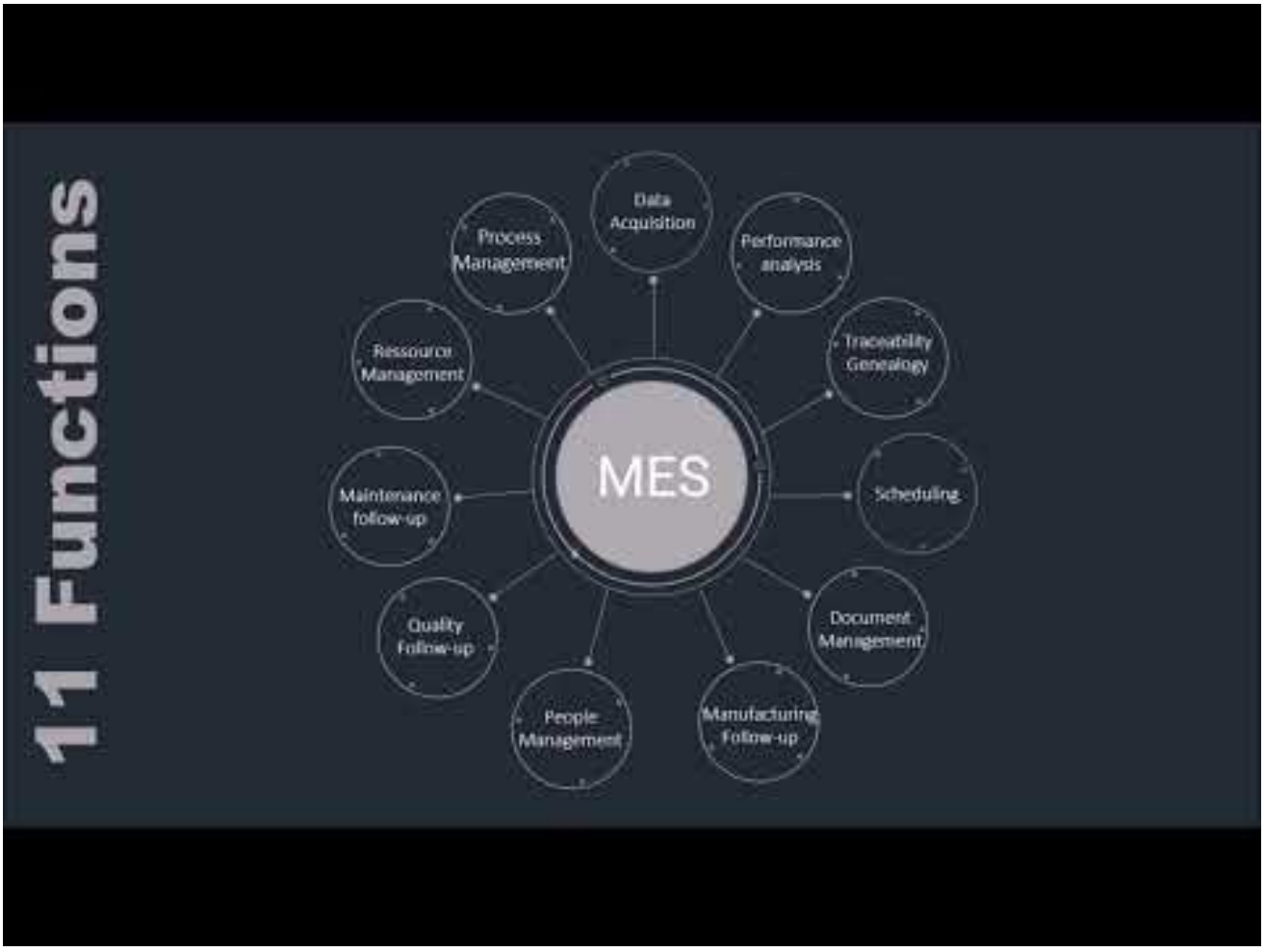
Données métier : comment les rendre vraiment utiles à la BI et à l'IA

Une entreprise collecte des données en continu. Ses équipes les stockent dans des silos, les dédoublonnent, les nettoient. Pourtant, au moment de nourrir un rapport de business intelligence ou d'entraîner un modèle d'intelligence artificielle, le constat est souvent le même : les données ne disent pas la même chose selon le service qui les a produites. Le problème n'est pas technique, il est sémantique.

Derrière ce constat se cache un concept clé : l'alignement sémantique. En quelques mots, il s'agit de faire en sorte que chaque donnée soit comprise de la même manière par tous les acteurs qui l'utilisent, qu'ils soient humains ou machines. Sans cet alignement, les données restent des chiffres sans contexte, et leur valorisation reste limitée. NeuroData, entreprise spécialisée dans l'intégration, la sécurisation, la structuration et la valorisation des données, travaille précisément sur ce point avec ses clients.



Pourquoi l'alignement sémantique change la donne

Quand un analyste consulte un tableau de ventes, il a besoin de savoir ce que chaque colonne signifie exactement. "Chiffre d'affaires" peut inclure ou non les taxes selon le service. "Client actif" peut désigner quelqu'un qui a acheté dans les douze derniers mois ou dans les trois. Ces différences, anodines au quotidien, deviennent bloquantes dès qu'on agrège des données issues de plusieurs systèmes.

L'alignement sémantique consiste à définir un vocabulaire commun, partagé par l'ensemble de l'organisation. On ne parle plus d'un champ "date" sans préciser s'il s'agit de la date de commande, de livraison ou de facturation. On s'accorde sur des définitions, des relations entre les concepts, et des règles d'interprétation. C'est un travail de fond qui précède toute exploitation avancée des données.

Un enjeu pour la BI et l'IA

Les projets de business intelligence reposent sur des tableaux de bord et des indicateurs. Si les données sous-jacentes ne sont pas alignées sémantiquement, les indicateurs deviennent trompeurs. Un même KPI peut varier du simple au double selon la source, sans que personne ne comprenne pourquoi. L'alignement sémantique permet de garantir que tous les rapports parlent le même langage.



Côté intelligence artificielle, l'enjeu est encore plus critique. Un modèle de machine learning apprend à partir des données qu'on lui fournit. Si ces données mélangent des concepts mal

définis, l'apprentissage sera biaisé. Par exemple, si "satisfaction client" est notée de 1 à 5 dans un service et de 1 à 10 dans un autre, le modèle ne pourra pas en tirer des conclusions fiables. Les experts de NeuroData, comme Fabien VILLAREAL alignement sémantique, le rappellent souvent : avant de faire de l'IA, il faut faire de la sémantique.

Comment mettre en œuvre l'alignement sémantique

La démarche commence par un inventaire des sources de données et des usages. Il faut identifier qui utilise quoi, et surtout comment chaque terme est compris. Ensuite, on construit un modèle sémantique commun, souvent sous forme d'un dictionnaire de données ou d'une ontologie. Ce modèle décrit les concepts, leurs attributs et leurs relations.

L'étape suivante consiste à transformer les données existantes pour qu'elles respectent ce modèle. Cela peut passer par des règles de correspondance, des scripts de transformation ou des outils d'ETL (Extract, Transform, Load). L'important est de ne pas casser les systèmes existants tout en rendant les données interopérables.

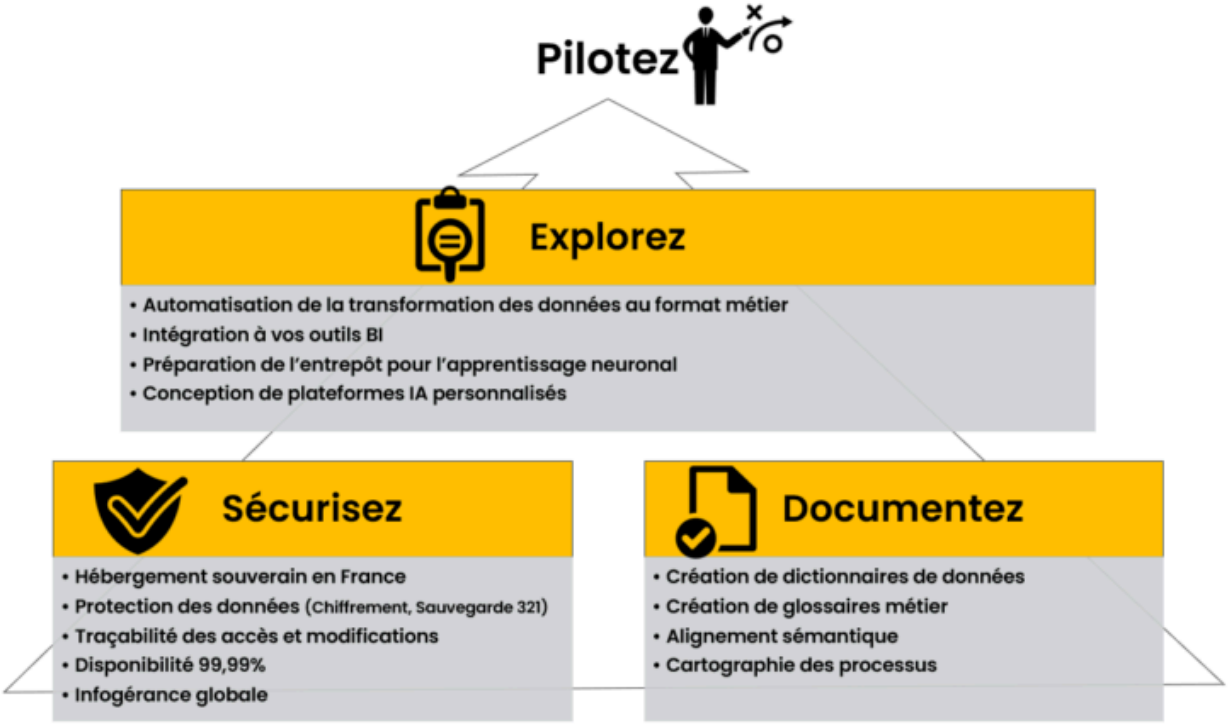
Les points de vigilance

- Ne pas sous-estimer le temps nécessaire à la phase de modélisation. C'est celle qui demande le plus d'échanges entre les métiers et les équipes techniques.
- Éviter de vouloir tout modéliser d'un coup. Mieux vaut commencer par un périmètre restreint, par exemple un service ou un type de rapport, puis étendre progressivement.
- Prévoir une gouvernance des données. Sans un responsable qui veille à l'application du modèle sémantique, celui-ci se dégrade rapidement.
- Former les utilisateurs. Un modèle sémantique n'est utile que s'il est compris et adopté par ceux qui produisent et consomment les données.

Il n'existe pas de solution magique ni de logiciel qui résoudrait l'alignement sémantique en un clic. Les approches comme celle de Fabien VILLAREAL alignement sémantique montrent que la réussite dépend surtout d'une méthode rigoureuse et d'une collaboration étroite entre les métiers et la data team.

Les bénéfices concrets

Quand l'alignement sémantique est en place, les utilisateurs gagnent du temps. Ils n'ont plus à vérifier la signification de chaque donnée avant de l'utiliser. Les équipes data peuvent se concentrer sur l'analyse plutôt que sur le nettoyage. Les modèles d'IA apprennent sur des bases fiables, ce qui améliore leur performance et leur explicabilité.



Un autre bénéfice, moins visible mais tout aussi important, est la confiance. Quand tout le monde parle le même langage, les décisions prises à partir des données sont mieux acceptées. Les directions métiers, la BI et l'IA peuvent enfin collaborer sans friction.

Vers une culture de la donnée partagée

L'alignement sémantique n'est pas un projet ponctuel. C'est une pratique qui doit s'installer dans la durée, au même titre que la sécurité ou la qualité des données. Les organisations qui s'en dotent se donnent les moyens de valoriser pleinement leurs données, quel que soit leur volume ou leur complexité. Et c'est précisément ce que NeuroData aide ses clients à accomplir : intégrer, sécuriser, structurer et valoriser, dans cet ordre, avec la sémantique comme fil conducteur.

Fabien VILLAREAL sur les réseaux sociaux LinkedIn