

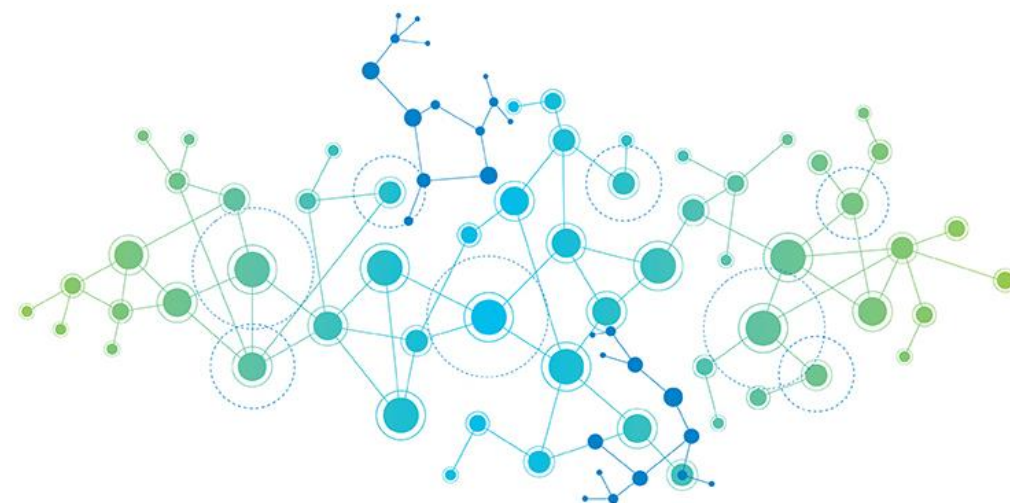
# Rendre les données génératrices de nouveaux services en santé : conception de modèles et gestion d'anomalies

Antoine Bordas

Doctoriales 2022

Mines ParisTech, PSL Université,

Centre de Gestion Scientifique, i3 UMR CNRS 9217



# Table of contents

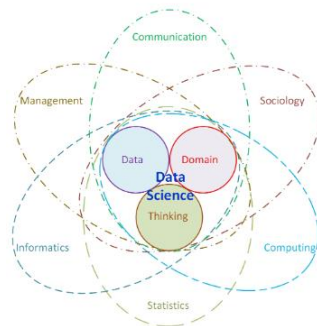
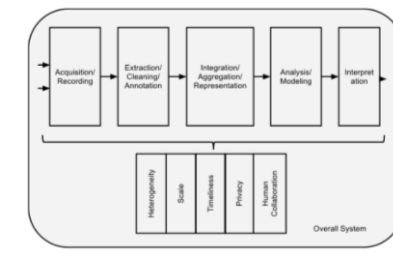
- 01 **Revue de littérature** : conception de modèles et gestion d'anomalies pour la valorisation des données
- 02 **Méthodologie** : éclairer le couplage entre modèles et anomalies
- 03 **Premiers résultats** : étude généalogique et un projet en externe
- 04 **Perspectives** et travaux à poursuivre

# Contexte et problématique : la valorisation des données dans un monde digital



Partage massif et **phénomène de « Big Data »** de plus en plus discuté (Marr, 2016), tant par les entreprises que les chercheurs, et **amené à se renforcer** (Sestino et al., 2020).

Parmi les caractéristiques du « Big Data », se trouve la notion de **valeur**, qui est la conséquence d'un long **processus « d'extraction de la valeur »** (Labrinidis and Jagadish, 2012).

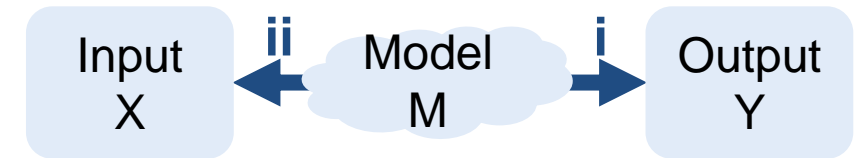


Les nombreux cas d'usages montrent le rôle majeur, dans ce processus de valorisation, de la nouvelle discipline qu'est **la data science** (Escoufier et al., 1995; Skiena, 2017).

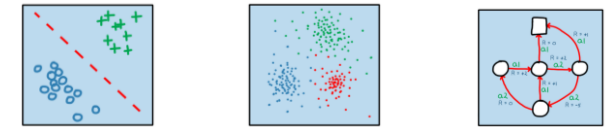
**Comment rendre les données génératrices de nouveaux services en santé ?**

# Revue de littérature : concilier modèles et anomalies pour la valorisation

- Les données sont vecteurs de valeur selon deux stratégies :
- comme point de départ pour la **création de nouvelles connaissances** (Fosso Wamba et al., 2015),
  - conception de la **donnée pour un usage** (Trabucchi and Buganza, 2018).

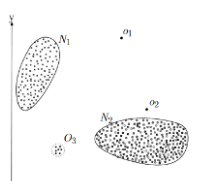
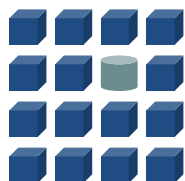


La data science est le dénominateur commun de ces deux approches, permettant la **conception d'un triplet optimal (X, M, Y)**. Ce paradigme est classique de cette discipline et s'appuie sur une **stratégie dominante** (James et al., 2013 ; Osman Kazakçı, 2015).



3 grandes catégories de modèles :  
apprentissage supervisé, non  
supervisé et par renforcement.

Panneau stop lu comme  
limitation de vitesse.



Des anomalies ?!

Les **anomalies sont omniprésentes** dans la littérature en mathématiques, tant pour la détection (Chandola et al., 2009) que la création (Eykholt et al., 2018). Mais cette littérature ne fournit que peu de clés de lecture et **n'explicite pas la logique d'action sous-jacente** (Li et al., 2015).

# Comment rendre les données génératrices de nouveaux services en santé ?

**QR 1 :** Comment modéliser et optimiser la conception de modèles dans une situation riche en données ?

- Comment modéliser le processus même de conception de modèles ?
- Quelles sont les caractéristiques à rechercher dans un modèle en data science ?

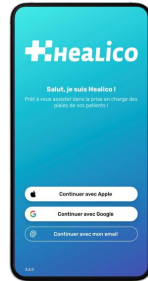
**QR 2 :** En parallèle de cet exercice de modélisation, quels rôles jouent les anomalies inhérentes aux modèles dans la définition de ces derniers ?

- Dans quelle mesure les anomalies ont-elles un rôle de catalyseur dans la conception de modèles ?
- Quelles sont les méthodes, processus et compétences nécessaires à une « bonne » gestion des anomalies ?

**QR 3 :** Quelles sont les méthodes et organisations assurant un couplage entre modèles et anomalies permettant la valorisation des nouvelles données ?

- Comment anomalies et modèles cohabitent lors de la conception de ces derniers ?
- Comment la conception de modèles et la gestion des anomalies permet de valoriser les données ?

# Méthodologie : éclairer le couplage entre modèles et anomalies



## Terrain principal :

- Recherche intervention chez Urgo Médical, dans l'équipe digitale en charge du projet Healico, rattaché au directeur transformation digitale et au data scientist.
- Healico, application mobile à destination des professionnels de santé pour la prise en charge des plaies.



## Etude de cas complémentaires :

- Saint-Gobain via l'encadrement d'un projet POC sur une problématique data science à partir d'images d'entrepôts.
- Autres cas de projets de data science avec des partenaires de la chaire TMCI.



## Généalogie du projet Healico :

- Comprendre les différentes techniques de gestion à l'œuvre dans un projet de data science
- Mettre en avant le rôle et l'importance de la gestion des anomalies dans ce contexte

## Méthodologie :

- Entretiens semi-directifs avec les anciens chefs de projet et de produits qui se sont succédés depuis le début
- Collection de documents internes pour appuyer les propos recueillis

## Apprentissages :

- Mise en lumière de l'existence d'une pluralité de modèles, cohabitant pour atteindre l'optimalité de (X, M, Y).
- Les modèles sont en fait des relais pour d'autres, le processus de passage de l'un à l'autre étant contrôlé par les anomalies.

# Premiers résultats : étude généalogique et un projet en externe (2/2)



## Projet Saint-Gobain :

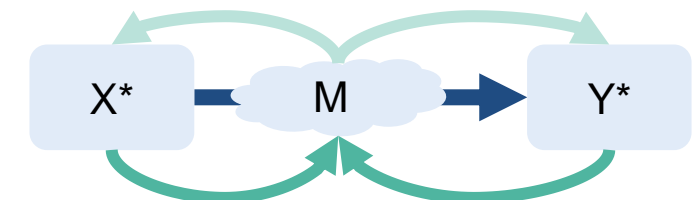
- Comprendre le processus à l'œuvre lors de la conception de modèles et mettre au jour les compétences nécessaires
- Mettre en avant des grandes caractéristiques des modèles de data science

## Méthodologie :

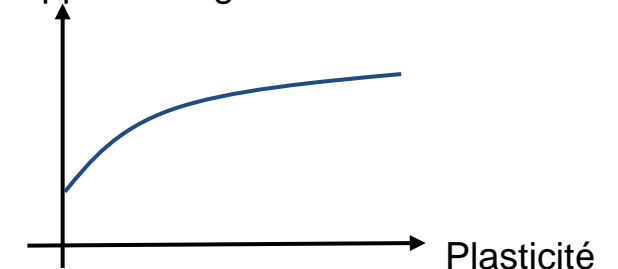
- Encadrement d'étudiants dans le cadre d'un projet du trimestre POC
- Suivi du processus mis en place par les étudiants pour développer un modèle de data science à partir de deux modèles génériques fournis

## Apprentissages :

- Mise en avant d'une faculté de rétroaction des modèles, génératrice d'apprentissages et clé pour la conception d'un triplet  $(X, M, Y)$  optimal.
- Explicitation de cette faculté d'apprentissage comme fonction croissante de la plasticité du modèle de départ.



Génération  
d'apprentissages





## Objectifs



- Approfondir les premières modélisations menées (double rôle des modèles ? anomalies comme variables de contrôles ?)
- Etude du nouveau paradigme de « Theory-Guided Data Science » (Karpatne et al., 2017)
- Approfondir les connaissances générées (caractérisation des modèles ? Couplage à d'autres sources d'apprentissage ?)

## Prochaines étapes



- Poursuite des analyses et confrontation à l'expérience via divers projets internes de data science
- Analyse du couplage physique/data (nouveau processus de conception ? conditions d'existence ? Implications organisationnelles ?)
- Reproduction d'études de cas similaires dans d'autres secteurs industriels
- Rédaction d'un papier

# Perspectives et travaux à poursuivre : expérimentation sur un modèle prédictif de la cicatrisation

## Objectifs



- Etudier les principes organisationnels de la conception de modèles pour la valorisation de nouvelles données
- Mettre au jour les grandes stratégies de gestion d'anomalies sous-jacentes
- Modéliser le rôle de l'écosystème dans la conception de modèles
- Concevoir un modèle prédictif de la cicatrisation des plaies

## Premières études



- Début de revue de (mathématiques) littérature (en médecine et sur la modélisation de la cicatrisation)
- Première étude des contraintes écosystémiques liées aux modèles prédictifs pour les services en santé
- Interview de professionnels de santé

## Prochaines étapes



- Travail sur la définition du modèle mathématique de la cicatrisation (collaboration avec E. Decenciere ?)
- Collecte de données et modélisation des principes organisationnels de la conception de modèle
- Poursuite de l'étude du rôle de l'écosystème (utilisateurs, réglementaire...)

# Perspectives et travaux à poursuivre : la santé digitale et apports d'un cas historique

## Objectifs



- Comprendre les problématiques, contraintes et opportunités de la santé digitale
- Proposer des méthodes permettant de répondre à la question de la valorisation des données en santé

## Premières études



- Revue de littérature et veille sur l'écosystème de la santé digitale
- Identification d'un cas historique (la Packard Instrument Company) intéressant à mettre en parallèle
- Premières modélisations duales des similarités entre e-santé et la PIC

## Prochaines étapes



- Approfondissement des apprentissages du cas historique
- Interview d'experts en santé digitale
- Etude de cas dans des entreprises du secteur de la e-santé

# Antoine Bordas

Mines ParisTech, PSL Université, Centre de  
Gestion Scientifique, i3 UMR CNRS 9217



Merci pour votre  
attention !