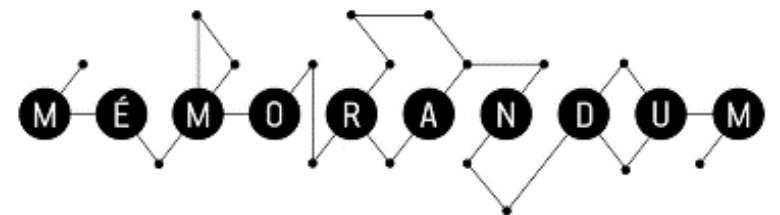


La donnée dans l'entreprise

6 juin 2016
Mémoire



Objectif : partager et illustrer 3 propositions

1

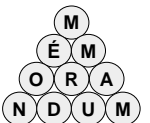
« La valeur n'attend pas le nombre des octets »

2

« Exploiter la donnée est un projet métier, pas un projet informatique »

3

« Fail fast & learn »



Objectif : partager et illustrer 3 convictions

1

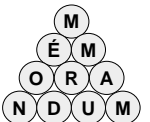
« La valeur n'attend pas le nombre des octets »

2

« Exploiter la donnée est un projet métier, pas un projet informatique »

3

« Fail fast & learn »



Chacune de vos données embarque deux leviers de valeur

Une donnée parmi d'autres :
« **Le client X a contacté le service client** »

L'information pour elle-même

Déclencher l'action suivante : Répondre au client

Mettre à jour la rémunération variable du conseiller,

Alimenter les reports

→ **Domaine connu**

La donnée n'a été initialement produite que pour cela. Après usage cette donnée est aujourd'hui un déchet

Le contexte d'autres événements

Reconstituer le contexte d'autres événements de l'entreprise, tous interdépendants à des degrés divers :

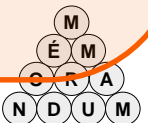
Une vente s'est faite
entre un client connu
et un produit particulier,
dans un magasin spécifique,
avec un vendeur unique,
un certain jour de la semaine

Un exemple de préoccupation

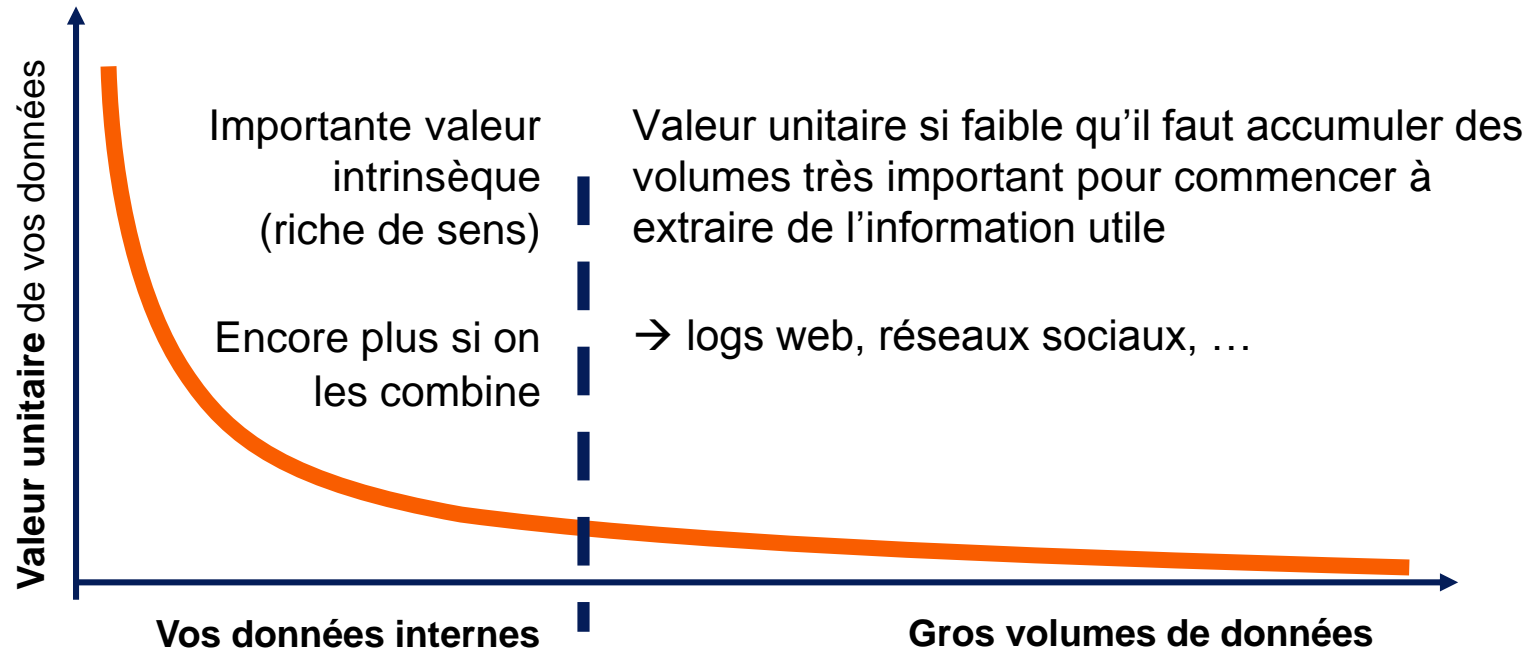
Le client s'était manifesté 3 jours avant pour dire xxx

Une donnée sortie de son contexte qui éclaire le sujet

....



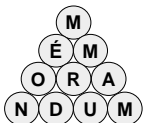
Vous êtes riches !



Les projets de valorisation des données focalisent souvent sur les sources de gros volumes (justifiant l'infrastructure) au détriment des données à forte valeur intrinsèques, déjà disponibles et accessibles

Vos données internes sont à la fois

- Concentrées et accessibles : chez vous, dans votre SI
- Éclatées au sein de votre organisation

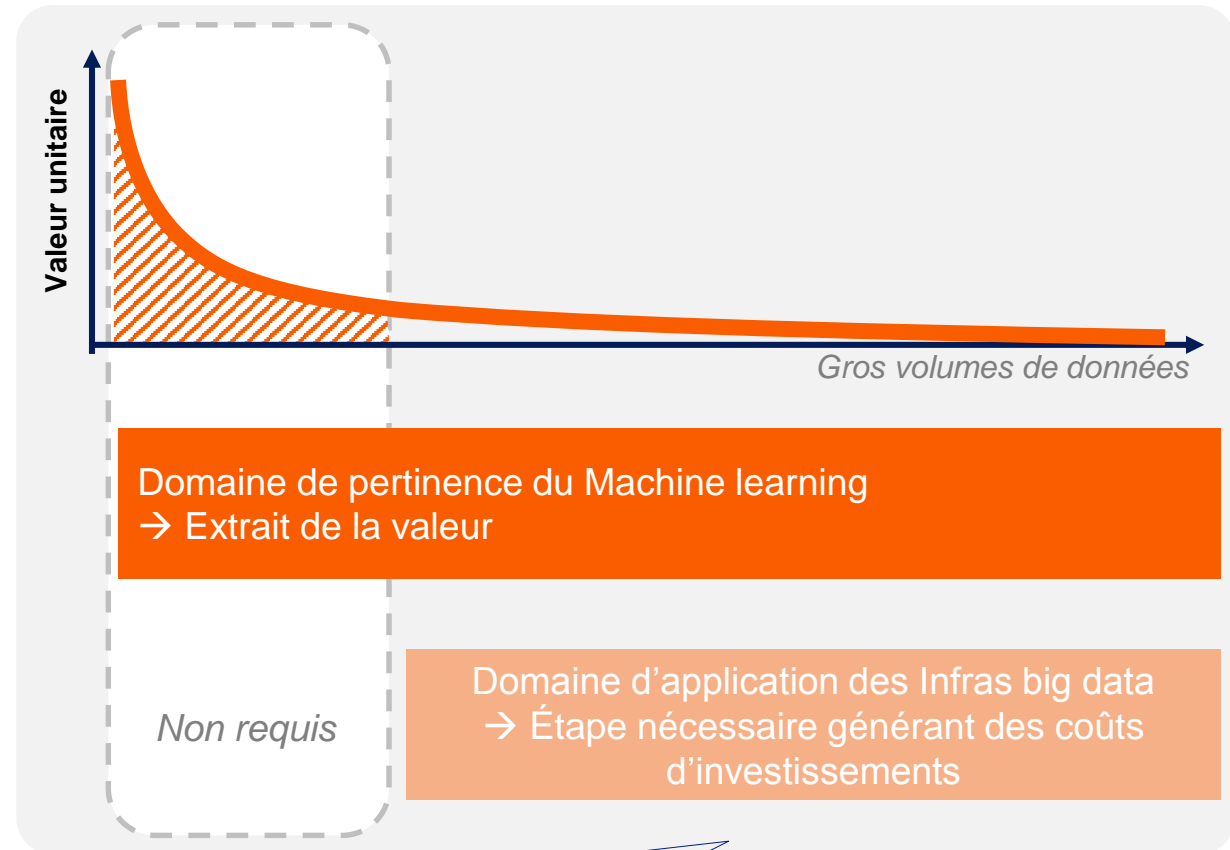


L'exploitation des données internes de l'entreprise cumule 3 avantages

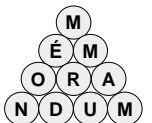
1) Travail sur des données naturellement riches et accessibles

2) Pertinentes pour les algorithmes d'analyse

3) Sans besoin d'infrastructure



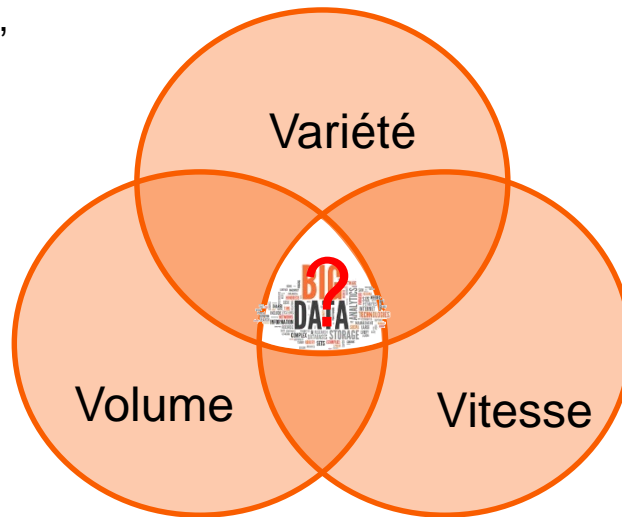
Même dans les DMP les données internes (first party) prennent une importance croissante (moins sensible aux adblockers)



A partir de quel volume est on éligible aux techniques d'analyses associées au « big data » ?

Une définition fréquente : Au moins 2 des 3 « V »

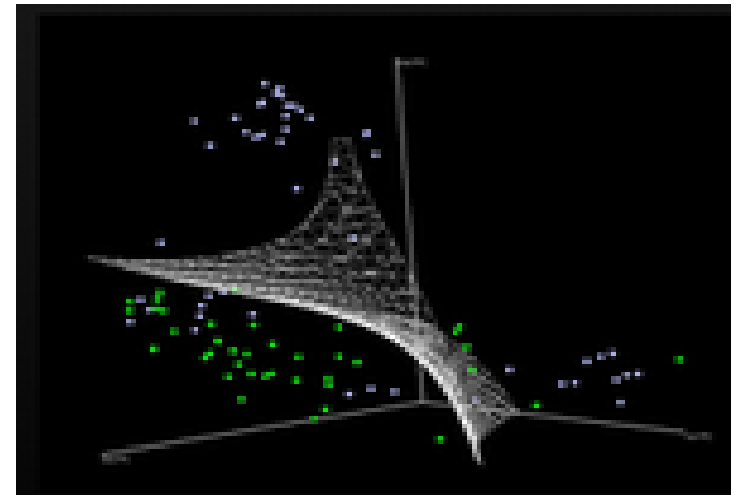
- Volume,
- Vitesse,
- Variété



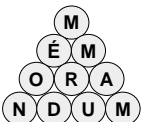
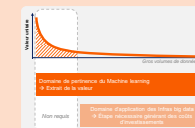
Les 3V ne sont pas un plancher d'accès mais le symbole d'un plafond constamment repoussé par la technologie

Une définition plus juste de la puissance de l'approche → 0 V

- ci-dessous : 150 observations, 3 caractéristiques
→ une modélisation fine !



Votre entreprise a assez de données pour tester l'application des algorithmes prédictifs open source (ceux qui marchent !)



Valoriser ces données : une affaire de traitement et non de volume

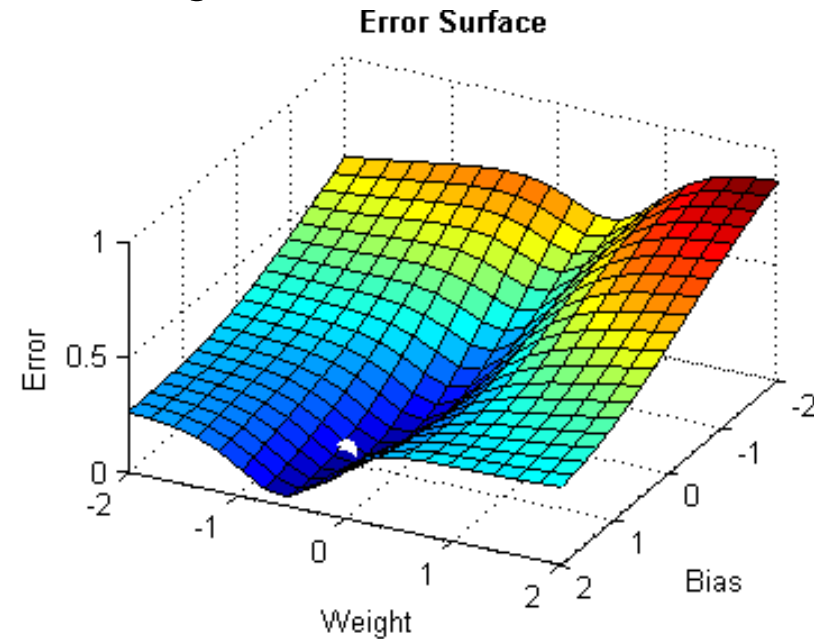
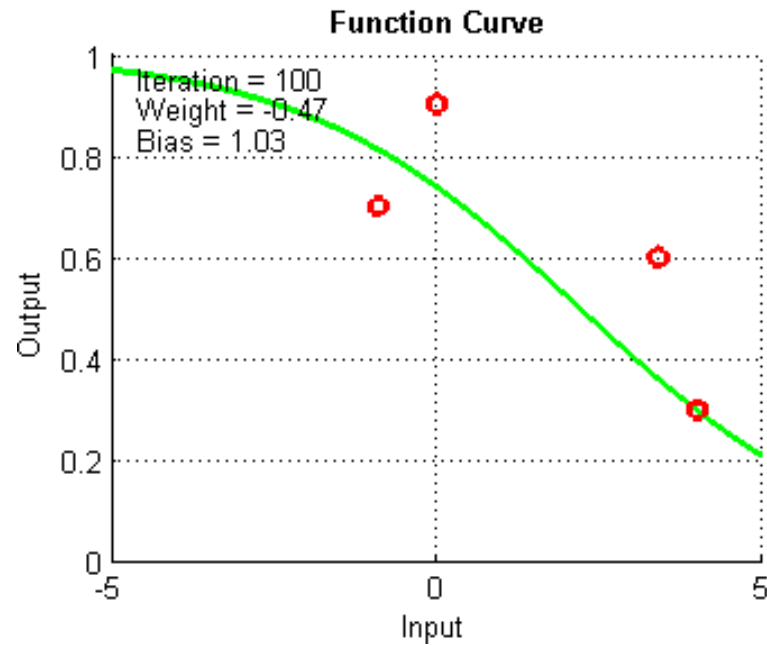
Big Data : une affaire de traitement et non de volume (tribune)

La Rédaction, 26 novembre 2015, 9:32

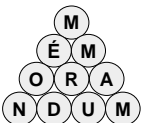
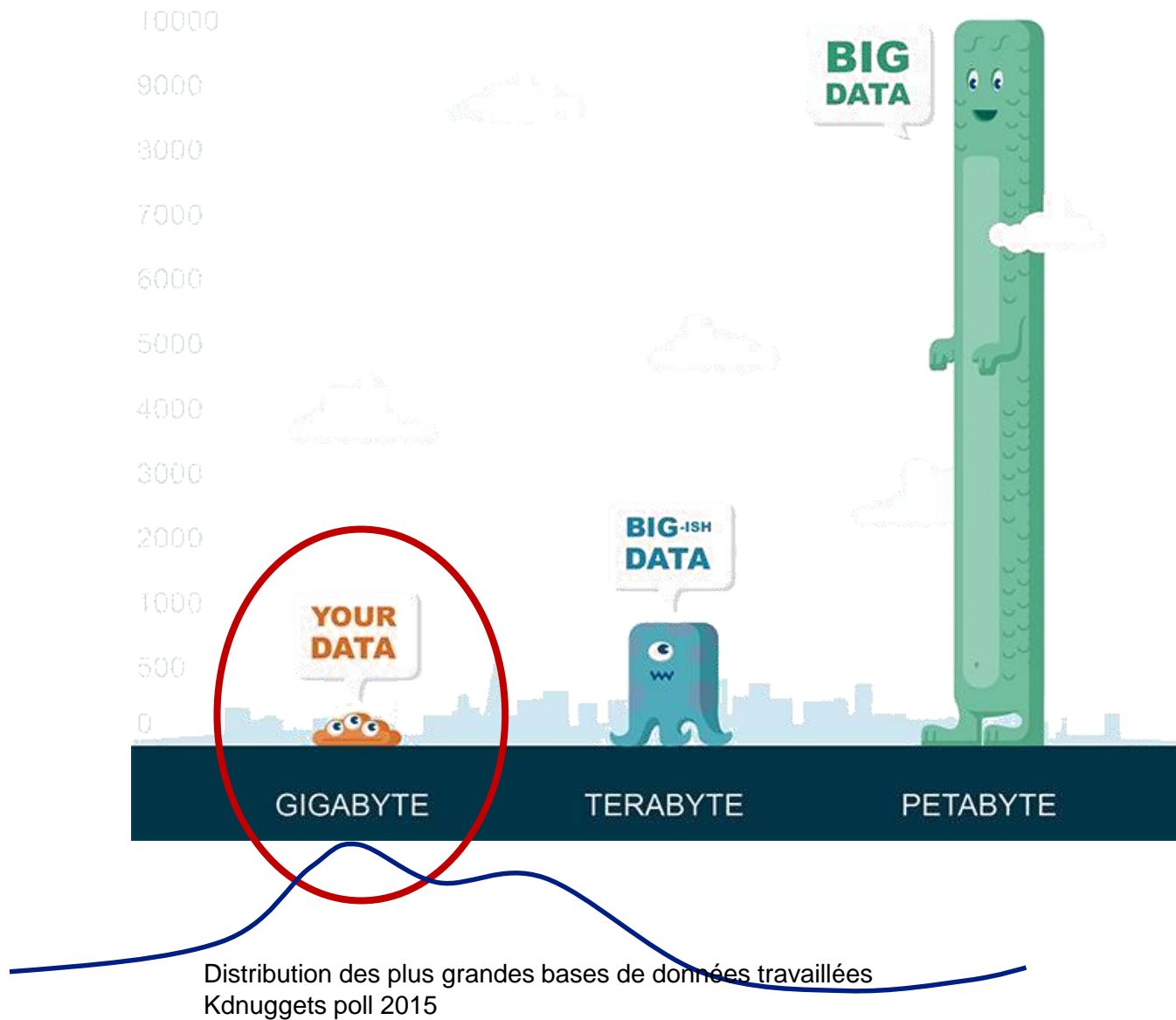


Le terme de Big Data a pu laisser penser qu'il s'agissait d'amasser de grands volumes de données. Mais la valeur de ces applications réside plutôt dans la teneur des traitements et la capacité à évaluer la pertinence des informations, analyse Claire Leroy, du CXP.

Apprentissage



Valoriser ces données : une affaire de traitement et non de volume



Objectif : partager et illustrer 3 convictions

1

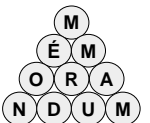
« La valeur n'attend pas le nombre des octets »

2

« Exploiter la donnée est un projet métier, pas un projet informatique »

3

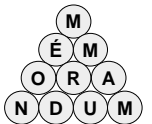
« Fail fast & learn »



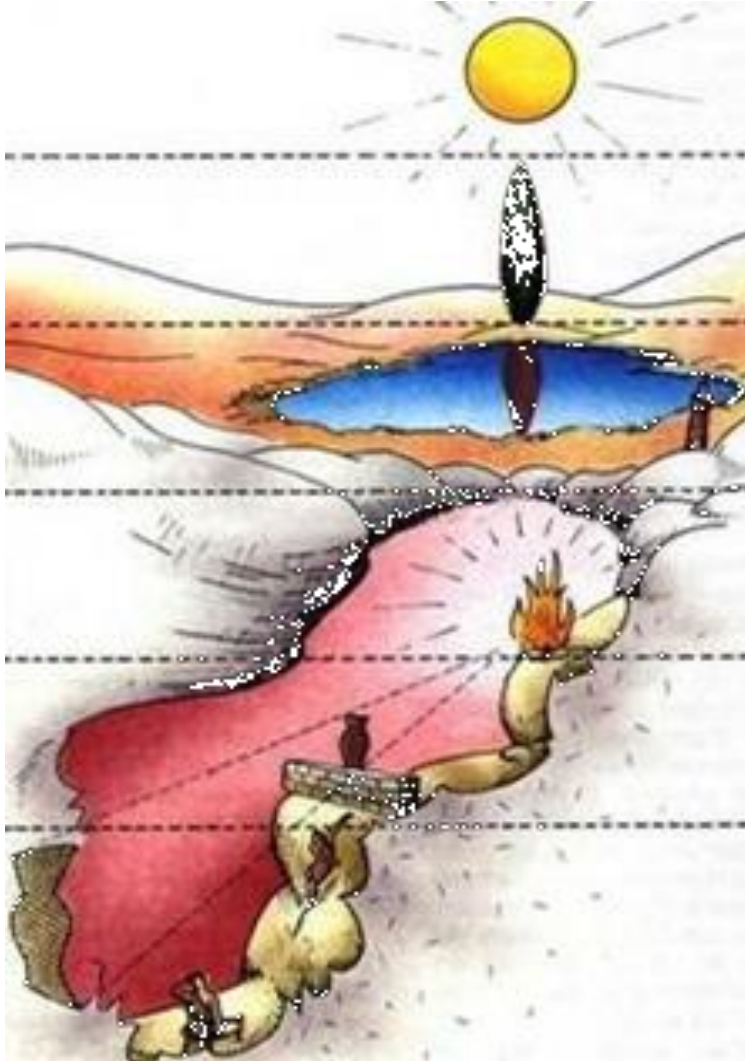
Les données les plus importantes sont dans les bases de données des métier



La réponse n'est pas un datalake !



La donnée n'est que l'ombre de la réalité : la manipuler et l'interpréter mérite une supervision métier

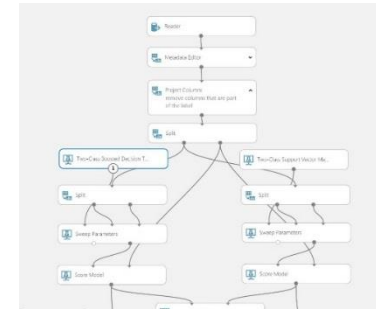


La barrière à l'entrée des techniques d'analyse se réduit progressivement

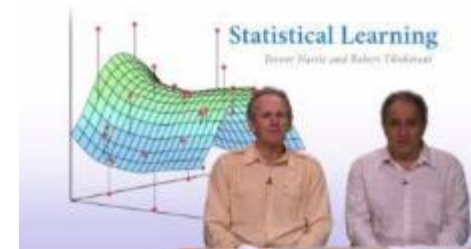
Les algorithmes de machine learning sont packagés et en libre accès dans différents environnements open source (R / Python)



Des interfaces graphiques se développent (Dataiku, Azure ML, ...) : possibilité de pratiquer sans coder

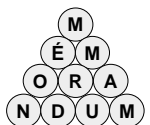


Des formations nombreuses et de bon niveau disponibles en ligne (MOOC)

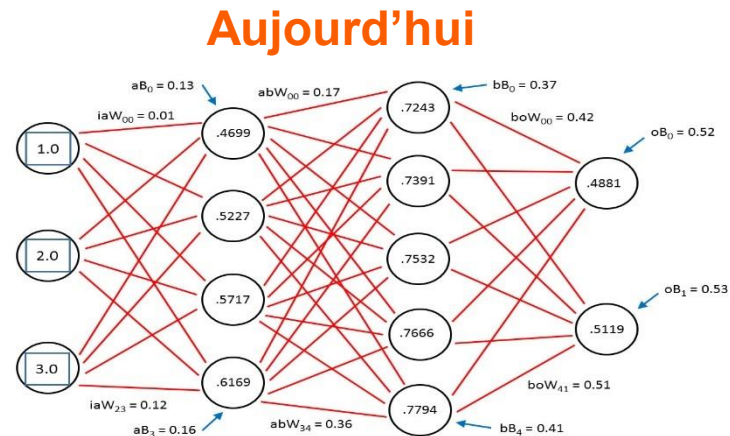
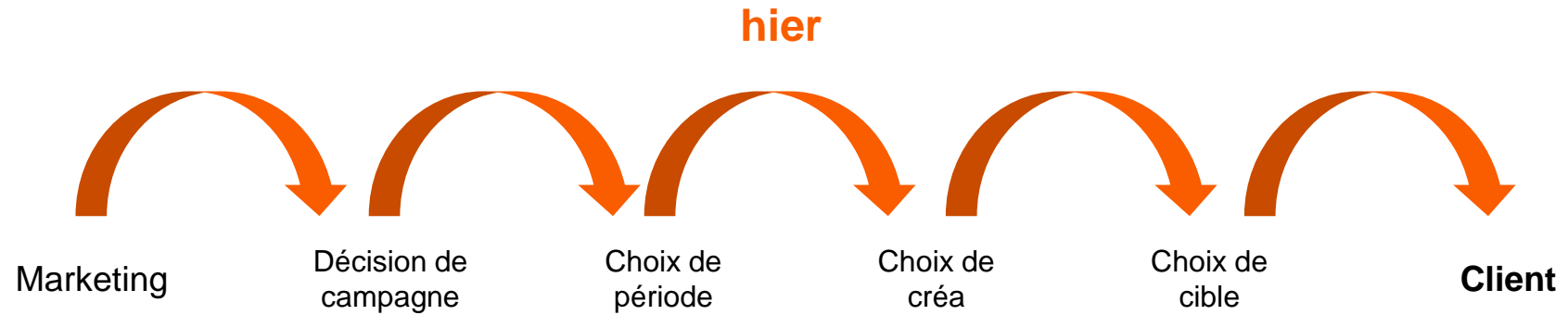


Une communauté active et ouverte

...



L'exploitation de la donnée en mode « big data » change profondément le processus de décision



Un circuit de décision transformé → un besoin de cadrage accru en amont

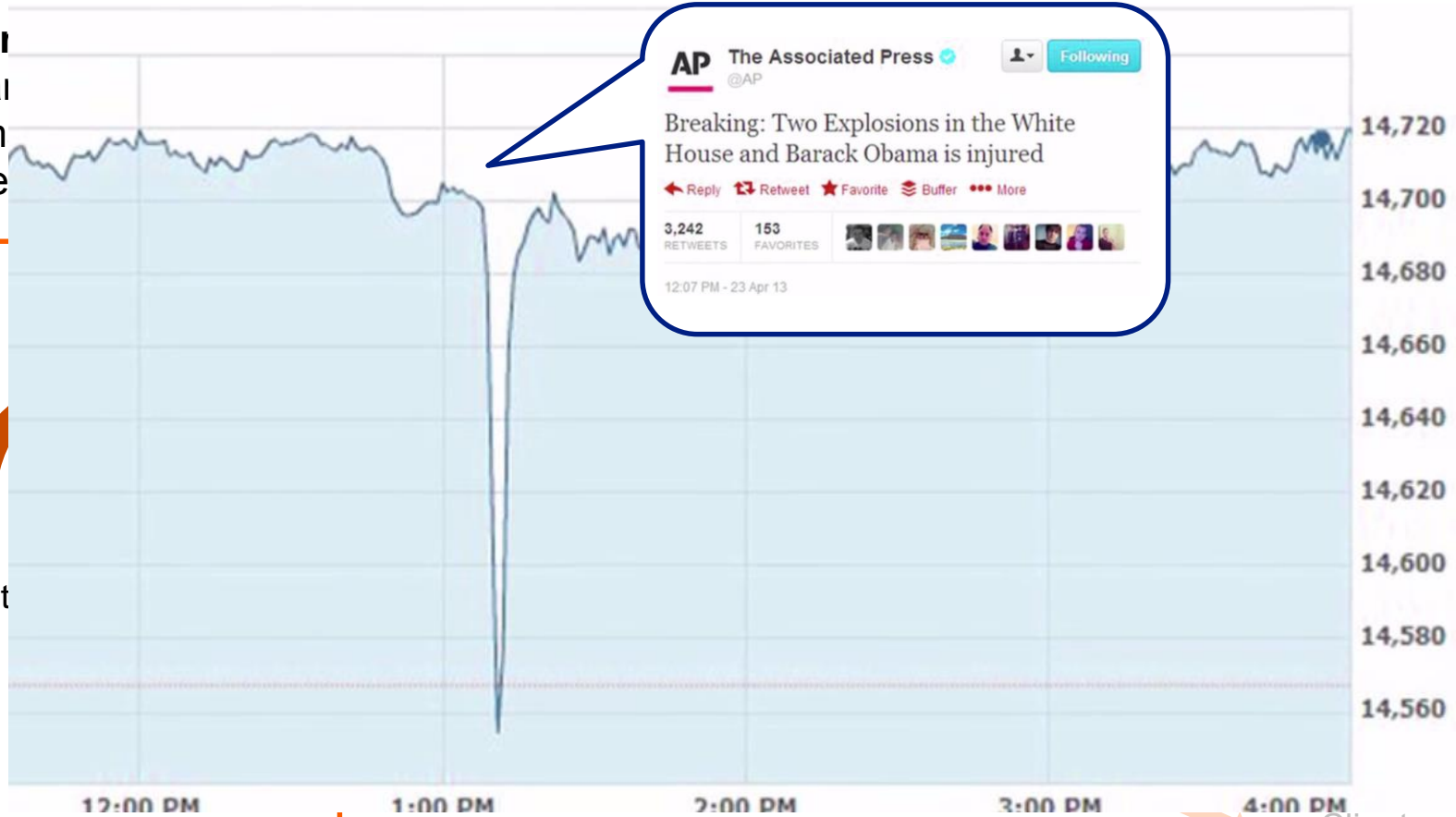
Besoin

→ mai

l'im

→ Pre

Market



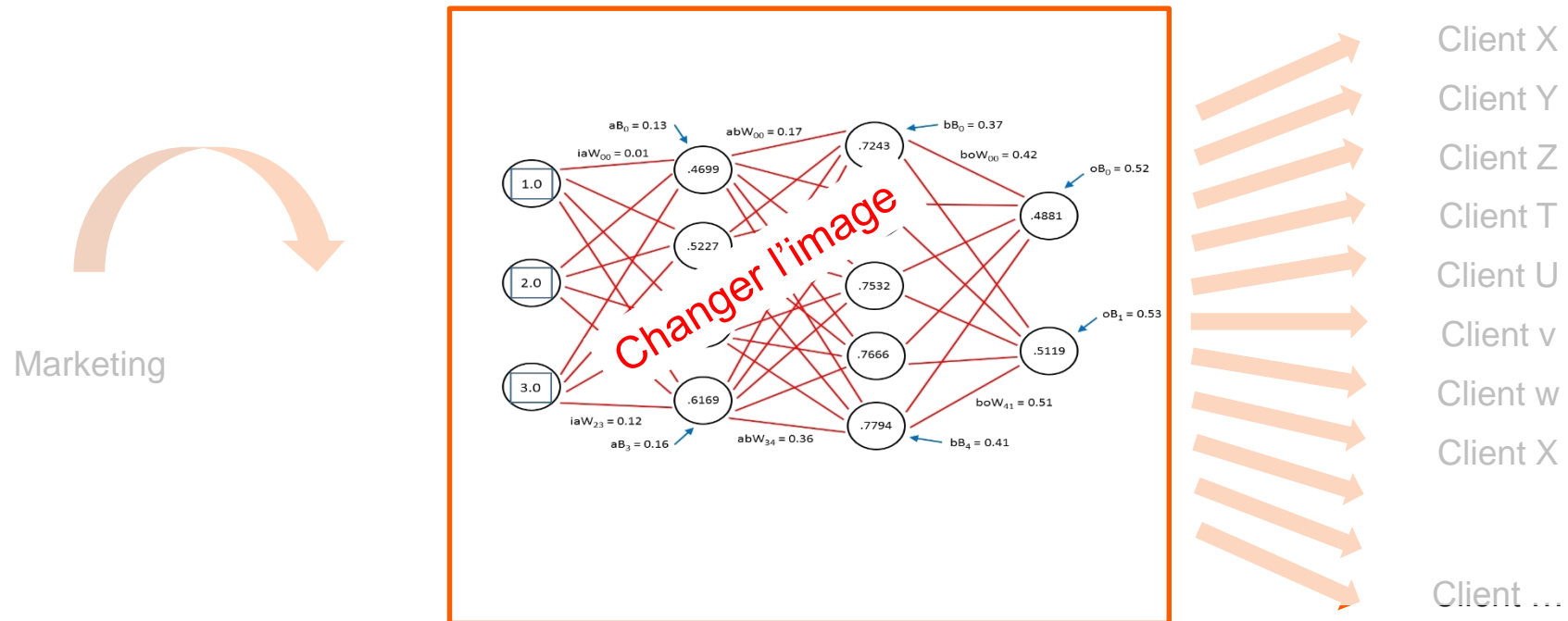
ent

Client ...

Un circuit de décision transformé → une culture data essentielle

Penser données

- Savoir identifier quand on manque de données, et qu'on a des données biaisées (exemple score crédit) et prendre les moyens de les acquérir : phase d'apprentissage, nouveaux services, crowd sourcing, open data, 2nd party, ...
-

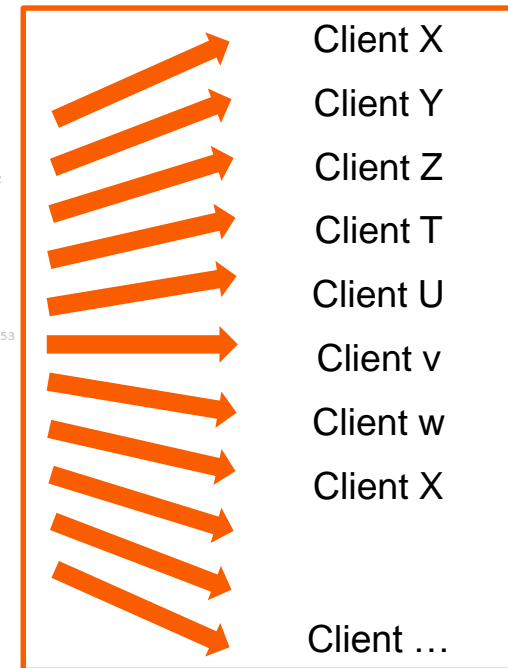
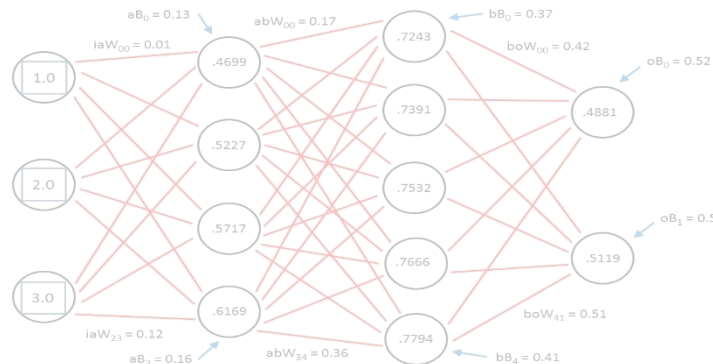


Un circuit de décision transformé

→ un pouvoir accru de ceux qui sont au contact du client

Pouvoir des vendeurs et des conseillers

- Les acteurs en bout de chaine (vendeurs, conseillers téléphoniques, ..) ont un pouvoir croissant : plus informé sur leurs clients, plus



ARRÊTEZ !! LA SEGMENTATION CLIENT NE SERT À RIEN !

OLIVIER FABRE, SOCIALINTELLIGENCE

25 MAI 2016

Dans le cadre de projets de transformation digitale, de vision 360° du client ou d'approches multicanal, nos clients nous demandent régulièrement de segmenter leur base clients. L'idée derrière cette demande est de pouvoir cibler simplement lors de campagnes marketing, des ensembles de clients homogènes (Personas), ayant potentiellement des goûts/appétences similaires et d'adapter au mieux leurs messages à leurs marchés.

Objectif : partager et illustrer 3 convictions

1

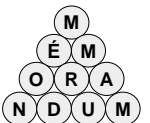
« La valeur n'attend pas le nombre des octets »

2

« Exploiter la donnée est un projet métier, pas un projet informatique »

3

« Fail fast & learn »



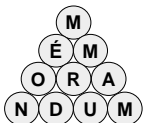
La difficulté principale est souvent plus dans la question que dans la réponse

Fail Fast

Trouver la bonne question à poser à un data scientist est aussi difficile que de choisir les questions qu'on pourrait poser à un Génie

- Espace infini de possibilités, impossibles à hiérarchiser, difficiles à formuler précisément

Heureusement avec un data scientist vous pouvez (et même devez) tester vos questions et les adapter au fur et à mesure que vous obtenez des réponses.

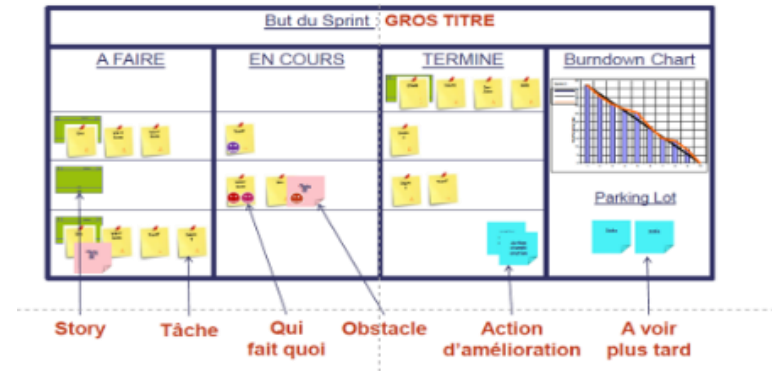


Exemple de leviers : Structurer les méthodes de travail et les livrables

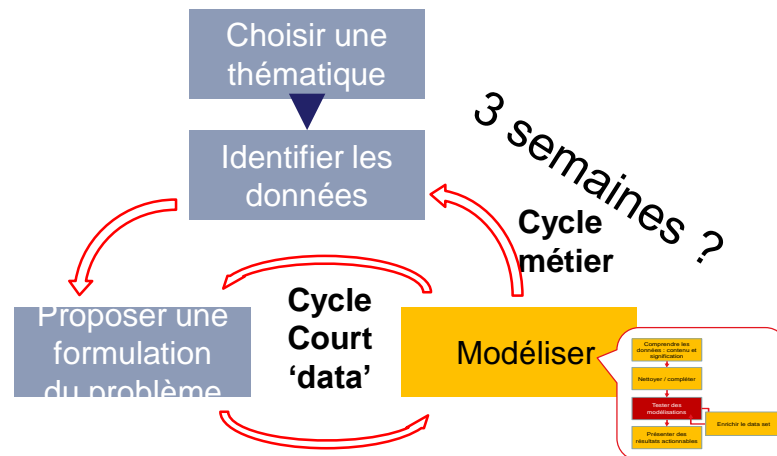
Fail Fast

Dérouler une structure de travail adaptée des méthodes agiles / scrum

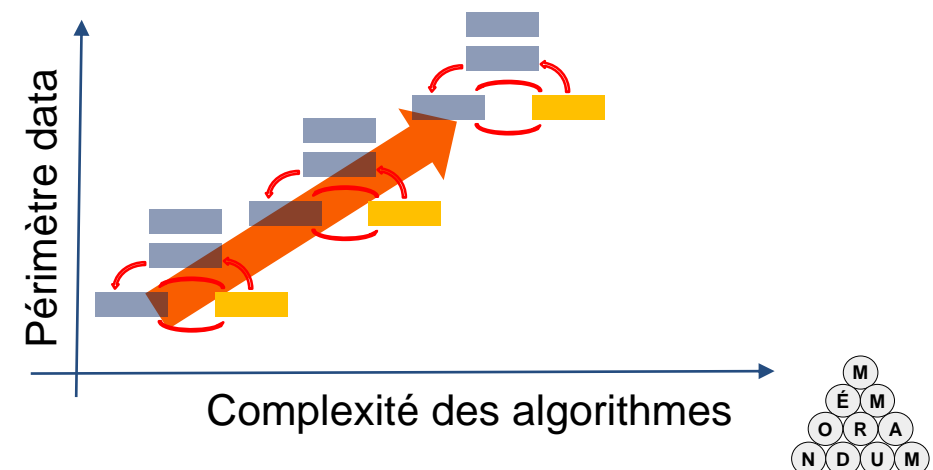
- Rend lisible l'avancée des travaux



Une démarche itérative qui fait progresser la question en même temps que la réponse



Des itérations qui partent de petits volumes et modèles simples



Extraits du livre blanc de bpifrance pour promouvoir le big data auprès des PME

Fail Fast

Les technologies de modélisation ont occupé financièrement de l'innovation montre que l'historique est faible pour eux se les fonds concept que la n bien au marketin chain.

Au fond, les modèles et algorithmes

Aremus & Associés
CAPITAL-INNOVATION

bpifrance

80% des dirigeants seront les figurants de demain.

Et vous ?

Oui. Votre business est déjà concerné.

La disruption sera vecteur de compétitivité pour les entreprises prêtes à se réinventer. A condition de faire vite et de définir une vision claire, sous peine de décroître puis sombrer.

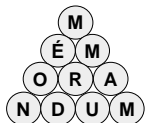
amen-
gitale
la fois
s des
et en
pour
rt-up.
logies
ement
énieur,
les et
allant
ssage
nées,
utili-
ile et

“ l'algorithme intime du client. De plus en plus temps réel : la disruption s'apprend. mais plus profondément encore d'entrer dans une transformation de l'expérience client sur des produits et services qui peuvent désormais bénéficier d'une évolution constante.”

Modèles, données et algorithmes

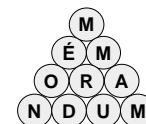
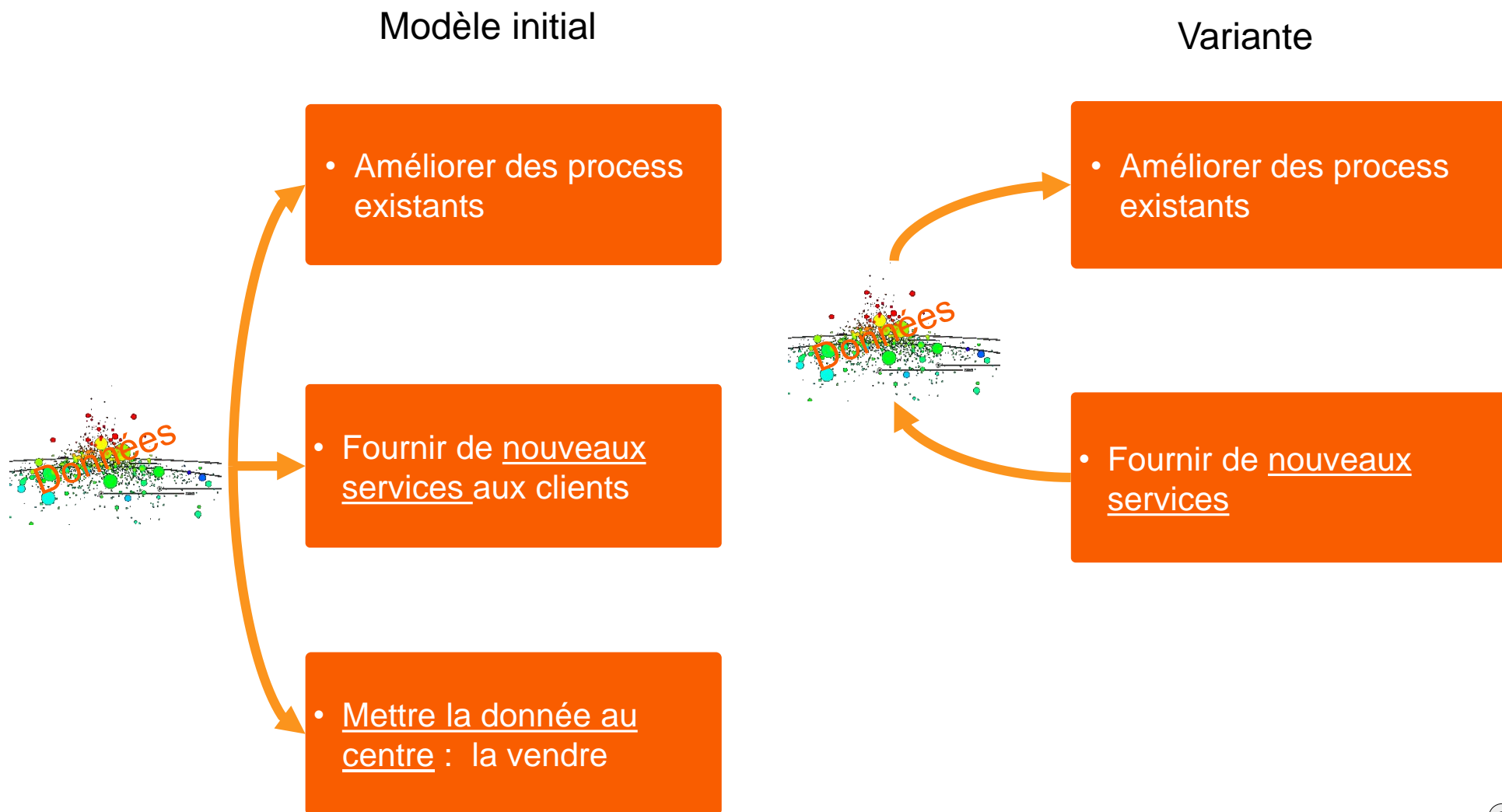
Les nouvelles frontières du numérique

Par Etienne de Rocquigny (Aremus & Associés)



D'une stratégie à l'autre : inverser la logique proposer des services qui produisent des données

Fail **Fast**



Si vous ne valorisez pas vos données, d'autres le feront à votre place

Fail **Fast**

Le monde change

Exemple de la base adresse nationale

Au début : un actif réservé et facturé



**Aujourd'hui un bien commun
(15 avril 2015)**



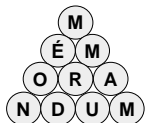
OpenStreetMap France
Cartographions le monde rue après rue...

Un nouvel acteur reconstitue la base en open source

La loi peut changer

« Je souhaite créer un nouveau statut de données, les données d'intérêt général, qui sont à l'origine des données commerciales qui appartiennent à des entreprises mais qui, si elles étaient ouvertes, pourraient bénéficier à la communauté et à l'intérêt général. »

Axelle Lemaire (2015)



Learn ?

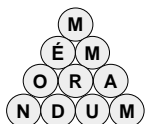
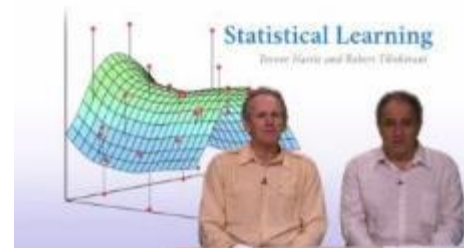
Beaucoup de leviers à disposition

Principe lancer une formation interne s'appuyant sur un mooc, ouvert à tous, pouvant être réalisé en partie sur le temps de travail

- Tout le monde apprend un peu
- Quelques uns vont au bout
- Certains sont bons et veulent continuer

L'entreprise

- a passé son message interne sur l'importance qu'elle accorde aux données
- Donne sa chance à tous
- Fait émerger un pool de salariés motivés .. Et reconnaissants



Développer et animer la compétence data

Learn

Qui se pose la question ?

De grandes entreprises avec des équipes éclatées ?

Des entreprises qui ont un ou deux spécimens ?

Des entreprises qui veulent faire émerger des talents ?

Qui ne se la pose pas ?

Des entreprises qui achètent de la magie



Objectif : partager et illustrer 3 propositions

1

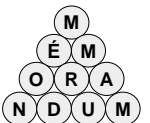
« La valeur n'attend pas le nombre des octets »

2

« Exploiter la donnée est un projet métier, pas un projet informatique »

3

« Fail fast & learn »



Cette idée était devenue une évidence. L'économie des données (des datas) était la clé de développement de toutes les entreprises liées au numérique et même de l'avenir de cette industrie. Ceux qui n'investiront pas suffisamment dans des outils pour recueillir, analyser en temps réel et exploiter des masses de données sont condamnés. Mais les choses sont en train de changer très vite [explique Quartz](#).

Les datas ont toujours une grande valeur économique pour qui sait les exploiter mais elles sont aussi en train de devenir un très grand risque, juridique et d'image, pour les sociétés qui les utilisent et les stockent. Cela signifie *«que dans l'avenir, les sociétés de données ne vont plus les détenir. Elles vont juste gérer leur flot»*.

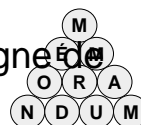
Les citoyens consommateurs n'ont pas confiance

Cela fait maintenant près de trois ans qu'Edward Snowden a révélé l'ampleur de la surveillance dont nous faisons tous l'objet et les citoyens de par le monde ont de plus en plus conscience des dangers de la récupération et de l'utilisation incontrôlées de leurs données personnelles. Selon une étude récente [citée par la Harvard Business Review](#), 97% des consommateurs aux Etats-Unis, au Royaume-Uni, en Allemagne, en Inde et en Chine craignent que les entreprises et les gouvernements utilisent leurs données personnelles à leur détriment.

La confiance est un élément essentiel de la relation entre les consommateurs et les entreprises. Mais il ne s'agit plus seulement de la confiance dans la capacité d'une société à fournir des produits et des services de qualité, mais aussi dans son éthique et dans sa façon de gérer les données qu'elle récupère sur ses clients. Une [autre étude internationale](#) montre que les trois-quarts des consommateurs ne font pas confiance aux entreprises sur la gestion de leurs données personnelles.

Obtenir le consentement

Cette défiance et cette crainte ont d'ores et déjà un coût considérable. Selon des données qui remontent à l'année 2015, [près de 200 millions d'ad-blockers](#) étaient installés dans le monde privant les diffuseurs en ligne de plus de 20 milliards de dollars de revenus par an. Même les utilisateurs de Facebook prennent parfois des mesures radicales pour protéger leurs données personnelles et un million d'entre eux accéderaient au réseau



Let's just say it: In the future, successful data companies will not own any data. The data economy is the engine of growth for all digital businesses. However, the landscape is changing. We are heading for a data crunch. Data has huge economic value but it is becoming businesses' biggest liability. As a consequence, in the future, data companies will not own data. They will just manage flows of it.

Now that we are approaching three years since the Edward Snowden revelations, citizens are more conscious of their personal data. This is true not just from a government surveillance perspective, but also when it comes to corporate surveillance. Recent research shows that 97% of consumers in the UK, US, Germany, India and China [are concerned that businesses or governments will misuse their data](#).

Trust is becoming an important currency in the business world. It's not just trust in an organization to deliver quality products and services; it's the trust that customers place in an organization to responsibly handle sensitive information. Another international study shows that [75% of consumers distrust brands with their personal data](#).

This distrust and sense of anxiety about how personal data is being handled is affecting how people behave. The latest data from mid-2015 shows that [198 million ad-blockers were installed worldwide](#). This is costing businesses \$22 billion per year. Even users of Facebook are taking radical steps to protect their personal identity, as it was recently reported that [a million people access Facebook on the "dark web"](#) every month.

It's not just digital citizens who are taking action. Regulators are, too. The EU council recently passed the General Data Protection Regulation (GDPR), which has severe fines for non-compliance: 4% of annual revenue or €20 million Euros, whichever is highest. The GDPR contains a number of new measures that will change the way businesses handle data.

Consent is a big one. Businesses will need to obtain consent for each specific use of that data. So the method of *"collect it now and figure out what to do with it later"* will become a high-risk strategy.

The other major change will be data portability. Businesses will have to provide their customers with a copy of their data. This will make data fluid and put the citizen in control of their personal data. In other words, not only

