

En bref

Valoriser la donnée n'est plus accessoire. De plus en plus souvent, l'entreprise est tenue de se réinventer autour de la donnée pour servir ses clients.

Si l'exemple le plus complet vient de façon inattendue d'une agence média, Havas Media Group, d'autres témoignages ont appuyé ce constat. En effet, le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise ».

En voici un compte-rendu.

Sommaire

Juridique

Data Asset : la valeur de la donnée associée à sa protection légale

Business

Data Asset : la délicate valorisation des rapprochements de données culturelles

Projets

Valoriser ses données en assurant leur réutilisation

Projets

Data Asset : le succès grâce aux données

Stratégie

Raphaël De Andreis (Havas Media Group) : « Je dois agréger et rendre utiles toutes sortes de données »

Data Asset : la valeur de la donnée associée à sa protection légale



Le 24 septembre 2015, CIO a organisé une conférence sur « Data Asset : la donnée, actif de l'entreprise » avec l'intervention de Christiane Féral-Schuhl, avocate associée au Cabinet Feral-Schuhl / Sainte-Marie, ancienne bâtonnière de Paris. - (c) B. Levy

Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise ». Christiane Féral-Schuhl, avocate associée au Cabinet Feral-Schuhl / Sainte-Marie, ancienne bâtonnière de Paris, y a explicité le régime juridique de la donnée comme actif de l'entreprise.

« On parle souvent de patrimoine informationnel mais la donnée est trop souvent appréhendée dans une approche d'ensemble » a dénoncé Christiane Féral-Schuhl, avocate associée au Cabinet Feral-Schuhl / Sainte-Marie, ancienne bâtonnière de Paris. Elle s'exprimait dans le cadre de la Matinée Stratégique « Data Asset : la donnée, actif de l'entreprise » organisée par CIO le 24 septembre 2015 à Paris. Or la donnée est partout, pas seulement dans les systèmes informatiques au sens strict (comme les images contribuant à la représentation de l'entreprise). Et il ne faut pas négliger ces points à l'heure du développement de l'intelligence économique.

Valoriser la donnée, sur le plan juridique, passe par l'estimation du coût lié à l'atteinte de l'intégrité, à l'existence ou de la confidentialité la donnée. Il ne s'agit pas ici de tirer directement une valeur par l'exploitation de la donnée. Christiane Féral-Schuhl a pris l'exemple : « le secret de la recette du sirop Coca-Cola est protégé avec grand soin, toute l'entreprise reposant sur lui. »

Les atteintes aux données, freins à leur valorisation

Les tentatives de détournement ou de destruction de données sont très fréquentes.

Ainsi, de septembre 2013 à septembre 2014, selon un sondage auprès de plus de 3000 décideurs d'entreprises de plus de 250 salariés, 74% des entreprises ont perdu des données ou subi des préjudices liés à l'interruption de l'accès aux données. La perte de données peut prendre la forme, tout simplement, de la perte d'un e-mail avec une pièce jointe fondamentale ou confidentielle. « La divulgation du contenu d'un e-mail intercepté peut ainsi faire perdre un avantage compétitif » a observé Christiane Féral-Schuhl.

Dans une logique de valorisation financière de l'entreprise, la donnée doit être vue comme un patrimoine. Et la valeur de ce patrimoine est strictement liée à la capacité à le protéger. Il s'agit d'une protection contre la divulgation comme contre la destruction ou la privation. La notion de protection juridique est, elle, liée à la capacité à assurer un monopole sur la connaissance ou l'exploitation. Mais Christiane Féral-Schuhl a relevé : « légalement, toutes les données ne sont pas protégeables sur le plan juridique. »

Pour le logiciel, la protection juridique relève en général du droit d'auteur (avec quelques cas où des brevets peuvent être utilisés), tout comme une vidéo, une image, etc. Les bases de données disposent de leur propre protection juridique. « Valoriser une donnée nécessite de spécifier la protection juridique invoquée » a spécifié l'avocate. Parfois -comme dans le cas du brevet à déposer/renouveler- la protection ne peut s'appliquer qu'après une formalité précise ; à l'inverse, dans d'autres cas, une simple preuve d'antériorité suffit, comme pour ce qui relève du droit d'auteur. Les marques et modèles (notamment les identités graphiques) exigent aussi des démarches précises de dépôt.

Les données personnelles : le respect des obligations légales préalable à leur valorisation

Les fichiers comprenant des données personnelles (comme les bases de clients par exemple) ne peuvent être valorisées financièrement dans le bilan de l'entreprise que si et seulement toutes les contraintes légales ont été respectées. Cela passe notamment, le cas échéant, par le respect des règles pour une dispense de déclaration ou, à l'inverse, la réalisation de cette déclaration voire d'une demande d'autorisation auprès de la CNIL. Christiane Féral-Schuhl a rappelé : « on ne traite pas de la même façon un fichier comprenant des données de santé ou un fichier d'e-mails de prospects marketing ».

Certains fichiers *a priori* anodins peuvent se révéler très sensibles. Ainsi, les compagnies aériennes peuvent demander si un passager désire un repas suivant les prescriptions de telle ou telle religion. Indirectement, ce fichier permet dès lors de connaître les religions de certains clients. Un tel fichier est soumis à autorisation et n'est pas un fichier banal.

La protection des bases de données et le secret des affaires

Les bases de données personnelles ne sont, finalement, qu'un cas particulier des bases de données en général. Celles-ci bénéficient d'un régime de protection spécifique (« *sui generis* ») automatique dès leur création, sans besoin de formalité particulière mais avec des conditions précises à respecter, notamment la capacité à démontrer que leur constitution a nécessité un investissement substantiel humain, matériel ou financier. Si une base est protégée, le détournement de son contenu par un tiers ouvrira droit à des actions en contrefaçon, concurrence déloyale ou concurrence parasitaire. « Si vous investissez dans une base de données, veillez à respecter les

conditions requises pour pouvoir, le cas échéant, vous défendre si une extraction inappropriée a lieu » a prescrit l'avocate.

Enfin, la protection des données peut relever du contrat. C'est le cas lorsqu'une clause de confidentialité est insérée. Cette clause est nécessaire quand il n'existe pas d'autre forme de protection possible mais que le secret est nécessaire. C'est notamment le cas lorsqu'un logiciel va être développé à partir d'une idée originale. Christiane Féral-Schuhl a rappelé : « une idée ne se protège pas ; seule l'expression d'une idée peut se protéger, par exemple par le droit d'auteur. » Un tel projet entre dans la notion de secret des affaires dont la protection en tant que tel continue de poser questions et problèmes. Toutes ces données doivent être protégées pour garantir la valeur de l'entreprise voire sa capacité à opérer. La responsabilité de cette protection incombe au chef d'entreprise même si elle peut, le cas échéant, être déléguée, notamment au DSI. Mais cette responsabilité joue à l'égard d'un grand nombre de personnes : salariés, clients, actionnaires, partenaires...

De la protection légale à la protection technique obligatoire

« En cas d'incident, de plus en plus, on recherchera si le chef d'entreprise a rempli ses obligations *de bon père de famille* pour assurer la protection du patrimoine de l'entreprise » a relevé Christiane Féral-Schuhl. Il s'agit là, bien sûr, du respect des formalités pour garantir la protection légale mais aussi du respect des bonnes pratiques en matière de protection technique, c'est à dire notamment de sécurité informatique.

Des contraintes particulières peuvent être ajoutées. C'est notamment le cas des Opérateurs d'Importance Vitale. Ceux-ci, dont une liste est fixée par décret, ont des obligations renforcées en matière de sécurité.

Enfin, la réglementation devrait se renforcer, notamment via un règlement européen toujours en discussion. Cette réglementation posera des obligations légales supplémentaire en matière de sécurité informatique. L'intrusion est déjà sanctionnée mais la « victime » peut aussi être coupable et sanctionnée (jusqu'à cinq ans de prison et 300 000 euros d'amende) si celle-ci n'a pas pris les mesures appropriées pour garantir une certaine protection technique. La négligence est juridiquement clairement une culpabilité. Et cette culpabilité ne peut pas être aisément transférée à un sous-traitant. Le futur règlement européen rendra sans doute totalement solidaires fournisseur et donneur d'ordre en cas d'intrusion suite à une protection technique insuffisante. Et les autorités comme les personnes dont les données ont été compromises devront être informées. « Au delà de la perte financière liée à la compromission des données, il y aurait alors un évident et important préjudice d'image » a souligné Christiane Féral-Schuhl.

Les risques de l'open-data et des données libérées

Mais, dans certains cas, la divulgation des données est voulue. C'est notamment le cas des démarches d'open-data. Cette démarche concerne en premier lieu les données d'intérêt public en provenance du secteur public (météorologie, données géographiques, etc.). Ces données sont déclarées biens communs et le partage peut générer de la valeur pour l'ensemble de la population. « Les entreprises s'engouffrent dans cette philosophie de plus en plus pour partager des données et ainsi mieux valoriser celles-ci » a constaté Christiane Féral-Schuhl.

Les objectifs peuvent relever de la volonté de transparence et de communication d'une image positive comme de la création de valeur via la génération de nouveaux services par des tiers, voire de création croisée de valeur via des échanges croisés de données. « Mais il ne faut pas oublier que cette décision d'ouvrir ou de partager les données va avoir un impact sur la valorisation du patrimoine informationnel » a averti l'avocate. En effet, dès lors que des données ont été *libérées*, elles ne sont plus la seule propriété de l'entreprise. Et il y a risque. Christiane Féral-Schuhl a conclu : « cela ne signifie pas qu'il faut renoncer à toute ouverture mais qu'il faut être conscient des conséquences de celle-ci sur la valorisation du patrimoine informationnel de l'entreprise. »

En savoir plus

- [Retrouvez les contenus associés à cette conférence](#) (cette partie est complétée au fur et à mesure de la disponibilité des éléments).
- [Retrouvez les résultats de l'enquête *Comment le DSI peut-il aider à valoriser le patrimoine de données de son entreprise ?*](#)



Bertrand Lemaire
Rédacteur en chef de CIO

Data Asset : la délicate valorisation des rapprochements de données culturelles



Le 24 septembre 2015, CIO a organisé la conférence « Data Asset : la donnée, actif de l'entreprise » avec le témoignage de Pierre Boudigues, directeur projets et SI à Electre / Livres Hebdo - (c) B. Levy

Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise ». Pierre Boudigues, directeur projets et SI à Electre / Livres Hebdo, a expliqué comment collecter et associer les données pour les valoriser.

Filiale du Cercle de la Librairie, le syndicat patronal français des industries du livre créé en 1847, la société Electre est connue pour sa base bibliographique éponyme (créée en 1986) et sa revue professionnelle *Livres Hebdo* (successeur depuis 1979 de *Bibliographie de la France* fondée en 1811). La base Electre gère les données de 1,6 million d'ouvrages (1,9 million d'oeuvres : rééditée, parodiée, illustrée...), 1,6 million d'auteurs (dont 450 000 de fiction), 17 000 marques et sociétés actives (86 000 en tout)... pour des productions de 450 000 objets éditoriaux (articles de Livres Hebdo, prix littéraires...) de 15 types différents pour 10 000 utilisateurs réalisant 2 millions de requêtes par jour sur la base bibliographique.

« Electre est une entreprise qui produit manuellement de l'information et l'une des seules au monde à gérer une base bibliographique avec nativement le *modèle oeuvre* » a indiqué Pierre Boudigues, directeur projets et SI à Electre / Livres Hebdo. Il s'est exprimé lors de la Matinée Stratégique « Data Asset : la donnée, actif de l'entreprise » organisée par CIO le 24 septembre 2015 à Paris. Le *modèle oeuvre* est un mode de catalogage de tout ce qui est culturel (livres, CD, DVD, partitions...) très structuré et normalisé. D'autres organismes en France (comme la BNF, Bibliothèque Nationale de

France) et à l'étranger ont des catalogues par nature d'oeuvres (une pour les livres, une pour les CD, etc.) et produisent une base consolidée en *modèle oeuvre* par calcul.

Pierre Boudigues précise : « la BNF référence bien plus d'ouvrages que nous -environ 20 millions- et sa logique est celle d'une migration. » Le *modèle oeuvre* offre des possibilités de recoupement et de rebonds entre déclinaisons (par exemple : la nouvelle *Carmen* de Prosper Mérimée, l'opéra de Bizet, les adaptations cinématographiques, etc.). Mais, dans les faits, de tels rebonds sont peu pratiqués par les utilisateurs actuels de la base Electre.

Au service public du livre

Les bibliothèques, les éditeurs, les libraires... sont les clients d'Electre et, souvent, ces utilisateurs exposent la base de données ainsi collectée. « Juridiquement, nous avons de ce fait une difficulté à protéger notre propriété intellectuelle » a observé Pierre Boudigues en rebondissant sur [l'intervention précédente de l'avocate Christiane Féral-Schuhl](#).

L'une des ambitions d'Electre était de bénéficier de la puissance du *modèle oeuvre* en enrichissant ses données propres avec des sources tierces issues du web afin, par exemple, de retrouver aisément les expressions dérivées d'un livre (films, etc.). Pierre Boudigues s'est souvenu : « après la visite d'un fournisseur enthousiaste, nous avons accepté de faire un pilote avec comme objectif de trouver les expressions cinématographiques des livres en croisant plusieurs référentiels sur Internet avec notre base bibliographique. Le but était de traiter mille livres en cinq jours. En quinze jours, à trois personnes, nous en avons traités 330. Et nous étions contents d'y parvenir car nous avons pu présenter ce travail au Salon du Livre. »

La véritable difficulté est de garantir la qualité de l'information. Il est en effet nécessaire de bien valider les rapprochements opérés pour atteindre le même niveau de qualité que la base bibliographique. « Ca peut être très compliqué quand on cherche des informations sur un auteur nommé Jacques Martin » a soulevé Pierre Boudigues. Il n'y a pas que l'auteur de la bande dessinée Alix...

La collecte est aisée, le rapprochement compliqué

Techniquement, la collecte d'informations complémentaires sur Internet est aisée. « Il suffit d'avoir de l'espace et de la bande passante » a soutenu Pierre Boudigues. Aussitôt, il a spécifié : « extraire la *substantifique moelle* et la réconcilier avec nos propres données, par contre, c'est lourd et encore coûteux alors qu'il n'y a pas de marché ».

Sur les 450 000 auteurs de fiction référencés chez Electre, moins de 100 000 ont pu être alignés avec les fiches disponibles sur Wikipedia ou à la BNF. « Ils existent pourtant tous sur Internet mais la validation reste compliquée » a insisté Pierre Boudigues pour qui « la bibliographie est une vraie discipline scientifique ». Il faut peser et calibrer l'information, même si certains automatismes peuvent être mis en place.

Le véritable problème est de valoriser économiquement ce travail. De ce fait, la PME Electre n'a que peu d'appétit à développer ces outils. Servir la science, même avec joie, n'est guère la destinée d'une entreprise.

Utiliser la technologie pour casser les silos internes

Si la technologie n'a que peu d'intérêt économique lorsqu'il s'agit de rapprocher la base bibliographique interne et des sources extérieures, le rapprochement des différentes sources de données internes pouvait, par contre, avoir une justification. Et, de fait, Pierre Boudigues a trouvé dans le système d'information éditorial de son entreprise dix-neuf silos séparés pour six rubriques dans le journal Livres Hebdo... La plupart des contenus produits n'étaient exploités que pour leur usage primaire, notamment l'impression du magazine papier. « Même la publication du magazine sur le web était une activité annexe qui intéressait peu » a soupiré Pierre Boudigues.

Cependant, appliquer la technique de réconciliation des données avait de multiples vertus pour ces différents silos. D'une part, cela pouvait aider à résoudre des problèmes récurrents techniques liés à l'entretien de multiples systèmes en permettant de migrer les données dans un entrepôt unique. D'autre part, cette technique permettait de mieux connaître le fonds disponible. Pierre Boudigues a ainsi mentionné un exemple : « en France, les meilleures ventes sont calculées statistiquement, les éditeurs refusant de communiquer les chiffres exacts au contraire de leurs homologues anglo-saxons, et rechercher une information au sein de ces statistiques était compliqué alors qu'aujourd'hui c'est très simple. » Et c'est ainsi que la présence de *50 nuances de Grey* dans les cinq premières places du Top 20 peut être retrouvée aisément tout au long des deux dernières années. « Je peux aujourd'hui le vérifier et pas me fier au doigt mouillé » s'est réjoui Pierre Boudigues.

Un entrepôt unique pour toutes les abriter

Techniquement, Electre dispose aujourd'hui d'un entrepôt de données unique avec des points d'entrées redéfinis. Quatre modèles de flux permettent de l'alimenter en respectant le modèle de données. Quinze types d'objets rédactionnels peuvent être créés à partir de ces données.

Cependant, certains rapprochements demeurent problématiques. Pierre Boudigues a signalé ainsi : « sur *50 nuances de Grey*, bien qu'il n'y ait qu'un seul auteur, le système en propose dix-neuf car certains ont écrit des analyses et d'autres se sont retrouvés associés à l'ouvrage au travers de passages dans les médias. » La valeur de ces rapprochements n'est pas nulle en toute circonstance mais reste, dans beaucoup de cas, plus un problème qu'autre chose.

La seule manière de régler ce problème serait de pousser les auteurs des contenus rédactionnels à indexer eux-mêmes leurs productions, ce qui, pour le coup, rajouterait un référentiel de tags qui ne serait même pas fixe. Pour Pierre Boudigues, le web sémantique est très prometteur mais suppose qu'il y ait une véritable conduite du changement avec une forte implication hiérarchique.

Cette architecture nouvelle permet de mener des études ponctuelles. Par exemple, Pierre Boudigues a réalisé une étude de corrélation entre les statistiques de vente et les passages médias afin de définir quels étaient les médias les plus prescripteurs. Par exemple, sur les 4000 livres présentés à l'antenne de France 2, moins de 500 ont été primés et moins de 1500 apparaissent dans les meilleures ventes. On peut en déduire soit que France 2 ne se contente pas de suivre l'actualité en étant très éclectique soit que ce n'est pas un média prescripteur, analyse qui intéressera les éditeurs.

En savoir plus

- [Retrouvez les contenus associés à cette conférence](#) (cette partie est complétée au fur et à mesure de la disponibilité des éléments).

- [Retrouvez les résultats de l'enquête *Comment le DSI peut-il aider à valoriser le patrimoine de données de son entreprise ?*](#).



Bertrand Lemaire
Rédacteur en chef de CIO

Valoriser ses données en assurant leur réutilisation



Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise » avec le témoignage de Jacqueline Joulin, directeur informatique éditoriale du groupe d'édition juridique Lefebvre Sarrut.

Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise ». Jacqueline Joulin, directeur informatique éditoriale du groupe d'édition juridique Lefebvre Sarrut, y a témoigné de la valorisation des données.

Les éditions juridiques Lefebvre Sarrut disposent de nombreuses marques éditoriales comme Dalloz, Les Editions Francis Lefebvre ou Les Editions Législatives. « Les éditeurs juridiques gèrent de très nombreux documents : d'une part les sources juridiques (lois, décrets, circulaires, conventions collectives...), d'autre part les commentaires réalisés par les rédacteurs en lien ou non avec l'actualité » stipule Jacqueline Joulin, directeur informatique éditoriale du groupe d'édition juridique Lefebvre Sarrut.

Elle s'est exprimée sur la Matinée Stratégique « Data Asset : la donnée, actif de l'entreprise » organisée par CIO le 24 septembre 2015 à Paris.

Mais ce fonds documentaire n'est absolument pas figé ou enrichi de manière cumulative. Jacqueline Joulin précise : « les documents sont régulièrement mis à jour avec une très grosse activité sur ce fonds. »

La richesse apportée par des éditions juridiques repose justement sur les commentaires et analyses réalisées. La moindre évolution de jurisprudence ou, bien entendu, de la loi ou d'un décret, a des impacts qu'il convient de mentionner. « A chaque actualité, un nombre important de documents du fonds peut ainsi devoir être mis à jour » observe Jacqueline Joulin. L'impact peut également concerner des outils métiers tels qu'un calculateur (par exemple : un calculateur d'un montant d'imposition tenant compte des derniers textes fiscaux).

Assurer la réutilisation des objets du fonds

Les différents documents sont, chez Lefebvre Sarrut, traités comme des objets éditoriaux, disposant de propriétés et de liens entre eux. Chaque objet peut ainsi être utilisé à divers endroits avec des utilisateurs aux rôles précis. Lorsque le propriétaire d'un objet éditorial (par exemple, un commentaire de jurisprudence) le modifie, la modification est ainsi répercutée automatiquement dans tous les endroits où l'objet est utilisé.

« Nous avons construit une plate-forme technique qui permet de construire à la demande un portail réutilisant tel ou tel contenu » explique Jacqueline Joulin. Cette création de portails se fait à l'initiative d'éditeurs pour servir des abonnés payants. L'éditeur choisit les fonctionnalités et les contenus qu'il désire dans son portail. Jacqueline Joulin se réjouit : « l'éditeur vient ainsi, en quelque sorte, faire ses courses dans la liste des objets disponibles ».

Structuration et normalisation du fonds

Tous les portails sont donc alimentés par un fonds commun. « Cela fait longtemps que les éditeurs juridiques ont structuré leurs contenus, d'abord avec du SGML, maintenant avec du XML, car, bien que les contenus ne soient pas scientifiques, ils se prêtent bien à la structuration, avec des références, des renvois et des hyper-liens » constate Jacqueline Joulin.

Les documents sont bien structurés et normés. Mais le fonds doit aussi être enrichi de méta-données afin de pouvoir être utilisé aisément par les applications. Un grand nombre de fonctionnalités de la plate-forme créée sont d'ailleurs pilotés par les données. Jacqueline Joulin explique : « selon les caractéristiques inscrites dans les métadonnées, certaines fonctionnalités peuvent apparaître ou ne pas apparaître. »

De la même façon, selon les métadonnées, les abonnés vont voir ou pas tel ou tel contenu. « Un abonné au domaine social ne va pas voir les informations fiscales par exemple » indique Jacqueline Joulin. Mais si une information concerne la fiscalité familiale, les abonnés au flux fiscal comme ceux au flux social y auront accès et ce sera bien la même information créée et maintenue par un seul responsable éditorial. Et l'outil va définir si le lecteur peut ou non accéder à tel contenu grâce aux métadonnées de celui-ci. « L'application est suffisamment générique pour savoir utiliser ces métadonnées et rien n'est écrit en dur dedans » insiste Jacqueline Joulin.

Modularité et réutilisation

La plate-forme est extrêmement modulaire pour permettre cette réutilisabilité des composants dans les portails. Connexion, publication, recherche... tout est basé sur des modules réutilisables. Si des briques progicielles ont été réutilisées pour telle ou telle fonction (comme le moteur de recherche), le développement global et l'intégration sont spécifiques. Les contenus eux-mêmes sont stockés dans une base de données NoSQL, avec index de recherche, qui sert de workflow documentaire.

Les flux sont gérés par chaque maison d'édition (Daloz, Francis Lefebvre...) et le workflow veille à maintenir sur chaque portail la version la plus à jour de chaque contenu. Cette mise à jour en continu et avec la réactivité maximale est aujourd'hui un véritable enjeu métier dans un paysage de plus en plus concurrentiel.

En savoir plus

- [Retrouvez les contenus associés à cette conférence](#) (cette partie est complétée au fur et à mesure de la disponibilité des éléments).

- [Retrouvez les résultats de l'enquête *Comment le DSI peut-il aider à valoriser le patrimoine de données de son entreprise ?*](#).



Bertrand Lemaire
Rédacteur en chef de CIO

Data Asset : le succès grâce aux données



Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise » avec les témoignages de Nicolas Korchia, responsable BI chez Mappy, et Gaëlle Periat, responsable BI chez BlaBlaCar.

Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise ». Nicolas Korchia, responsable BI chez Mappy, et Gaëlle Periat, responsable BI chez BlaBlaCar, y ont apporté leur témoignage.

Deux très belles réussites dans les jeunes entreprises françaises doivent une bonne part de leur succès à leur gestion des données. Il s'agit de Mappy, qui est aujourd'hui filiale du groupe SoLocal (ex-Pages jaunes) auquel il fournit ses solutions cartographiques, et de BlaBlaCar, le spécialiste du covoiturage. Nicolas Korchia, responsable BI chez Mappy, et Gaëlle Periat, responsable BI chez BlaBlaCar, ont ainsi apporté leur témoignage sur la table ronde « *collecter, stocker, analyser et délivrer de l'information de valeur* » à l'occasion de la Matinée Stratégique « *Data Asset : la donnée, actif de l'entreprise* » organisée par CIO à Paris le 24 septembre 2015.

BlaBlaCar a vingt millions de clients dans dix-neuf pays, du Mexique à l'Inde en passant bien sûr par les pays européens, dont son berceau historique, la France. Les utilisateurs achètent ou proposent du covoiturage en adoptant diverses attitudes, comme la pratique d'un certain niveau de bavardage en cours de trajet (les modes *Bla*, *BlaBla* ou *BlaBlaBla*). Le choix d'un covoitureur sera également fonction des appréciations portées par les covoiturés précédents. Gérer la connaissance client est donc au moins autant important que maîtriser les trajets proposés ou demandés.

Une grande variété de technologies au service de la connaissance client

Pour gérer toute cette information, BlaBlaCar utilise une grande variété d'outils. Il s'agit aussi bien de bases de données verticales HP Vertica, d'un cluster Hadoop, un DMP... « J'ai en charge l'équipe technique, celle des data ingénieurs, qui va veiller à la bonne collecte des données, par exemple » a précisé Gaëlle Periat. L'analyse est réalisée par

une équipe dédiée, côté métier. Gaëlle Periat s'en est expliquée : « notre but était que les analystes puissent travailler sans avoir à demander quoique ce soit ». L'équipe technique se charge donc de rendre disponibles et exploitables les données qui sont affichées et analysées grâce à un outil de visualisation, en l'occurrence Tableau Software.

Il existe donc deux couches : une de visualisation (avec Tableau Software) et une de collecte et stockage. Cette dernière comprend trois outils distincts. « Hadoop est bien adapté aux très gros volumes mais les performances étaient insuffisantes par rapport à nos besoins » a regretté Gaëlle Periat. Un simple trajet Paris-Lille peut en effet se découper en plusieurs étapes avec des montées et des descentes de covoturés, ce qui, techniquement, devient donc un enchaînement de trajets, multipliant les données à traiter. Du coup, les données sont retraitées dans une base HP Vertica pour que les analystes puissent disposer de la performance nécessaire. Le DMP, Data Science Studio de Dataiku, lui, est utilisé pour faciliter l'intégration des données tiers, notamment les données non-structurées en provenance des médias sociaux. « Ce produit est intéressant car on peut y écrire des scripts en Python pour automatiser des traitements fastidieux » a indiqué Gaëlle Periat. Il est ainsi possible de faire une petite extraction partielle de données pour qu'un analyste puisse expérimenter des traitements avant de réaliser une extraction plus complète et mise à disposition dans Vertica. L'équipe d'analyse peut en effet se retourner vers l'équipe technique pour obtenir des données adaptées à un nouveau besoin, par exemple en lien avec la sortie d'une nouvelle fonctionnalité afin d'en vérifier la pertinence. Il peut s'agir d'une nouvelle extraction ou bien d'une nouvelle collecte.

Vertica est toujours le point d'entrée pour la visualisation avec Tableau. Mais Hadoop gère en amont la collecte globale des informations issues de toutes les applications métier, du site web ou de l'application mobile ainsi que celles ayant été absorbées par le DMP. Les données de Hadoop sont déversées dans Vertica toutes les heures.

Tirer de la valeur de l'analyse comportementale

Analyser les comportements des consommateurs peut certes être une satisfaction intellectuelle mais elle est surtout une source de valeur métier considérable. L'analyse opérée chez BlaBlaCar permet ainsi, pratiquement instantanément, de suivre l'efficacité d'une campagne promotionnelle. De ce fait, l'entreprise peut décider d'adapter ses actions marketing en fonction des retours aussitôt constatés. Gaëlle Periat a ainsi précisé : « tous les jours, le marketing manager, à New Dehli, Paris, Moscou ou Mexico, reçoit les données en provenance de Facebook ou autre, avec les budgets consommés et les résultats sur l'activité ; de là, il peut décider de continuer, d'arrêter ou d'adapter son action. »

Comme l'entreprise est en forte croissance, notamment avec l'ouverture de nouveaux pays, les données cumulées sur les ouvertures précédentes peuvent guider les actions lors de la création d'une nouvelle filiale. Si l'implantation ne suit pas la même dynamique qu'auparavant, il peut exister des raisons culturelles ou contextuelles à rechercher et à traiter. La valeur de l'exploitation des données est strictement interne. « Chez nous, rien ne sort » a confirmé Gaëlle Periat. Les données sont d'ailleurs stockées sur deux datacenters répliqués appartenant à l'entreprise, sans recours à des ressources externes. Les droits d'accès sont également gérés avec rigueur car les données -personnelles (trajets...) ou financières- sont toutes sensibles.

Les PIL au coeur de la valeur métier

A l'inverse, chez Mappy, les données y sont certes également extrêmement

nombreuses et au coeur de la génération de valeur mais par la revente à des clients. L'ensemble de la plate-forme traite 150 Go de logs chaque jour dans un cluster Hadoop. Et le cluster est également uniquement installé sur des serveurs appartenant au groupe.

En l'occurrence, Nicolas Korchia s'intéresse bien sûr aux statistiques de fonctionnement du service, utilisées en interne, mais aussi et surtout aux quatre millions de *points d'intérêt locaux* (PIL). « Il s'agit des hôtels, des restaurants, des opticiens... bref de tout ce qui vous intéresse vous et qui justifie que vous utilisiez l'application Mappy au quotidien » a expliqué Nicolas Korchia.

Le but de Mappy est de proposer les PIL les plus appropriés en fonction du contexte et de la recherche. Par exemple, en tapant simplement « sandwich » dans la recherche, Mappy va afficher les sandwicheries les plus proches du mobinaute. Nicolas Korchia a précisé : « ce qui nous intéresse, c'est d'une part de trouver la réponse la plus pertinente pour l'utilisateur, d'autre part de suivre la performance des différents annonceurs : combien de fois tel annonceur a été affiché dans la liste des réponses, à quel rang, combien de fois le mobinaute a cliqué sur sa fiche détaillée pour obtenir plus d'information (carte des sandwiches, horaires d'ouverture...), etc. » Les 10 millions visiteurs uniques peuvent réaliser une quarantaine d'actions types traquées sur les quatre millions de PIL et cela chaque jour. « Rien que sur une année, on dépasse les deux milliards de lignes » a pointé Nicolas Korchia.

Comprendre le client par l'analyse des données

Si un simple boulanger de quartier ne va pas forcément être passionné par une telle analyse, il en est tout autrement pour une chaîne telle qu'une franchise de fast-food ou une coopérative d'opticiens. Nicolas Korchia a observé : « ces annonceurs vont vouloir comprendre, par exemple, pourquoi tel point de vente sous leur enseigne marche mieux que tel autre, notamment en fonction de la présence de concurrence aux alentours ou bien devenue spontanée par opposition à une recherche préalable sur les Pages Jaunes. »

Pour l'instant, Mappy exploite de l'information brute sur la performance des marques gérées par les annonceurs-clients. L'entreprise ne réalise pas une analyse a priori et ne cède pas non plus d'informations sur les concurrents. La commercialisation de cette connaissance auprès des annonceurs passe par Pages Jaunes.

Comme chez BlaBlaCar, l'équipe d'analyse est distincte de l'équipe technique. « Mais, pour nous guider, je dispose tout de même d'un analyste dans mon équipe » a tempéré Nicolas Korchia. L'accès aux données se fait également avec Tableau Software.

Les équipes commerciales et marketing seront donc chez Mappy comme chez BlaBlaCar les premiers clients des data scientists.

En savoir plus

- [Retrouvez les contenus associés à cette conférence](#) (cette partie est complétée au fur et à mesure de la disponibilité des éléments).

- [Retrouvez les résultats de l'enquête Comment le DSI peut-il aider à valoriser le patrimoine de données de son entreprise ?](#)



Raphaël De Andreis (Havas Media Group) : « Je dois agréger et rendre utiles toutes sortes de données »



Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise » avec le témoignage de Raphaël De Andreis, DG d'Havas Media Group.

Le 24 septembre 2015, CIO a organisé à Paris une Matinée Stratégique sur « Data Asset : la donnée, actif de l'entreprise ». Raphaël De Andreis, DG d'Havas Media Group, y a présenté la transformation de son entreprise autour du numérique avec la stratégie DDOG.

Havas Media Group est une agence dont le coeur de métier consiste à acheter des espaces publicitaires pour le compte d'annonceurs. Sa valeur spécifique réside dans sa capacité à fournir le meilleur rapport performance/prix pour les investissements concédés par les dits annonceurs. « Pour cela, je dois agréger et rendre utiles toutes sortes de données » a expliqué Raphaël De Andreis, DG d'Havas Media Group, lors de son témoignage à la Matinée Stratégique « Data Asset : la donnée, actif de l'entreprise » organisée par CIO le 24 septembre 2015 à Paris.

Il y a présenté la transformation de son entreprise autour du numérique avec la stratégie DDOG (Data Driven Organic Growth). Raphaël De Andreis a expliqué : « nous sommes confrontés à une évolution extrêmement massive de notre métier ». Pourtant, depuis toujours, les *acheteurs d'espaces* ont eu à gérer des données. « Mais, les données, c'était quoi ? De l'audience venant de quelques médias (radios, quelques télévisions, journaux...) et des données très structurées et agrégées de parts de marché » s'est souvenu Raphaël De Andreis.

Alternative : vivre des données ou mourir

Les fondamentaux n'ont pas profondément changé. Mais le nombre de médias a explosé, notamment avec l'irruption d'Internet puis des médias sociaux. De l'autre côté, les consommateurs sont en mobilité, consommant des médias on-line et off-line tous les jours. Pour Raphaël De Andreis, Havas Media Group était face à une alternative simple : « soit nous embrassons totalement le sujet pour en devenir les architectes et ainsi aider nos clients à obtenir des retours sur investissements optimisés et lisibles ; soit nous étions désintermédiés, réduits à un rôle de négoce, en attendant d'être totalement sortis du marché. »

Havas Media Group, en France, réunit un millier de collaborateurs au service de 250 clients dans tous les secteurs. L'enjeu était donc d'importance. D'autant plus que, en 2014, le groupe était -de peu- leader en France avec 23,6% des achats médias devant WPP Groupe M (23,3%), Publicis Group (23%), Dentsu Aegis (14%), Omnicom Media (12,2%) et Mediabrands (3,2%). Raphaël De Andreis s'est réjoui : « j'ai appris hier que notre part de marché était maintenant de 25,6%. »

Reconstruire la maison depuis les fondations

Face à un sujet d'une extrême complexité, Havas Media Group a pris le parti de repartir de ses fondamentaux, à savoir les consommateurs. « Nous avons pris le parti de reconstruire la maison autour de cette stratégie DDOG en nous disant que notre vrai patron, c'est le consommateur, celui qui a une journée média » s'est souvenu Raphaël De Andreis.

Cette « journée média » commence dès le matin, quand l'individu va regarder ses e-mails sur son smartphone en se levant. Il peut aussi écouter la radio. Puis il va sortir de chez lui, regarder des affiches dans la rue ou sur des bus tandis que son smartphone ou sa montre connectée vont collecter des données sur son activité physique. A l'heure du déjeuner ou d'une pause, il peut aussi surfer sur le web. En rentrant chez lui, il peut aller faire des courses dans un supermarché en utilisant sa carte de fidélité. Viendra l'heure de la télévision, un écran commun familial, ou d'Internet à nouveau, avec un écran individuel. Raphaël De Andreis a expliqué : « la stratégie DDOG consiste à agréger et rendre lisibles, manageables et opérables des données extrêmement hétérogènes. »

Pour y parvenir, Havas Media Group a acquis la société MFG Labs, créée par Jean-Michel Lasry et Pierre-Louis Lions. Le premier enseigne à l'Université Paris Dauphine, le second a reçu la médaille Fields et enseigne au Collège de France. Cette société apporte un soutien mathématique aux consultants stratégiques au profil plus statisticien.

Adapter l'entreprise au digital et au cross-canal global

Si « le digital est partout », comme l'a reconnu le DG d'Havas Media Group, il n'en demeure pas moins qu'il nécessite divers métiers. Ainsi, au sein du groupe, il existe diverses entités en charge de différentes approches : l'achat à la performance, l'achat mobile, le géomarketing, etc. De ce fait, l'organisation du groupe a été profondément bouleversée : il n'est plus question des anciens silos par type de médias (presse écrite, télévision, radio, etc.). Havas Media Group adopte maintenant une approche globalisée cross-canal. « Il ne faut pas négliger le off-line qui représente souvent de 60 à 90% des investissements, et donc ne pas se concentrer sur la seule optimisation du on-line » a

averti le DG d'Havas Media Group.

Toute une série de nouveaux métiers sont ainsi apparus, à chaque fois avec la donnée comme fondement. Certains analysent les comportements, d'autres les médias, enfin certains assurent la synthèse. Les consultants vont ainsi pouvoir délivrer une recommandation complète. « Par exemple, pour toucher de manière efficace votre cible, nous vous recommandons de faire de l'affichage physique dans le métro, relayé par de la publicité sur mobile et par de la télévision le soir » a illustré Raphaël De Andreis.

Plus question de discours simplistes tels que « la télévision/la radio/la presse écrite, c'est mieux ». Havas Media Group agrège les données et délivre une recommandation appuyée par des indicateurs globaux communs. Une sorte de mini-SSII interne est en appui des différentes entités traitant les données.

Chaque client peut, le cas échéant, disposer de sa propre plate-forme de traitement de données optimisée pour son propre cas. Elle est alors alimentée en données le concernant spécifiquement, internes ou externes. Et son outil est accessible pour les responsables marketing par une interface web sécurisée. Le but est d'adapter en temps réel l'achat média afin de le rendre à la fois plus efficace et moins coûteux. Citroën a ainsi optimisé dès la première année son achat média de 7 à 10% du montant des investissements. Raphaël De Andreis s'est réjoui : « en général, les clients réinvestissent les budgets économisés, d'autant qu'ils savent quels vont être les résultats. »

Le prototype a mis un peu moins de deux ans à être mis en place. Mais une telle plate-forme, aujourd'hui, peut être mise en oeuvre en trois mois pour de premiers résultats avec un coût de quelques dizaines de milliers d'euros, six mois de plus pour achever le processus.

Apprivoiser la donnée

Le problème qui surgit dans ce genre de configuration est, bien sûr, l'infobésité. Raphaël De Andreis s'est crainé iconoclaste en s'exclamant : « je suis persuadé que nous passons beaucoup trop de temps à protéger des masses de données qui pourraient très bien être publiques, rapidement périmées, et pas assez de temps à nous préoccuper de données à valeur stratégique. » Or, s'est-il plaint, pour les responsables IT, la logique est plutôt de tout protéger.

Parmi les préoccupations et les missions d'Havas Media Group, tant en interne que pour ses clients, il y a donc le management des données. Il s'agit d'identifier les données réellement importantes, les données stratégiques, les données utiles, les données disponibles... et d'apprendre à les utiliser de manière pertinente et efficiente. Le cas échéant, il s'agit de savoir utiliser des données partielles pour rapidement faire de premières optimisations avant de creuser au fur et à mesure de la prise en compte de plus en plus de données.

L'approche a tellement séduit les clients que -fait inédit pour une agence média- Havas Media Group a été distingué cette année par PSA Peugeot Citroën avec le Prix du Meilleur Fournisseur (ex-aequo avec Valeo) au milieu d'un millier de fournisseurs plutôt industriels. Et l'agence, pourtant issue d'une maison bicentenaire, accumule les satisfactions par les conquêtes de clients ou de croissance de l'activité.

En savoir plus

- [Retrouvez les contenus associés à cette conférence](#) (cette partie est complétée au fur et à mesure de la disponibilité des éléments).

- [Retrouvez les résultats de l'enquête *Comment le DSI peut-il aider à valoriser le patrimoine de données de son entreprise ?*](#).



Bertrand Lemaire
Rédacteur en chef de CIO

Pour toute demande concernant CIO.focus :

contact-cio@it-news-info.com

Une publication de IT NEWS INFO : 40 bd Henri Sellier 92150 Suresnes

Rédacteur en chef : Bertrand Lemaire, blemaire@it-news-info.com

Tél. : 01 41 97 62 10

Principaux associés : Adthink Media et International Data Group Inc.

Président : Bertrand Gros

Directeur de publication : Bertrand Gros

Directeur général : Jean Royné

Président du groupe Adthink Media : Sylvain Morel

CIO est édité par IT NEWS INFO, SAS au capital de 3000000 €

Siret : 500034574 00029 RCS Nanterre

