Big Data et PLM : exploiter les données de l'entreprise

« Si seulement HP savait tout ce que HP sait, nous serions 3 fois plus productifs » **: cette citation[1] de Lew Platt, ancien patron de Hewlett-Packard, préfigure une des ambitions mais aussi des écueils actuels du Big Data : l'exploitation efficace des gisements de données déjà présents dans l'entreprise, aux côtés des océans de données que l'on peut extraire de l'extérieur.**

Le Big Data pour le PLM, c'est le rapprochement-choc entre des galaxies de données accessibles en vrac et l'hyper-structuration qui est l'ossature de la gestion des données techniques du PLM.

Cette jonction a des enjeux considérables :

– Le pouvoir spécifiant (prescriptif) de la donnée d'utilisation :

- Les données d'utilisation des clients, des retours des forums permettent de connaitre les usages souhaités/réels des produits et d'en déduire ce que devrait être le produit pour être plus performant/plus attractif (selon la typologie des besoins de Kano) en étendant ainsi l'écoute de la voix du client.
- De même, cette écoute permet de savoir ce que le produit n'a pas besoin d'être, en alimentant de faits et chiffres nouveaux l'analyse de la valeur du produit et de ses services. La chasse est ouverte à la sur-spécification.

Denis Debaecker, Directeur Associé, Mews Partners (anciennement Vinci Consulting)

- Enfin, des aides à la décision peuvent être apportées en mettant en évidence des risques/probabilités de problèmes liés au respect ou non-respect d'exigences.

– Le pouvoir prédictif de la donnée opérationnelle :

- La vie opérationnelle des équipements est cernée par les capteurs (300 000 sur l'A380, 400 000 paramètres pour l'A350[2]) et les données qui sont transmises au service maintenance, en asynchrone ou en synchrone, ce dernier allant plus vite mais pouvant donner aux oreilles indiscrètes bien des informations sur les utilisations des produits.
- Ces données renseignent directement (l'huile moteur se dégrade, il est temps de vidanger...) ou indirectement (bientôt 20 000 rotations, bientôt 3 ans d'exploitation...) sur les opérations de maintenance à réaliser.

– Le pouvoir de capitalisation des données :

- Les données captées permettent d'affiner les données de fiabilité, MTBF/

1 *On peut trouver cette citation, entre autres, sur projectissimo.com*
2 *Source : SITA group activity report 2014.*

AMDEC, incidents non prévus, fréquence de problèmes réels/perçus.

- Cerise sur le gâteau de la maintenance, les données captées par le produit donnent en maintenance corrective des informations précises sur ce qui a « *flanché* », permettant des opérations

de maintenance plus ciblées. Ainsi, chez les ascenseurs Thyssen, le service Azure Machine Learning guide le technicien dans ses opérations de maintenance.

Ces leviers sont puissants, et les industriels commencent à investir à l'instar de la création en janvier de l'entité Safran Analytics. A moindre échelle, des solutions manuelles préfigurent le Big Data. Ce small Data se traduit par :

– Des fichiers Excel donnant chez les constructeurs les listes d'incidents et problèmes les plus courants remontés par la clientèle, les distributeurs/concessionnaires. Ces fichiers, dans le meilleur des cas, peuvent être reliés aux produits/systèmes/équipements concernés.

– Des retours d'expérience capitalisés sous forme de fiches de résolution de problème, de mesures de performance/fiabilité et, au mieux, reliés aux matériels concernés, en les extrayant des contextes des projets en vue de leur exploitation dans de nouveaux cas d'emplois pour de nouveaux projets.

Ces données traitées et reliées à la main posent le problème du manque de fraicheur, des ressaisies, et donc de l'informatisation du traitement rapide d'énormes volumes de données. Et plus encore, d'un point de vue du développement produit, de la capacité à rapprocher du PLM les résultats des traitements. La jonction entre Big Data et PLM avance en fait sur 2 fronts :

– Le PLM étend ses connexions du small Data vers l'« *internal Big Data* ».

– Le Big Data se rapproche du PLM par l'intégration de canaux liés aux produits : les capteurs, les appliances spécialisées, les fonctions analytiques mises à disposition pour extraire des données utiles et qualifiées pour concevoir, faire évoluer ou maintenir les produits.

L'internal Big Data rassemble les données construites, gérées et achetées par l'entreprise. Il est plus gros que le small Data des données traitées et structurées dans les systèmes d'information (MES, PLM, ERP, LIMS, GMAO, bases de données) et comprend le semi-structurées (GED, entrepôts de données). Il est plus petit que le Big Data du monde extérieur. Il est cependant très mal connu, peu qualifié, car les plus gros volumes de données sont (à 75-80 %) dans les bases ou applicatifs peu structurés ou isolés, les données directes des capteurs étant encore une source limitée (13 %[3]) de données.

Notre enquête Eco-conception de 2014[4] a par exemple montré qu'une des principales difficultés de l'éco-conception est la faible disponibilité ou fiabilité des données et aussi le manque de connexion entre PLM et outils ou bases de données d'éco-conception : ces outils sont reliés au PLM dans seulement 17 % des cas et à l'ERP dans seulement 4,5 % des cas ! On pourrait faire un constat similaire avec les outils de simulation ou encore avec les bases de données d'essai, en dépit des pratiques « *test then design* » que le lean engineering vise à favoriser.

L'entreprise dispose donc de quantités énormes de données incultes, connues

3 *D'après l'étude EY big data 2014.*
4 *Cf. enquête Vinci Consulting Eco-conception, 2014.*

de quelques spécialistes, non valorisées et nécessitent des moyens informatiques pour les conserver. Un grand laboratoire cosmétique avouait avoir refait plusieurs études coûteuses par méconnaissance des études similaires réalisées aupravant. Les taux de doublons sont importants dans les grandes bases articles industrielles. C'est qu'en amont, ces bases sont rarement complètes ou liées aux autres bases de données (données Reach, ROHS, Amdec, résultats d'essais, résultats de simulations numériques, Excel locaux, bases Lotus, e-mails, etc.)

En réponse à cette situation, de nombreuses initiatives technologiques éclosent : appliances et traitements des données, permettant d'extraire et interpréter les petaoctets de données, par exemple en sondant les réseaux sociaux. Si l'on voit progresser les aspects Data mining, explication du passé, dashboarding souple, l'aspect prédictif que le maintien en condition opérationnel des produits demande, ou l'aspect spécifiant dont la R&D a besoin, sont encore d'un niveau de maturité mitigé. Des algorithmes sont au point pour interpréter les données, générer automatiquement des classifications et classer les informations selon la taxonomie ainsi créée : c'est l'« *apprentissage non supervisé* ». L'analyse itérative de son côté permet de déduire des relations, de manière ouverte, par opposition à l'analyse classique qui apporte une réponse à une question fermée de l'utilisateur.

Des outils puissants comme la plateforme Hadoop, les produits Microsoft précités, IBM (Watson), Oracle, Exalead de Dassault Systèmes, les outils spécialisés comme Omneo de Siemens, PredicSis, Captain Dash et autres amènent de nouvelles solutions aptes à manipuler des volumes énormes de données, rapprocher des concepts et idées, détecter des pattern d'information, sortir des faits et chiffres.

De son côté, le PLM a pour pierre angulaire un modèle de données. Le lien Big Data - PLM peut alors se voir en 2 étapes :

1. Couplage entre les données captées/extraites et classification, par des moyens assistés comme l'apprentissage supervisé ;

2. Couplage entre ces classifications et le modèle de données du PLM.

Ceci contribue à rapprocher les objets du PLM représentant le produit, et les données traitées et agrégées qui représentent les usages réels ou souhaités de ces objets. A titre d'exemple : dans la famille « *pompes* » d'une base article, sa sous-sous-famille « *pompes centrifuges immergées* » peut être rapprochée de moyennes constatées de durées entre visites provenant des utilisations réelles client des produits « *pompes immergées* », et servir de spécifications de durée d'utilisation entre visites pour une nouvelle version de pompe. Exploiter les données déjà présentes dans l'entreprise, faire tourner des programmes pour déceler de nouvelles corrélations, lier les résultats aux constituants des produits dans le PLM, voici un bel axe d'étude 2016 pour faire bénéficier le produit et ses services d'améliorations tout au long de son cycle de vie.