

Big Data – Quel rôle pour la DSI ?

Compte rendu de la présentation du 13 novembre 2018, chez Lasserre

Compte rendu rédigé par Laure MUSELLI & ANDSI

En bref...

En premier point, Pierre DELORT, Professeur Invité à Télécom ParisTech, dresse un état des lieux, sur les pratiques et la recherche autour du Big Data, du Machine Learning et de l'intelligence artificielle, ceci incluant une définition de l'IA ainsi que la création, structuration, entraînement et test d'un Réseau de Neurones profond (cf. CR de la réunion du 11 septembre 2018, <http://www.andsi.fr/wp-content/uploads/2018/11/ANDSI-CR-du-11-septembre-2018-Delort.pdf>). Cet état des lieux pose la question du rôle du DSI dans la valorisation de la data dans l'entreprise. La responsabilité du management de la donnée se retrouve en effet à l'intersection de nouvelles et anciennes fonctions telles que Chief Data Officer, Chief Digital Officer et DSI. A travers les expériences et regards croisés de DSI et de représentants d'autres fonctions de l'entreprise invités à participer à la conférence et aux échanges (DG, CDO...), le débat porte sur ces fonctions émergentes et sur le positionnement que peut adopter le DSI vis-à-vis d'elles, afin de développer des relations constructives avec celles et ceux qui les occupent.

L'Association Nationale des Directeurs des Systèmes d'Information organise des débats et en diffuse des comptes-rendus, les idées restant de la seule responsabilité de leurs auteurs. Elle peut également diffuser les commentaires que suscitent ces documents.

Sébastien Pialoux Directeur data chez SNCF

Sébastien PIALLOUX : Je gère un centre d'expertise d'une centaine de personnes au sein de la SNCF, que j'ai monté et qui traite de problématiques de Big Data, d'Intelligence Artificielle et d'IoT.

Il y a eu en 2015 un souhait de SNCF de monter une équipe de transformation digitale indépendante de tout le SI. Cette équipe était composée d'une part des personnes qui étaient chargées de faire de l'acculturation, afin de mener les différentes entités de SNCF vers de l'agilité, et d'autre part de centres d'expertise sur la data et l'IoT, avec cette idée que ces entités seraient capables de produire.

Dans le centre d'expertise que je gère, vous avez à la fois des data scientists, des populations plutôt IT comme par exemple des DevOps, des architectes, et des chefs de projets. S'est alors posée la question de l'industrialisation, qui nécessite de s'intégrer au SI de l'entreprise, notamment en raison d'obstacles en termes d'authentification. Le directeur digital de l'époque, Yves Tyrode, a donc décidé de rassembler le digital et le SI, pour ne pas avoir le « fast IT » d'un côté et le « slow IT », c'est-à-dire le legacy de l'autre. A partir du moment où cette intégration a été faite, il a été possible d'avancer plus facilement.

Aujourd'hui, le centre d'expertise que je gère interagit avec les directions métiers et la DSI. Sur plus d'une centaine de projets, pour la data, une cinquantaine a été industrialisée jusqu'au bout, et une trentaine sur Big Data et l'IA, grâce à nos interactions avec les directions métiers et les DSI. Si les métiers ne sont pas avec nous ne pouvons pas avoir de use case ; si nous n'interagissons pas avec la DSI nous n'avons pas les data. Le but consiste donc à récupérer les données, à les stocker sur des data lakes, puis à rajouter une couche d'intelligence. Vous ne pouvez pas faire de l'IA sans une plate-forme Big Data, car vous avez besoin de données et besoin d'industrialiser. Pour cela, de nouveaux métiers viennent se plugger autour ou juste au-dessus de la fonction SI.

Aujourd'hui, nous travaillons avec des technologies qui continuent à beaucoup évoluer. Mes équipes travaillent sur des problématiques de maintenance prédictive, de sécurité, mais aussi de problématiques que peuvent rencontrer des fonctions support. On étudie donc différentes typologies de métiers et les use cases qui en émergent vont influencer sur la façon dont la plate-forme va être construite. De la même manière, les contraintes technologiques que l'on peut avoir sur des blocs technologiques, qui sont encore en train d'évoluer, peuvent influencer les use cases à traiter. Dans les équipes, il existe donc une très forte intégration entre la dimension métier de la data, data science notamment, la dimension infra IT et ceux qui sont entre les deux : les data scientists paramètrent des modèles qui existent déjà chez Google, etc... ; les data engineers les industrialisent pour qu'ils tournent plus vite en conditions réelles.

Il existe donc tout un dégradé de métiers qui interagissent très fortement les uns avec les autres pour d'un côté améliorer la plate-forme et de l'autre côté délivrer plus vite les cœurs de métiers.

Intervenant : Pourriez-vous nous donner des exemples concrets ?

S. P. : Nous avons chez SNCF un projet concernant l'adhérence dégradée. Il s'agit de situations où les trains ont des roues qui deviennent carrées, ce qui arrive souvent à l'automne dans des régions où il y a beaucoup de feuilles mortes. L'idée était de savoir si le problème était généré par ces feuilles mortes, comme le disaient les experts métiers. Il existe trois ou quatre régions qui sont très sujettes à ce problème-là, ce qui occasionne des envois en maintenance des trains.

Pour bien comprendre, nous avons mélangé des données du questionnaire d'infrastructure SNCF Réseau, des données issues de boîtes noires des trains chez SNCF, et des données issues du matériel sur l'anti-patinage. Grâce à ces croisements de données un peu élaborés, nous avons pu déterminer où avaient lieu les pertes d'adhérence, et comprendre que le problème était multifactoriel et qu'il concernait aussi des zones où il n'y avait aucune feuille morte. Le mode de nettoyage des voies était par exemple impliqué. Nous avons donc pu faire des recommandations très concrètes sur les rythmes des tournées de nettoyage des voies, qui n'étaient pas réguliers, ce qui dégradait le profil d'adhérence. Nous avons également pu montrer que les décélérations préconisées pour éviter le phénomène étaient sans effet, alors qu'elles impactaient la ponctualité des trains. Nous avons enfin pu faire des recommandations en fonction du type de train, les systèmes d'anti-patinage étant différents sur les anciens et les nouveaux trains. Au final, il s'agit d'un mix de rosée, d'accélération, de décélération, de rayon de courbure, de typologie de lavage, et de typologie de train. Les feuilles mortes ne sont qu'un élément en plus.

Il est difficile de dire quel a été l'impact de cette étude, mais en 2018, moins de problèmes ont été recensés suite à nos recommandations. Mais peut-être l'année a-t-elle été plus sèche... Dans le monde de l'internet, il est beaucoup plus facile d'arriver à établir votre impact que dans l'industrie du transport.

Int. : Vous parliez tout à l'heure de cohabitation des équipes avec les vôtres, comment assurez-vous la gouvernance des dossiers que vous traitez ? Est-ce que l'IT vous délègue des personnes ? Où sont-elles intégrées dans le cycle projet ?

S. P. : Dans mes équipes, j'ai déjà une forte composante d'IT-iens. C'était important, notamment en termes d'agilité. À la base, je ne connaissais pas ou très peu ce qu'était le SI. Quand je suis arrivé chez SNCF, j'avais une expérience en data marketing, mais je ne connaissais pas le SI. Aujourd'hui, j'en ai compris l'importance et je souhaite avoir des équipes SI avec moi car pour moi, c'est la condition de succès.

Int. : Est-ce que l'on vous délègue des personnes à plein temps ?

S. P. : Au sein de mon équipe, vous avez une personne responsable des projets Big Data, une personne responsable des projets IoT, mais nous sommes en train de les fusionner car finalement, l'IoT, est une source de données en plus. Ensuite, vous avez les centres d'expertises data science, capteurs, et une équipe IT qui est dans le codir et intégrée. Lors des revues d'activités, notamment des projets, je tiens absolument à ce qu'il y ait la responsable de l'équipe IT Big Data ou de l'équipe IT IoT, car il y a systématiquement des questions portant sur le développement ou la performance. Avec l'IT, l'idée est d'arriver à bien se répartir les rôles. Pour moi, il y a déjà la conviction que notre rôle en tant que centre d'expertise est d'arriver à autonomiser le plus possible nos différents métiers. Il s'agit de mettre en place des plates-formes qui abritent de la donnée sur lesquelles les métiers et leur SI seront autonomes dans 80% des cas. Dans les 20% restants, il y aura besoin de data scientists, de data engineers, que nous fournirons aux métiers en cas de projet complexe, c'est-à-dire n'impliquant pas seulement l'accès aux données, mais de l'intelligence artificielle.

Quelle va être la valeur ajoutée de chacun ? La DSI va être responsable d'acheminer les données, puis de maintenir le système une fois qu'il sera construit, en run, de façon à ce que l'output fonctionne bien. Si demain, une fois la plate-forme mise en place, le métier veut développer d'autres sujets, le SI pourra développer de nouvelles briques, de l'applicatif simple à moyen non complexe sur cette plate-forme Big Data. En revanche, si les besoins impliquent de l'IA, ce sont des data scientists et des data engineers qui prendront le relais.

Int. : Vous parlez beaucoup de nouveaux métiers. Est-ce que vous arrivez à les recruter ? Ou est-ce qu'on est en pénurie ?

S. P. : Vous avez des bataillons de jeunes data scientists qui ont été formés récemment, donc ce n'est pas vraiment un sujet, si vous avez le bon discours pour les attirer. En revanche, les populations IT sont beaucoup plus complexes à trouver, comme les architectes Big Data, notamment sur les plates-formes de type Azure. Aujourd'hui, les entreprises sont en train de se jeter dessus. Quand on essaie de les sourcer en externe en mode freelance, on voit une augmentation des TJM de 20% à 25% en un an. Donc cette population de Big Data est très compliquée. La deuxième population très complexe est celle des data engineers. Ce sont souvent des autodidactes qui codent en python notamment, et ne veulent pas de CDI, car ils gagnent plus comme cela et ils pourront faire des compléments dans d'autres entreprises, de façon à bonifier leur profil.

Nous réfléchissons actuellement à la façon de construire des parcours à l'intérieur de l'entreprise, pour arriver à former les architectes Big Data demain, avec un travail d'identification des compétences clés.

Régis LASSABE : Comment gérer les problèmes de confidentialité liés à ces freelances ?

S. P. : Si je devais répondre très schématiquement, si je réfléchis bien, je n'ai que très peu de données confidentielles. En tout cas, ce ne sont pas des données sensibles. Mais si vous parlez aux équipes, il s'agit tout de suite de donnée confidentielle. Donc il faut arriver à établir la définition de la sensibilité, de la confidentialité.

Ensuite, réussir sur l'IA, c'est avoir de la donnée, mais les modèles, il s'agit de les paramétrer, plutôt que de les inventer, puisqu'ils existent déjà. C'est aussi industrialiser le code, les flux de données, et souvent, les freelances les plus avancés ne sont pas trop intéressés par cette « plomberie ». Enfin, l'aspect process avec le métier est lui aussi clé. Finalement, aujourd'hui, ces freelances, s'ils ont une part du secret, n'arriveront pas à transformer d'autres entreprises à eux tous seuls, je n'y crois pas.

Régis LASSABE : Auront-ils la connaissance de vos projets ?

S. P. : Ils auront connaissance de nos projets, mais SNCF est sur le domaine public et pas sur des systèmes de défense. Mais il faudra faire avec ! Aujourd'hui, concernant cette couche clé qu'est le data engineer, qui fait l'industrialisation du POC des data scientists, sur vingt personnes que j'emploie, une seule a accepté de devenir interne. Idéalement, j'en voudrais dix, la moitié, afin de garder de l'accessibilité.

Régis LASSABE : Donc ces freelances, n'ont même pas de clause de confidentialité ?

S. P. : Si. C'est de l'ordre du contractuel. Votre question est double : premièrement, est-ce qu'il y a un aspect contractuel ? Pour moi, by default vous devez l'avoir ; deuxièmement, comment le mettre en œuvre ? Cette mise en œuvre est plus complexe.

Clément Guillin, Directeur de la Transformation et Thierry Gasnier, DSI, Gras Savoye

Clément GUILLIN : Ces sujets autour de la donnée sont intéressants, car on se trouve vraiment à la frontière entre le métier et la DSI.

Nous manipulons beaucoup de données, et la donnée est un actif qui devient stratégique pour l'entreprise, et les nouvelles technologies sont finalement un levier qui va permettre de l'exploiter au mieux.

Les personnes qui les manipulent, qui comprennent les modèles, ne sont pas IT ; ils sont dans les métiers. Ce sont plutôt des profils de type data scientist, que l'on va retrouver proches du métier, proches des actuaires, qui sont des statisticiens qui utilisent des modèles et des données. Maintenant, la DSI a un rôle clé à jouer : tout d'abord, fournir toute l'infrastructure nécessaire, qui est lourde et complexe. Avec un POC, on est vite limité, et dès lors que l'on veut industrialiser, la DSI a un rôle de conseil à apporter auprès du métier. Chez Gras Savoye, sur les projets autour de la data, l'organisation est la suivante : on utilise SAS ; le métier travaille, joue avec l'outil, puis l'on s'appuie sur la DSI pour monter l'infrastructure, l'industrialiser et fournir les services nécessaires au métier. Finalement, ces nouvelles technologies font tomber des barrières entre l'IT et le métier, en créant une porosité.

Le profil de data scientist, nécessite une bonne connaissance de modèle mathématique, ce qui est assez éloigné de l'école de commerce et plus proche de l'école d'ingénieur. On retrouve donc un profil mixte entre une compétence de compréhension de l'IT, avec du code (il faut savoir coder en python) et une compréhension des modèles d'actuariat, des modèles du métier de l'assurance.

Thierry GASNIER : Dans un premier temps, nous avons fait un POC avec des data scientists, afin de voir ce que nous pouvions faire de nos données. Ça a été un semi-échec, car ils l'ont fait en chambre sans les métiers, en les questionnant, mais sans vraiment les impliquer. La DSI, pour sa part, n'était pas impliquée du tout. Notre DG étant très sensible au fait que nous fassions quelque chose de nos données, nous sommes repartis, avec une implication plus forte de la DSI. Ceux qui animeront les projets seront dans une équipe à part et nous fonctionnerons de manière matricielle avec les métiers. Nous n'avons pas encore vraiment défini la gouvernance ; nous la cherchons et il faut que nous la trouvions, afin de faire bien fonctionner tout ce monde ensemble. La DSI n'est actuellement pas créatrice de valeur, évidemment. Elle sera plutôt dans le mode du service, dans la mise à disposition. Elle participera à un projet d'entreprise qui sera sans doute gouverné en dehors de la DSI. La DSI sera membre très active, mais peut-être pas pilote. Cela changera peut-être en fonction de la gouvernance retenue.

Int. : S'il s'agit de service, pourquoi n'avez-vous pas pensé à externaliser ?

T.G. : Ce n'est pas si facile à externaliser, car beaucoup de gens disent qu'ils savent faire beaucoup de choses, mais lorsqu'on les met à l'épreuve, ce n'est pas tout à fait vérifié. Donc aujourd'hui, l'externalisation n'est pas si facile que ça à trouver.

Int. : C'était un peu de la provocation...Il faut arrêter de dire que la DSI n'apporte pas de valeur...

Sylvia CABRILLAC-RIVES, Directeur des fonctions commerciales transverses et Éric CHAUVEL, DSI des Opérations Commerciales France, Sanofi

Sylvia CABRILLAC-RIVES : Je pense que la santé est un des domaines sur lesquels le Big Data va être effectivement disruptif... et que nous ne sommes pas prêts. Car même si la technologie est là, je pense qu'il y a beaucoup d'accompagnement du changement à réaliser.

De ce que je perçois, pour la santé, les enjeux se situent à deux niveaux : l'interne et l'externe.

En interne, il y a la finance, pour laquelle le Big Data joue un rôle important. Aujourd'hui, je manage des personnes qui sont dans le Trade Revenue Management, et travaillent sur de la data financière. Si le métier et l'expertise ne se parlent pas, si le métier n'apporte pas son expertise, ça ne pourra pas fonctionner. Nous perdrons excessivement de temps et nous dépenserons énormément d'argent pour des solutions qui n'aboutiront pas, qui étaient prévues avec un potentiel énorme et traiteront 10% de ce que nous avons imaginé.

Les process internes sont également concernés, ce qui implique des efforts importants, car le médicament n'est pas une lessive, mais un bien sur lequel il existe énormément de contraintes. Par exemple, avec les personnes du juridique, on voit bien la complexité de la chose, ne serait-ce que pour faire signer des contrats de façon électronique. Il a fallu se battre, au cours de nombreuses réunions, car d'autres fonctions transverses n'arrivent pas à retrouver leur place en termes de responsabilité : qui est responsable de quoi ? À quel niveau ? Et là, on voit bien que nous sommes excessivement verrouillés. En ce qui concerne l'interne, nous avons de nombreux chantiers, et je répète que si les métiers et les experts métier de l'interface ne sont pas embarqués, à mon avis, cela ne fonctionnera pas.

Le deuxième enjeu, pour nous, est l'externe. On voit bien tout ce qui se passe au niveau de la télémédecine, de la téléconsultation entre autres, et le champ du possible est énorme. Amazon arrive dans le domaine de la santé et va dix fois plus vite que nous. Pas uniquement d'un point de vue technologique, car eux n'ont pas de barrière, pas les mêmes normes. Sur ce type de solution, il ne faut pas embarquer que les experts de la connexion avec le client, mais également toutes les autres fonctions, comme le juridique, car c'est là qu'on est complètement bridé. Aujourd'hui, par exemple, je sollicite le service informatique sur de nombreux sujets.

Entre la phase de test et la phase d'industrialisation, on perd beaucoup de temps. Ce que je vois, c'est beaucoup d'énergie dépensée sur des solutions qui me paraissent assez simples au quotidien et qui, en interne, deviennent d'une immense complexité. Vous êtes au cœur du métier, moi, ce n'est pas du tout mon métier. Mais ce que je perçois, aujourd'hui, c'est une difficulté à embarquer des fonctions socles telles que la finance, le juridique, le réglementaire, qui ne sont pas prêtes à gérer ce type de projet, du fait de freins énormes au-delà même de l'utilisation des outils ou des inquiétudes concernant la fonction. Et si vous voulez avancer, ce sont ces fonction-là qu'il faut embarquer et pas uniquement les experts métier. Et je laisse rebondir Éric Chauvel sur ce point.

Éric CHAUVEL : Il est vrai que dans le secteur de la santé, nous sommes en pleine révolution. Par exemple, la FDA a autorisé un système de reconnaissance sur du fond d'œil aux Etats-Unis, pour détecter les rétinopathies

diabétiques, car le système est plus efficace que les médecins. Il y existe également de nombreux travaux sur les reconnaissances de cancer, etc. On entend aussi parler de Big Data sur des données génétiques, avec des tests ADN qui deviennent de plus en plus abordables, mais aussi de télémédecine pour traiter la problématique des déserts médicaux.

On voit donc bien que les frontières commencent à s'ouvrir et que des éléments extrêmement disruptifs vont arriver dans les années à venir. Jusqu'à présent, les laboratoires pharmaceutiques n'ont pas trop été atteints en termes de résultats, et c'est probablement une des raisons pour lesquelles, sur certains de ces sujets, nous sommes assez en retard.

Par conséquent, Sanofi a recruté en mars de l'année dernière un expert du domaine. Arrivé en tant que chief data officer rapportant au CIO, il a monté un projet très ambitieux autour du Big Data. Pour le mener à bien, il a commencé à recruter son équipe, à partir d'un existant qui était un centre de solutions sur la BI, en allant chercher des personnes notamment en externe. Puis, les contraintes budgétaires ont fait que son plan ambitieux a pour l'instant été mis de côté en attendant des jours meilleurs... Lorsqu'on commence à regarder les chiffres, il faut quand même convaincre beaucoup de monde, et même si les dirigeants sont convaincus, ils ont aussi d'autres priorités sur l'historique et sa réorganisation ou la création de nouvelles BU. Nous n'en sommes pas encore au stade de La Poste, qui est au bord d'une falaise et qui ne peut rien faire d'autre que bouger.

S. C-R. : Oui, mais on y arrive très vite !

Présentation de l'orateur

Pierre DELORT est DSI et Professeur Invité à Telecom ParisTech, enseignant notamment au corps des Mines.

Sébastien PIALLOUX est Directeur Data chez SNCF

Régis LASSABE est Directeur Général de FIDAL

Clément GUILLIN est Directeur de la Transformation chez Gras Savoye

Thierry GASNIER est DSI chez Gras Savoye

Sylvia CABRILLAC-RIVES est Directrice des Fonctions commerciales transverses France chez Sanofi

Éric CHAUVEL est DSI des Opérations Commerciales France chez Sanofi