemple ne sont pas adaptées au cadre général de la réutilisation des données publiques puisque, notamment, elles ne permettent pas d'intégrer les obligations de non-altération et de non-dénaturation du sens des données qui constituent des garanties pour les usagers et les clients finaux des réutilisateurs. En effet, avec les Creative Commons, l'exploitation des données peut être autorisée soit sans modification soit avec toutes les modifications possibles. La loi française sur la réutilisation permet de modifier les données, notamment en les croisant, les mélangeant avec d'autres tout en posant certaines limites liées à la nature particulière des informations publiques.*

Les licences Creative Commons ont été créées sur le fondement du droit d'auteur. La LIP se fonde sur le droit des données publiques et permet par exemple la diffusion des contenus sur lesquels il n'y a pas ou il n'y a plus de droits d'auteur, ajoute Thomas Saint-Aubin.

Les licences et les données culturelles

Dans son rapport *"Partager notre patrimoine"*, Bruno Ory-Lavollée, conseiller-maître à la Cour des comptes invite les institutions à réfléchir à leur « stratégie » en matière de réutilisation de leurs données culturelles : recensement des contenus, recherche des droits, examen des différentes conditions d'usages envisageables en fonction de la nature des données, rédaction de licences types, adaptables au cas par cas etc.

Compte tenu de la mosaïque institutionnelle qui caractérise le monde de la culture en France, de la libre administration des collectivités territoriales, et des spécificités de différentes institutions culturelles et de la variété des types de collections, le risque n'est pas négligeable de voir se développer une multitude de pratiques différentes d'un territoire à l'autre, d'une institution à l'autre, voire d'une collection à l'autre.

Une complexité qui atteindra sans doute son paroxysme lorsqu'un même fonds est dispersé entre plusieurs institutions. Le rapport Ory-Lavollée invite le ministère de la culture à « harmoniser l'action des détenteurs de fonds ». Qu'en sera-t-il au niveau de chaque collectivité ? Interrogé sur la question par la Gazette, Bruno Ory-Lavollée parie sur le bon sens des élus et des professionnels : *A l'échelle d'une ville, par exemple, ils devront veiller à ce qu'il n'y ait pas trop de pratiques hétérogènes, afin d'éviter des incohérences. Mais mieux vaut encore passer par une phase provisoire de désordre, avant une remise en ordre, que de ne rien faire du tout.*

RENNES METROPOLE EN ACCES LIBRE

L'open data : c'est quoi?

(...)

Quels types d'information existent dans les services publics ?

- Des données de description du territoire (cartes, cadastre...)
- Des fonds documentaires (études, réglementation, statistiques...)
- Les données de la décision publique (projets, enquêtes, délibérations, subventions, budgets...)
- Le fonctionnement des réseaux urbains (eau, énergie, transports, logistique, télécoms...)
- La localisation et les horaires d'ouverture des services et des commerces
- L'occupation des ressources et des capacités (voirie, bâtiments, espaces, parkings...)
- Des mesures (environnement, trafic...)
- Des événements (culture, sports...)
- Des informations touristiques, culturelles, des données d'archives
- Les flux urbains (circulation...)
- Des données de surveillance...
- Des données électorales

Ces informations, ces données ont des caractéristiques qui les rendent particulièrement intéressantes : ce sont des informations précises, de qualité et mises à jour. Elles sont la plupart du temps géolocalisées, c'est-à-dire qu'elles ont une localisation précise sur un territoire.

Ces données doivent pouvoir être mises à disposition et rendues facilement accessibles dans des formats exploitables pour permettre une réutilisation, un traitement, une analyse par les citoyens eux-mêmes. Il faut des données brutes c'est-à-dire sans retraitement.

Il s'agit de "l'ouverture des données publiques" ou selon le terme anglo-saxon "Opendata".

Bien évidemment, cette démarche exclut les données personnelles ou nominatives.

La réutilisation des informations publiques, une obligation légale

S'agissant en tout cas des informations détenues par les organismes publics, ou par des entreprises chargées d'exploiter un service public, la législation européenne depuis 2003 et française depuis 2005 est claire : ces informations publiques doivent pouvoir être rendues accessibles et réutilisées, à des fins commerciales ou non, d'une manière non-discriminatoire et

non-exclusive, et à des coûts qui n'excèdent pas leur coût de production. Les exceptions sont peu nombreuses, en dehors bien sûr des informations nominatives.

La réutilisation des informations publiques, une stratégie économique

La motivation affichée par la Commission européenne sur son site consacré à l'information publique est avant tout économique. En effet, selon une étude réalisée en 2006, chaque année dans l'Union Européenne, la réutilisation d'information du secteur public (gratuitement ou contre paiement) génère un chiffre d'affaires estimé à au moins 27 milliards d'euros, peut-on lire sur le site de la commission européenne.

Les informations du secteur public, mieux et davantage utilisées, devraient permettre de créer de nouvelles activités et d'offrir aux habitants de nouveaux services. Le marché des applications mobiles, en partie fondé sur des informations du secteur public, pourrait représenter jusqu'à 15 milliards d'euros à l'horizon 2013. Toutefois, une grande partie de ces informations, en Europe, sont sous-exploitées, ou ne le sont pas du tout.

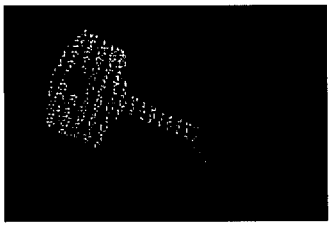
Au-delà de la dimension économique des pays comme la Grande-Bretagne ou le Canada en font également le fondement de leur stratégie de réforme des services publics : c'est grâce à la réutilisation des données publiques que des idées neuves pourront fleurir, et que les organismes publics eux-mêmes pourront proposer des nouveaux services, que de nouvelles formes de proximité pourront émerger.

(...)

la gazette.fr

DOCUMENT 5

Open data : la transparence démocratique demeure virtuelle



Les collectivités engagées dans l'open data disposent de peu de retours permettant d'améliorer la gestion publique.

Par S. Maréchal

Les collectivités pionnières ont enclenché un mouvement irréversible, mais les effets de l'open data sur la transparence de l'action publique territoriale se font attendre. La culture de la donnée doit sortir du microcosme technologique pour intéresser les citoyens.

En septembre 2011, Arnaud Montebourg, président (PS) du conseil général de Saône-et-Loire, médiatisait la mise en ligne du plus gros volume de données budgétaires jamais réalisée par une collectivité. Se doutait-il que, sept mois plus tard, c'est la consultation de cartes postales numérisées qui serait plébiscitée par les internautes ? Et que les applications lauréates du concours organisé sur l'ensemble de ces données seraient liées au tourisme ? « La dimension valorisation du territoire est aussi un enjeu de notre démarche. Pour la partie transparence de l'action publique, nous avons créé les conditions préalables avec un site web dédié afin que les citoyens sachent où chercher, mais nous n'en avons pas encore perçu tous les effets », reconnaît Pierre Martinerie, conseiller général chargé de l'économie et du développement durable.

Faibles retours - Le constat est identique dans la quinzaine de collectivités qui ont rejoint le mouvement de l'open data français, lancé il y a deux ans par Rennes métropole : malgré les fortes attentes, les résultats sont maigres en termes de propositions d'amélioration de la gestion publique. « Ces premières collectivités ont libéré les données dans l'urgence, en mode 'guérilla', avec une forte volonté de communication et d'exemplarité, affirme Philippe Ourliac, délégué général de l'Observatoire des territoires numériques (Oten). Il faut laisser retomber le soufflet et que chaque collectivité définisse une politique de réutilisation selon ses priorités. Car, pour parvenir à une véritable transparence de l'action publique, on doit entrer dans une phase d'ouverture massive des données dans toutes les collectivités et les administrations. »

Une culture à diffuser - Le mouvement de l'open data territorial est donc avant tout constitué de défricheurs qui apprennent sur le tas. Les données mises en ligne sont disparates, sans véritable cohérence, et dans des formats qui ne répondent pas toujours aux exigences techniques des militants. « La dynamique est née dans les directions des technologies de

l'information et de la communication, et des systèmes d'information géographique, mais la culture de la donnée est encore peu développée dans les autres directions. Les acteurs locaux de la démocratie participative ne savent pas eux-mêmes quel usage ils peuvent faire de ces données », remarque Amandine Brugière, chef de projet d'un programme de réutilisation des données publiques à la Fondation internet nouvelle génération (Fing).

Elargir le public concerné - Pour sortir le sujet de son microcosme technologique, la Fing sensibilise désormais des associations non « geeks » qui réfléchissent aux innovations démocratiques. Elle co-organise des évènements, telle la Semaine européenne de l'open data qui se tiendra du 21 au 26 mai 2012 à Nantes avec des associations ou partenaires engagés localement dans l'ouverture des données publiques, comme l'association LiberTic. Ce sera l'occasion de partager les premières expériences d'appropriation par des associations, comme Les Petits Débrouillards, qui diffusent la culture scientifique aux enfants et adolescents, ou Décider ensemble, qui promeut la culture de concertation avec les citoyens. De son côté, le département de Saône-et-Loire réfléchit aux passerelles à mettre en place pour relier son « entrepôt » de données à sa plateforme participative Sirius 71, qui recueille les idées des habitants pour l'évolution du département.

Pourtant, les demandes spontanées des premiers réutilisateurs sont loin d'être satisfaites. « A Paris, par exemple, nous essuyons le refus de la mairie de mettre en ligne les données sur la présence des élus au conseil municipal. Il n'est pas simple, non plus, d'obtenir la localisation des bureaux de vote pour réaliser la carte des résultats des élections quartier par quartier », regrette Tangui Morlier, cofondateur de Regards citoyens, association reconnue pour ses applications de suivi de l'activité des élus au Sénat et à l'Assemblée nationale.

Fournir des indicateurs - Cependant, la transparence ne passe pas nécessairement par la mise en ligne de données brutes au format électronique. « Entre la collectivité qui ne partage rien et celle qui propose des informations sur document PDF, nous préférons la seconde, faute de mieux. C'est une première étape vers la transparence », explique Nicolas Patte, journaliste de données à Owni.fr, site d'information friand de data journalisme. La mise à disposition de jeux de données le plus exhaustifs possible sera donc déterminante pour tisser, enfin, le lien entre démocratie, politique et technologie.

« Au-delà de l'open data, nous nous préparons au big data, c'est-à-dire l'analyse de données détenues à la fois par le secteur public et ses partenaires privés », confie Laurent-Pierre Gilliard, directeur général adjoint d'Aquitaine Europe Communication, l'agence numérique de la région. Les données relatives à l'énergie, l'eau ou la collecte des ordures ménagères seront croisées avec celles des collectivités pour produire des cartographies quartier par quartier, type d'habitat par type d'habitat. Avec des données proches de son quotidien, le citoyen disposera des indicateurs pour juger de la pertinence des politiques engagées par les élus et, à l'heure des réseaux sociaux, du moyen de partager ses sources. Tout comme il a disposé quasiment en temps réel des vérifications des chiffres énoncés par les candidats à la présidentielle lors du débat télévisé qui les a opposés le 2 mai. L'open data local est encore loin du compte.

Simon Chignard : « Il existe un vrai décalage entre les promesses politiques et les attentes des citoyens »

Simon Chignard est consultant indépendant, président de l'association Bug, vice-président de la Cantine numérique rennaise et auteur du livre « Open data. Comprendre l'ouverture des données publiques », Fyp éditions, avril 2012.

Quels premiers constats tirez-vous du mouvement de l'open data lancé par les collectivités locales ?

La dynamique initiée par Rennes métropole a fait réfléchir des villes qui avaient, elles aussi, des demandes de réutilisation de la part de leurs habitants. Le temps est encore à la pédagogie pour trouver la bonne manière d'organiser cette mise en ligne de jeux de données. Ce qui frappe le plus ceux qui se sont lancés, c'est que le chemin parcouru est aussi important que le résultat. Le retour sur investissement de la mise en ligne de données géolocalisées est démontré pour la création de services liés à la vie quotidienne des citoyens et l'implication des usagers est un facteur clé de succès.

Mais les promesses de l'open data ne sont-elles pas la transparence de l'action publique et la réduction de la dépense ?

Il existe un vrai décalage entre ces promesses politiques et les attentes des citoyens. Les militants de l'open data se livrent eux-mêmes à un combat de valeurs, qui peut donner des résultats antinomiques. Ils n'ont ni la même définition de la transparence, ni les mêmes attentes. Certains voient dans l'ouverture des données un moyen, pour la société civile, d'évaluer la performance du secteur public et, par conséquent, de revoir les frontières entre services opérés par le public et ceux à confier à des acteurs économiques. D'autres estiment qu'elle permet au citoyen d'exercer son droit de regard et, donc, de contrôle sur la conduite des acteurs publics. D'autres, enfin, la conçoivent comme la possibilité de transformer une contrainte légale en une opportunité sociale et économique. Si la transparence se limite aux données, sans être élargie au dialogue avec les autres acteurs, nous aurons raté quelque chose.

Faut-il que les collectivités territoriales lancent des concours de réutilisation des données financières ?

Organiser un concours c'est bien, mais il doit l'être en fonction des objectifs publics. Si l'on cherche la transparence de l'action publique sous l'angle des données budgétaires, l'open data n'est pas forcément la meilleure manière d'y parvenir. La Grande-Bretagne, dont la priorité, justement, était la diminution de la dépense publique, découvre que l'ouverture des données n'est pas nécessairement le bon outil. Lorsque l'on libère des données liées aux transports ou à l'emplacement des arrêts de bus, on est rapidement dans le champ du participatif. Lorsqu'il s'agit de données budgétaires, on est bien dans le champ de la transparence, mais pas dans celui du participatif. Ces informations sont déjà disponibles sous d'autres formes, les livrer en données brutes n'a pas de réelle valeur ajoutée. Les taux de téléchargement sont peu significatifs et peu d'applications sont développées.

Ouvrir ses données : un choix de droite ou de gauche ?

Les premières collectivités ayant créé leur propre plateforme open data sont dirigées par une majorité de gauche. Les rares collectivités gérées par une majorité de droite qui ont décidé d'ouvrir leurs données ont préféré rejoindre data.gouv.fr, mis en place par l'Etat. Ce clivage politique est cependant à relativiser, sachant que huit des onze plus grandes villes de France sont dirigées par une majorité de gauche. Et, parmi celles qui ne se sont pas encore lancées, il reste autant de villes « de gauche » que « de droite ». Cependant, les collectivités et intercos les plus avancées dans leur démarche (Rennes métropole, communauté urbaine de Bordeaux, Nantes métropole, régions Aquitaine et Pays de la Loire) forment un « arc atlantique ». Une tendance que le consultant Simon Chignard (lire l'entretien ci-dessus) expliquerait par la forte culture de la relation avec le tissu associatif qui existe dans cette zone géographique.

JDN
JOURNAL DU NET

Qu'est-ce que le Big Data ?

Pour répondre aux nouveaux enjeux de traitement de très hautes volumétries de données, les entreprises peuvent faire appel à des solutions spécialisées dans le Big Data. JDN Solutions fait le point.

En savoir plus

- **Chronique :** Les innovations technologiques Big Data (Brian Gentile - JasperSoft)
- **Dossier :** Business Intelligence

Que recouvre la notion de Big Data ?

Le Big Data fait référence à l'explosion du volume des données dans l'entreprise et des nouveaux moyens technologiques proposés par les éditeurs, en particulier de la Business Intelligence, pour y répondre. Le terme même de Big Data a été évoqué la première fois par le cabinet d'études Gartner en 2008 mais des traces de la genèse de ce terme remontent à 2001 et ont été évoquées par le cabinet Meta Group racheté en 2005 par...le Gartner.

A quels enjeux répondent les solutions de Big Data ?

Les fournisseurs de solutions ont commencé à élaborer leurs offres à la fin des années 2000 et le mouvement s'est accéléré ces derniers temps avec l'arrivée sur ce créneau d'acteurs historiques comme Oracle, IBM et tout récemment Microsoft. En juin 2011, Informatica, acteur dans le domaine de l'intégration de données, a également annoncé son offre.

Les objectifs de ces solutions sont de traiter un volume très important de données aussi bien structurées que non structurées, se trouvant sur des terminaux variés (PC, smartphones, tablettes, objets communicants...), produites ou non en temps réel depuis n'importe quelle zone géographique dans le monde.

L'un des principaux points forts de la base de données NoSQL est sa performance

Sur quelles technologies reposent les solutions Big Data ?

Pour analyser les hautes et disparates volumétries de données, les solutions peuvent s'appuyer sur trois types de technologies distinctes : une accélération matérielle à l'aide de mémoires dynamiques DRAM ou Flash, le recours à des bases de données massivement parallèles (Massively Parallel Processing) ou encore les solutions utilisant des formats de bases de données non relationnelles basées sur NoSQL.

Quel est l'intérêt de proposer du Big Data basé sur NoSQL ?

L'un des principaux points forts de la base de donnée NoSQL est sa performance. Ce n'est d'ailleurs pas pour rien que des géants du Web comme Facebook, Twitter et ou encore LinkedIn ont choisi de migrer une partie de leurs données dessus. Les atouts de NoSQL sont au moins au nombre de trois : cohérence (visibilité par tous les nœuds d'un système des données identiques à un instant T), haute disponibilité des données même en cas de panne, et possibilité de partitionner tout système distribué.

Quels fournisseurs s'appuient sur NoSQL et Hadoop pour leurs solutions Big Data ?

Les fournisseurs proposant des solutions basées sur ce type de bases de données sont de plus en plus nombreux. Ils y associent également le plus souvent le framework Open Source d'analyse de données en masse Hadoop, comme c'est le cas pour Microsoft et IBM. Mais également Oracle qui de son côté propose une appliance (boîtier) dédiée au Big Data reposant à la fois sur Hadoop mais aussi une version personnalisée de NoSQL.

A lire ailleurs

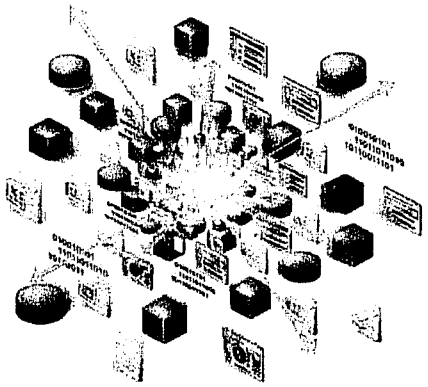
- **Article :** Big Data : 5 choses à savoir avant de se lancer (en anglais)
- **Encyclopédie :** Wikipedia (en anglais)

Le Big Data fait-elle partie des préoccupations centrales des DSI ?

Pour le cabinet d'études Gartner, aucun doute là-dessus. Elle arrive même devant le Cloud Computing et l'analyse en mémoire dans le palmarès des 10 tendances technologiques stratégiques pour l'année 2012. Mais le Big Data se fait tout de même voler la vedette par d'autres problématiques clés dont l'intégration dans le système d'information des tablettes ou encore l'Internet des objets.

LeMondeInformatique.fr
Toute l'info et les tendances du monde IT

Dossier : Le décisionnel à l'heure du big data



(...)

Big data, un nouveau défi pour les entreprises

Ce n'est plus une surprise, le concept de big data devrait avoir la part belle des investissements en 2012. Et pour cause, ce phénomène est la réponse à la multiplication des données à traiter, surtout celles qui ne sont pas ou peu structurées en provenance des réseaux sociaux, des sites de vidéos en ligne, des emails, etc. « La consumérisation IT a fortement contribué à l'explosion des données dans les entreprises et donc du big data » indique à juste titre Fabrice Endlicher, directeur des ventes Europe du Sud et Benelux du distributeur à valeur ajoutée Stordis.

En 2011, 1,8 zétabytes de données se sont échangés selon le cabinet d'études américain IDC, ce qui équivaut à remplir 57,5 milliards d'iPad d'une capacité de 32 Go. L'augmentation effrénée des volumes de données est aussi à mettre en parallèle avec l'explosion de la data mobile (usage intensif des smartphones), des usages convergents et multi-terminaux en entreprise. Bref, selon IDC, d'ici à 2020, l'accroissement du volume d'information à gérer sera 50 fois supérieur.

Fort de ce constat, on comprend mieux le phénomène du big data. Mais, cette croissance exponentielle des données interpelle sur la gestion de leur cycle de vie, leur qualité, leur sécurité et surtout leur traitement. Selon une étude menée par l'éditeur LogLogic et le cabinet de conseil Echelon One en janvier-février auprès d'un panel international de 207 entreprises, 67% de ces dernières avouent que la gestion de toutes ces données représente un enjeu important voire sensible. Sensible car les données semi-structurées et non structurées sont problématiques à exploiter contrairement aux informations transactionnelles du système d'information.

Les entreprises sont et seront confrontées aux limites des systèmes existants de base de données relationnelles qui ne sont plus à même de traiter et d'être ainsi rentables pour faire face à ces volumes énormes de données non structurées. « Il y a 10 ans, les données étaient à 70% structurées pour 30% d'informations non structurées. Demain ça serait l'inverse. Cela signifie que le datawarehouse risque de s'isoler de plus des 2/3 des données. Ce n'est donc plus acceptable » s'alarme Gilles André, directeur général de Polyspot. De plus, le poids de la donnée non

structurée est extrêmement important ce qui dégrade les performances. Par exemple, l'exécution de certaines ressources et requêtes (plusieurs milliards de lignes) critiques pour l'activité de l'entreprise peut réclamer plusieurs heures à partir des systèmes traditionnels alors qu'il faut au mieux quelques secondes sur des plateformes nouvelles générations très évolutives de type « big data ».

Parallèlement à la vétusté des systèmes déjà existants en interne, les entreprises font aussi face à un déficit en compétences nécessaires pour exploiter les possibilités qu'offre le croisement des big data avec l'analyse de données. Il faut dire que les opérations à réaliser (chargement de données, extraction, transformation, traitement, etc.) réclament une certaine expertise dans ce domaine. D'ailleurs, selon les résultats d'une enquête effectuée par le groupe EMC auprès d'environ 500 data scientists, un spécialiste du big data, au niveau mondial, la majeure partie des entreprises aujourd'hui souffrent de cette carence en compétences. Ainsi, 32% des personnes interrogées estiment qu'elles manquent de compétences et de formation, 32% d'entre elles confient un manque de budget ou de ressources, 14% des entreprises avouent qu'elles n'ont pas une structure organisationnelle adaptée et 10% des personnes relèvent une pénurie d'outils et de technologies.

Un nouvel eldorado pour les éditeurs

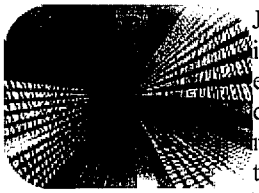
Globalement, le big bata reprend les mêmes problématiques que le décisionnel (Business Intelligence) même si, théoriquement, la BI se concentre davantage à l'exploitation de données à des fins de monitoring et d'aide à la décision, alors que le big data, s'ouvrant en plus à des informations présentes à l'extérieur de l'entreprise, est plus focalisé sur l'utilisation d'outils d'analyse et d'algorithmes afin de générer des données prospectives devant mener à l'innovation. Cela dit, big data ou Business Intelligence, tous deux ont pour objectif des analyses de données pertinentes et en temps réel de préférence.

Ainsi, tous les acteurs spécialisés dans l'analytique, l'intégration, l'extraction et l'analyse de données s'intéressent à ce nouvel eldorado pour pousser leurs offres. On peut citer Japersoft, Pentaho ou Quantum 4D pour leur solution décisionnelle mais aussi Terradata avec Aster Data Analytic Platform, EMC avec Greenplum Data Appliance Computing, Oracle avec Exadata, HP avec Vertica Analytics Platform, IBM avec Netezza et BigInsights, SAS avec Business Analytics, SAP avec Sybase IQ VLDP, Talend Open Studio for big data, etc.

Du côté des entreprises, elles se montrent attentives. « Nos clients confirment leur intérêt pour le big data, avec des Business Cases confirmés par les pilotes en cours dans les secteurs du retail, de la logistique et autour des usages du social média qui va constituer une vraie rupture technologique. Néanmoins, les technologies sont toutes encore assez récentes, et il est encore tôt pour confirmer qu'elles seront les solutions majeures dans 3-5 ans » reconnaît Eric Blum, le CTO de BMC Software pour l'Europe et le Moyen Orient.

[...]

Vous avez dit Big Data ?



Jean-François Forgeron - Véritable phénomène, « pétrole du XXIème siècle » (1), c'est paraît-il LE défi informatique de la décennie. Il s'agit du « big data » (2), expression qui recouvre ces ensembles de données que nous produisons, ou que nos équipements produisent (c'est « l'internet des objets ») en temps réel, qui sont d'origines diverses et souvent non prédictibles. Cette masse a nécessairement conduit à inventer de nouveaux dispositifs de gestion. Les technologies requises, telles que les bases NoSQL ou les composants logiciels comme Hadoop (utilisé par Facebook) ou MapReduce (d'origine Google) sont déjà là. Qu'il s'agisse d'améliorer la santé publique en faisant progresser les études épidémiologiques, de lutter contre la cybercriminalité (3) ou encore de mieux connaître les habitudes de consommation de la clientèle, aucune organisation ne peut raisonnablement longtemps résister. Les usages, souvent issus du marketing intelligent et des avancées de la business intelligence, vont bientôt fleurir...

Au cœur de ces réflexions, les questions juridiques semblent pour une large part à peine posées (4). Elles sont pourtant au centre des enjeux, puisqu'on peut difficilement imaginer le développement du mécanisme du big data sans monétarisation, et celle-ci sans encadrement juridique. Bref, le commerce appelle le droit.

Dans le contexte d'une technologie émergente, il pourra sembler logique de s'interroger dans l'ordre suivant : tout d'abord sur l'appropriation des données, puis sur les limites posées par la nature des données et la licéité des traitements envisagés, et enfin sur la contractualisation des rapports entre fournisseurs et clients. Voici quelques pistes.

A qui sont les données ?

Il est admis que « la notion de propriété des données n'a pas de statut juridique en tant que tel » (5). La question de la propriété des données est donc juridiquement complexe. Schématiquement, elle peut être divisée en deux parties : la protection des bases de données et la titularité des droits sur les données. La première est classique et tranchée, en Europe, par les dispositions de la directive du 11 mars 1996 sur les bases de données (transposée en France par la loi n° 98-536 du 1 juillet 1998 concernant la protection juridique des bases de données) qui ajoute à la protection par le droit d'auteur un droit "sui generis" du producteur de la base contre une extraction ou une réutilisation de celle-ci ou d'une "partie substantielle du contenu de celle-ci". La seconde suppose à nouveau une classification, entre les données privées et les données publiques, qui relève de l'open data. Les données privées, soit parce qu'elles sont le produit d'une entreprise ou qu'elles relèvent de la sphère privée des individus, ne peuvent faire l'objet, librement, d'une appropriation par un tiers. Les dispositifs juridiques de lutte contre la concurrence déloyale ou le parasitisme économique et la violation du secret de fabrique (article L.621-1 du code de la propriété intellectuelle), ou la contrefaçon de droits de propriété intellectuelle (y compris pour les creative commons), ou encore de protection de la vie privée montent la garde. Et quand on imagine que les réseaux sociaux sont des pourvoyeurs de données privilégiés, on fera attention à ce qu'on écrit sur son mur, si on ne veut pas que les informations qu'il contient soient exploitées par d'autres ! (6) Quant aux données publiques (7), leur usage est régi par les dispositions de la licence ouverte (8) qui autorise expressément l'exploitation commerciale des données, y compris en combinaison avec d'autres données et par inclusion dans un produit ou une application. Les bases de données géantes du Big Data sont donc largement appropriables à condition de vérifier l'origine des données concernées.

Quelles limites pour les traiter ?

Si le terrain de jeu du Big Data est loin d'être restreint, il n'est pas sans limites. Elles tiennent, en premier lieu, à la nature des données et aux traitements envisagés, et quand il s'agit de données personnelles, la vigilance est nécessaire. D'autant qu'il s'agit de la catégorie de données à laquelle on ne peut s'empêcher de penser immédiatement en matière de Big Data. En Europe, le traitement de données à caractère personnel est régi par les dispositions de la Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (transposée en tous les états membres). Parmi les principes qui gouvernent la réglementation, ce sont peut-être ceux relatifs aux finalités, à la proportionnalité et à la pertinence des traitements qui risquent le plus de souffrir, si l'on y prend pas garde. Dès lors qu'il s'agit de mener toutes sortes d'analyses et de recoupements en temps réel, difficile en effet d'assurer que tous les

traitements seront réalisés avec la transparence requise. Et si l'on ajoute que le futur règlement européen s'accompagne de nouvelles mesures de protection des droits des personnes, c'est donc une véritable stratégie juridique de mise en œuvre du Big Data que les entreprises devront adopter pour conjuguer avec justesse les enjeux économiques et les libertés individuelles.

Mais les limites ne s'arrêtent pas là. Aux côtés des données personnelles, il existe divers ensembles de données soumises à des règlementations sectorielles qui en imposent la confidentialité, qu'il s'agisse de données relatives aux secteurs de la banque et des assurances, d'informations commercialement sensibles (ICS), du secret de la défense nationale... et qui peuvent rendre les recoupements particulièrement malaisés. Là encore, la connaissance du cadre juridique est une vertu en relation directe avec la gestion du risque.

Quels contrats ?

Framework et outils d'analyse (notamment de text mining) sont sur le marché et les fournisseurs de solution également. Le mode de contractualisation, de type client/fournisseur est somme toute assez classique, peut être avec un renouveau des clauses de participation aux résultats, comme il en existe dans les contrats relatifs aux solutions de yield management construits autour d'un ensemble qui associe les outils et la prestation autour d'un objectif spécifique d'optimisation financière, propre à chaque client, anticipé et chiffré. Mais nul doute que les pratiques contractuelles issues des services Cloud s'invitent aussi à la partie, pour supporter les offres de services qui allient à l'hébergement dans le nuage la mise à disposition d'outils d'analyse. Là encore, ce qui compte c'est de ne pas être dans le brouillard (9) et de disposer d'engagements clairs au moyen d'un contrat qui soit un véritable outil d'anticipation des risques.

Finalement, alors que le Big Data est souvent résumé autour de la loi des « 4V » (10) qui caractérise les données à traiter (pour Volume, Vélocité, Variété et Valeur), et si on osait en suggérer un cinquième pour « Validité » ?

(1) Voir Jacques Secondi, "Big data : Le pétrole du XXIème siècle", Le Nouvel économiste, publié le 22-2-2012.

(2) Voir Wikidépia, 2-5-2012.

(3) Voir Valery Marchive, "La sécurité se met résolument au «Big Data»", LeMagIT, publié le 12-3-2012.

(4) Voir Didier Gazagne, "Technologies Big Data : quelle data-strategy pour l'entreprise ?", le 28-2-2012.

(5) Voir Alain Bensoussan, "La propriété des données", Blog expert Le Figaro, le 18-5-2010.

(6) Voir Mathieu Prud'homme, "Caractère public de propos sur Facebook : attention à la preuve du paramétrage", le 27-4-2012.

(7) Disponibles sur http://www.data.gouv.fr/

(8) Disponible sur http://ddata.over-blog.com/

(9) Voir Alexandre Fiévée, "Cloud computing : l'informatique dans le nuage, pas dans le brouillard", le 12-4-2012.

(10) Voir Laurence Dubrovin, "Big Data - L'offre est là, les besoins doivent émerger", L'Œil Expert, publié le 3-4-2012.

ZDNet.fr

Big Data : décisionnel de M. Jourdain ou réelle innovation ?

Sommaire : Le "Big Data" est le dernier concept à la mode qui promet l'analyse d'énormes volumes de données de toutes natures, structurées ou non structurées, et véhicule la promesse d'une terre promise...

Par Frédéric Charles pour Green SI |

Le "Big Data" est le dernier concept à la mode qui promet **l'analyse d'énormes volumes de données de toutes natures, structurées ou non structurées**, et véhicule la promesse d'une terre promise et de nouvelles opportunités pour les entreprises. Comme souvent dans ce type de situations, de nombreux vendeurs tapent à la porte de la DSI pour expliquer que les architectures et les technologies à mettre en œuvre demandent une révolution du SI qu'il faut engager au plus tôt... et de préférence avant la date de calcul des commissions des commerciaux.

Alors **mode passagère encore pleine de paillettes marketing, ou réel sujet à traiter** pour préparer ses investissements 2013 ?

Pour GreenSI il est important de revenir aux fondamentaux : les "DATA"!

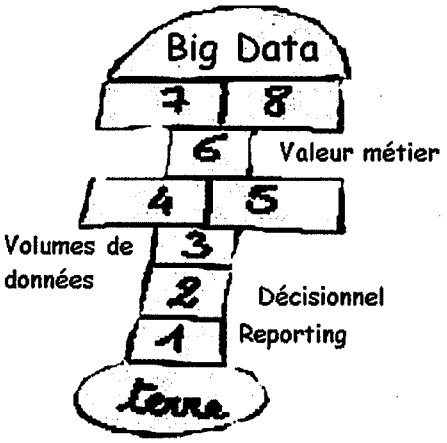
Elles "appartiennent" aux métiers et transitent dans les tuyaux et applications de la DSI. **Donc si pression "big data" il y a, elle commence par la production (ou la capture) par les métiers de ces quantités énormes de données.** La DSI vérifiera ensuite si ses tuyaux sont adaptés. Donc première vérification à faire avec nos directions fonctionnelles, l'augmentation ou pas des volumes, en anticipant les 3 prochaines années et en cherchant surtout les données nouvelles qui seraient pertinentes. Mais attention, car certes il y a chaque année des volumes croissants de données à traiter, mais pour arriver au point de rupture qui ferait que nos architectures décisionnelles ne sont plus adaptées, **il faudrait que ces volumes augmentent exponentiellement, et pas linéairement.**

Dans cette catégorie on rencontre par exemple :

- **Les logs des sites web** et des navigations entre les pages, actuellement rarement stockées voire analysées,
- **Les données des médias sociaux** et toutes les interactions des clients sur le web,

- **Les nouvelles données capturées avec capteurs** embarqués et mobiles ou répartis sur un territoire : lecture de tags RFID, localisations GPS, mesures de capteurs des bâtiments comme l'énergie, ...

Mais quand on creuse, il n'y en pas tant que ça. Donc en 2013 cela risque de ne pas être le cas :



- **pour toutes les entreprises :** certainement pour celles où la R&D et le marketing sont essentiels pour l'avenir de l'entreprise comme pour les laboratoires pharmaceutiques ou la grande distribution, mais plus difficile à croire dans des entreprises qui n'auraient pas aujourd'hui déjà des entrepôts de données conséquents,
- **pour toutes les maturités d'entreprises par rapport à l'importance des données** : Google et Facebook ont clairement compris que leur modèle et leur avenir passent par l'exploitation en masse des données, mais est-ce le cas des DG des entreprises qui ne sont pas encore passées par la case "analyse décisionnelle" et en sont encore à la case "simple reporting" ?
- **pour tous les domaines fonctionnels** : peut-être la partie client, marketing, avec le développement des données sur la relation client sur internet au sens large comme vu précédemment, mais plus difficile à croire dans le domaine de la finance par exemple même si on nous explique que la détection de la fraude demande à se développer avec de nouvelles données,
- **pour celles qui ne peuvent en tirer un avantage concurrentiel** : ce qui paraît une évidence, mais bonne à rappeler. Car le big data n'est pas dans le domaine du réglementaire qui s'impose à l'entreprise. Si l'entreprise se met à produire ou "aspirer" des déluges de données ce n'est pas juste pour le plaisir de saturer ses serveurs ou de remplir son Cloud... **La valeur métier est donc essentielle à apprécier et indispensable pour dimensionner le niveau de ressources à y consacrer.**

Toutes les entreprises ne sont donc pas égales vis-à-vis du Big Data. Et pour la plupart des DSI, GreenSI pense que 2013 sera le moment des études et de la réflexion, peut-être de la R&D et du prototypage voire de la stratégie, mais certainement pas des révolutions architecturales. Sur le moyen terme, le rythme de développement de l'entreprise numérique déterminera si il concernera ou non, toutes les entreprises.

Et si vraiment vous voulez caser le mot Big Data dans une conversation avec votre DG, faite comme M. Jourdain qui faisait de la prose sans le savoir. Car **fondamentalement le Big Data, c'est une architecture de données, des méthodes et du décisionnel, ce que l'on fait déjà depuis 20 ans... mais avec moins de données.** Le changement actuel ne semble lié qu'à la scalabilité de nos investissements technologiques passés qui trouvent une limite dans le monde de l'internet. Une limite que de nouvelles technologies peuvent aider à dépasser quand on aura atteint **le seuil de leur valorisation métier.**

Et pour ceux qui se rapprochent de ce seuil ou qui ont déjà atteint la case Big Data, s'ouvre devant eux l'adoption de nouvelles technologies :

- pour le **stockage** et le traitement de la volumétrie "big" et exponentielle,
- pour le **requetage** qui demande une nouvelle structuration et souvent une analyse proche du temps réel,
- pour le **traitement** qui amène une complexité d'analyses nouvelles, prédictif par exemple ou même tout simplement la **visualisation** des données en grand nombre et la manipulation de ces visualisations,

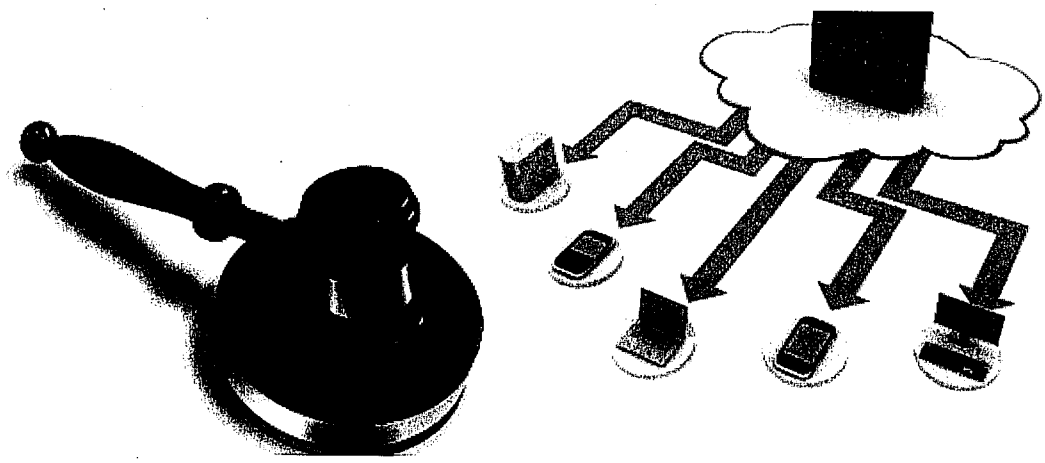
Les ruptures techniques apparaissent donc entre le relationnel et le "**No SQL**". Le framework **Hadoop** (open source) commence à s'imposer pour permettre du traitement en parallèle, qui avec le travail en direct sur les données de production ou des méthodes d'acquisition très performantes, sont les clefs des nouvelles architectures.

Mais rassurez vous, GreenSI pense que pour la majorité d'entre nous, on a le temps de s'adapter si on s'y prépare et de jouer à M. Jourdain en attendant. Et pour le commercial qui vous attend dans le couloir derrière la porte? Oh, il finira bien par partir... ;-)

Bienvenue dans le Blog de Référence de la DSI

Reférence DSI

Bienvenue sur Référence DSI, le blog des Directeurs Informatiques



Jusqu'où s'étend la responsabilité du DSI ?

Le DSI a en charge la mise en place de l'accès à toute l'information nécessaire à l'entreprise. Peut-on considérer que le Système d'Information (SI) de l'entreprise englobe désormais le WEB ? Tout le WEB ? Le Cloud ? Quel est le rôle du DSI ?

Il n'est plus réaliste de limiter le Système d'Information aux seules frontières de l'entreprise. Aujourd'hui, si les entreprises deviennent peu à peu étendues, le Système d'Information est depuis longtemps à dimension variable. Quelle est la nature du SI ? Quel rôle le Directeur des Systèmes d'Information (DSI) doit tenir dans ce contexte changeant ?

Les limites de l'infrastructure :

On définit souvent l'infrastructure comme le support nécessaire à la gestion du contenu. Il n'y a donc pas de limite géographique à l'infrastructure, le Système d'Information s'appuie sur des éléments d'infrastructure internes et des éléments externes. Que les serveurs soient locaux, hébergés ou virtualisés dans le CLOUD, ils constituent l'infrastructure composante « physique » du Système d'Information. Le rôle du DSI est de veiller à ce que tous les éléments de son infrastructure soient sous contrôle, disposent de redondance, qu'un SLA soit contractualisé et que l'ensemble des processus de pilotage de l'infrastructure soient réellement actifs. C'est évidemment beaucoup de contractualisation avec des tiers.

Les frontières de l'information :

La classification des informations en informations « structurées » et « non structurées » ne permet pas de fixer la frontière du SI sur le plan des informations. En effet, il est nécessaire de distinguer les informations nécessaires à la gestion, à la capitalisation en connaissance et celles qui ne le sont pas. La notion de propriété de l'information ne permet pas davantage de limiter le champ du SI. Avoir le pouvoir « d'effacer » une information assure sans doute que celle-ci est une composante du SI de l'entreprise, mais la réciproque n'est pas vraie. De nombreuses informations utilisées désormais dans le SI proviennent du WEB et sont délivrées par des WebServices aux différents logiciels de gestion. Le rôle du DSI, pour la partie information du SI, sera de garantir la disponibilité des sources d'information externes et de capitaliser les données internes afin de parvenir à constituer le SI des informations.

Le rôle juridique du DSI dans la gouvernance :

On voit désormais le rôle prépondérant des éléments externes (infrastructures et données) dans le SI. Parmi les nombreuses expertises attendues chez un DSI, celle de la contractualisation concernant la disponibilité de toutes les ressources du SI n'est pas la moindre. D'autant que le monde de l'OPEN DATA procure de nombreuses informations mises à disposition gratuitement. La contractualisation avec les sources n'en est que plus complexe. Même si ces sources n'ont aucune raison de ne plus mettre à disposition leurs données, le DSI est celui qui doit veiller à disposer de secondes sources et qui doit vérifier la véracité des informations fournies afin de ne pas fragiliser le SI de l'entreprise.

La contractualisation de services

Face à ces nouveaux enjeux imposant un SI étendu à une entreprise étendue, le DSI doit pouvoir s'appuyer sur des experts et sur ses pairs. En effet, la contractualisation d'infrastructure ou d'information devient aujourd'hui d'abord et surtout la contractualisation de services. La mise en place de ces nouvelles relations, qui laissent de côté les approches « propriétaires », conduit à des changements importants dans la gouvernance du SI, dans le rôle d'expertise du DSI et dans la répartition des responsabilités entre l'entreprise et ses fournisseurs.

Sans aucun doute, ces évolutions récentes donnent un poids supplémentaire au DSI dans la gouvernance de l'entreprise.