

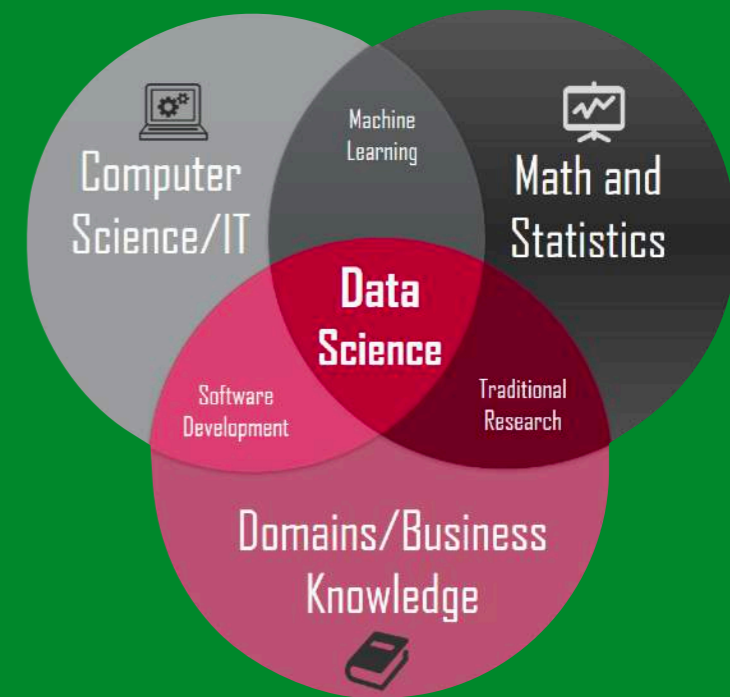
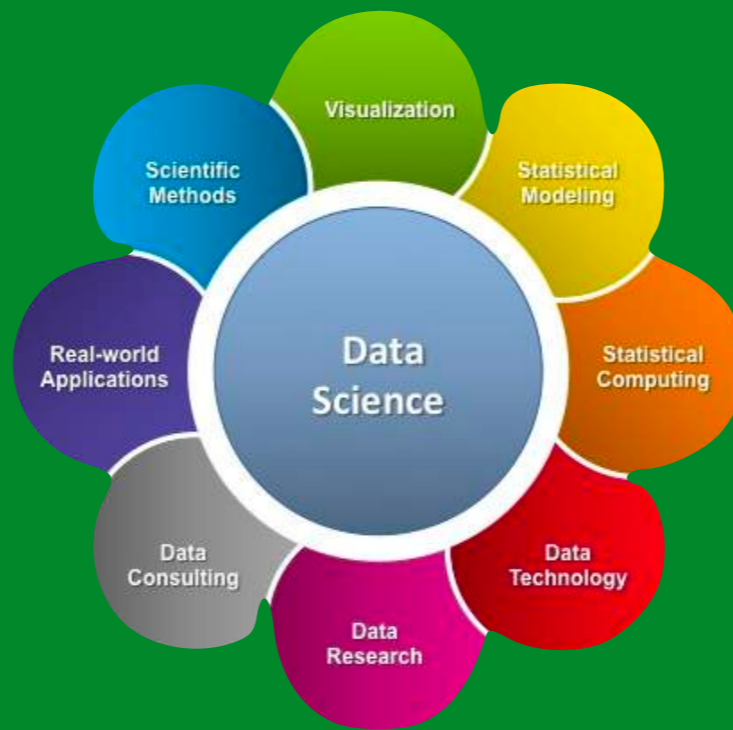
# des données à la décision

Sophie Demasse

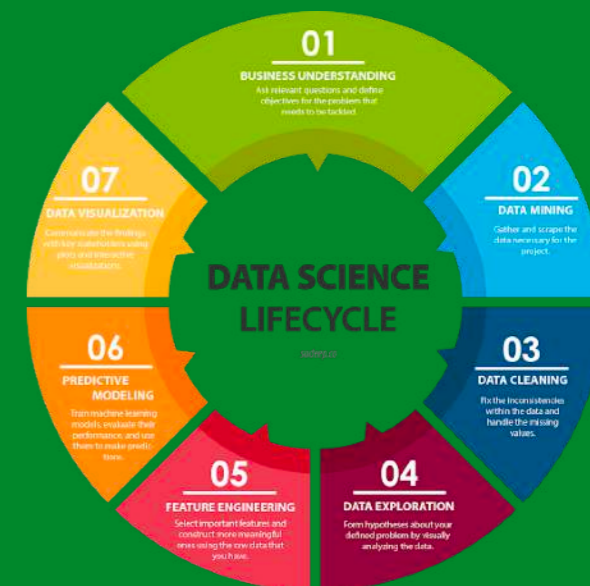
# 1

régression linéaire, etc.

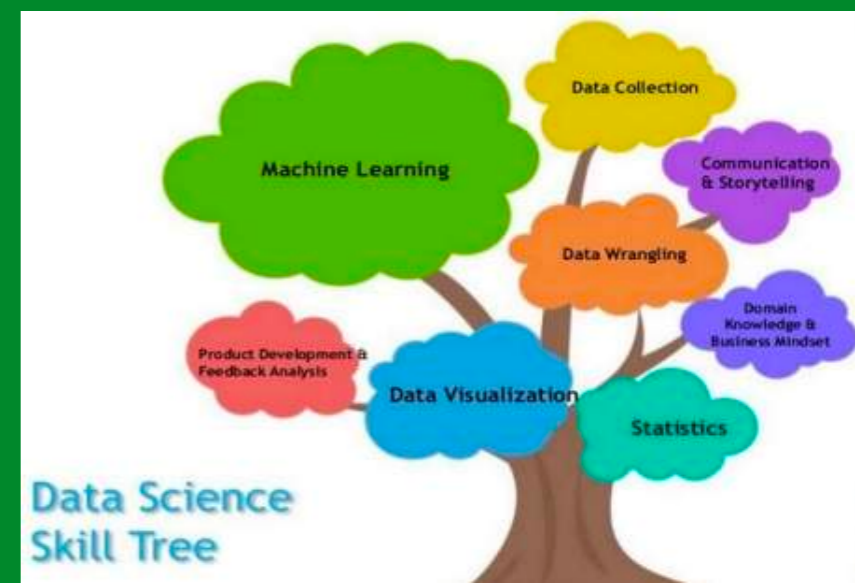
- la science des données est un creuset
- elle emprunte à l'apprentissage automatique mais ne s'y résume pas
- l'aide à la décision emprunte à la science des données mais ne s'y résume pas
- automatisations, la théorie et la pratique



# data science ?



*Despite the excitement around “data science”, “big data”, and “analytics”, the ambiguity of these terms has led to poor communication between data scientists and those who seek their help. (Harris, Murphy, and Vaisman 2013)*



**le traitement automatique des données**  
est l'essence de **l'informatique** et **l'extraction de**  
**connaissances** l'une de ses premières applications

**data science ?**

**l'extraction de connaissances à partir des données**  
est l'essence de la **statistique** et **le traitement**  
**automatique** en est une branche appliquée

- *(Tuckey, 1962) "data analysis": les statistiques appliquées comme une science, au-delà de la statistique inférentielle formelle*
- *(Chambers, 1993) "inclusive concept of learning from data": accent sur la préparation et la présentation des données*
- *(Breiman, 2001) prédiction plutôt que compréhension, algorithmes (d'apprentissage) plutôt que modèles mathématiques*
- *(Donoho, 2017) 1. collecte, préparation, exploration 2. représentation, transformation 3. calcul 4. modélisation 5. visualisation, présentation 6. inception*

**Donoho D. (2017) "50 Years of Data Science"**

# augmentation, diversification, progrès des capacités de:

- ◆ stockage
- ◆ acquisition
- ◆ communication
- ◆ calcul (architecture)
- ◆ calcul (algorithme)
- ◆ développement

ex: architecture distribuée

ex: portable

ex: internet

ex: gpu

ex: NN backpropagation

ex: R, common task framework

## évolution conjuguée des

statistiques et optimisation  
apprentissage automatique  
systèmes distribués  
contextes applicatifs

*Machine Learning!*  
*Big Data!*

données, connaissances, décisions

# ex: google map

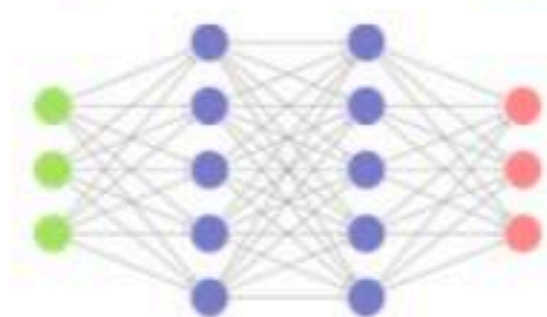
données

prédiction

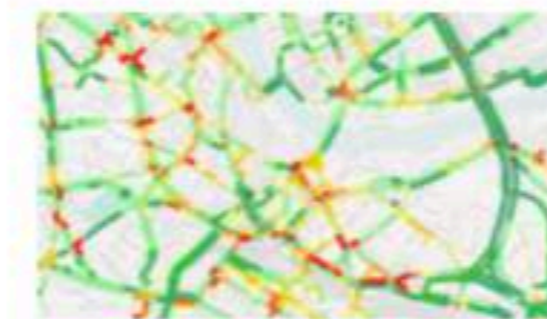
décision



Data



Predictive model



Predicted travel times



Routing algorithm



Shortest path



# données ?



## DÉFINITIONS

---

**donnée**  nom féminin

1 Ce qui est donné, connu, déterminé dans l'énoncé d'un problème.

*Les données du problème.*

2 Élément qui sert de base à un raisonnement, de point de départ pour une recherche. *Données statistiques.*

3 **INFORMATIQUE** Représentation conventionnelle d'une information permettant d'en faire le traitement automatique. *Données*

*alphanumériques. Banque, base de données. — Ces informations.*

→ **ANGLICISME** data.

# donnée brute

- ◆ grandeur mesurée ou calculée
- ◆ image, son, texte, code, signal, flux
- ◆ (description, contexte, temporalité)

# traitement automatique

- ◆ circuit analogique, système optique
- ◆ système numérique: mathématique/informatique

# pour quoi ?

- ◆ communiquer
- ◆ connaître
- ◆ percevoir
- ◆ comprendre
- ◆ généraliser
- ◆ présager
- ◆ décider

# pour qui ?

- ◆ humains
- ◆ machines

ex: militaire, politique,  
business, scientifique,  
clinicien, robot, logiciel

développements parallèles en ingénierie et en intelligence artificielle

# aide à la décision: continuum

systeme d'information partiellement ou totalement automatisé

systeme d'aide à la décision,  
systeme d'information géographique,  
systeme d'aide au diagnostic clinique,  
business analytics,  
systeme expert,  
systeme à base de connaissances,  
workflow en programmation quantitative,

prescription  
décision

prédiction

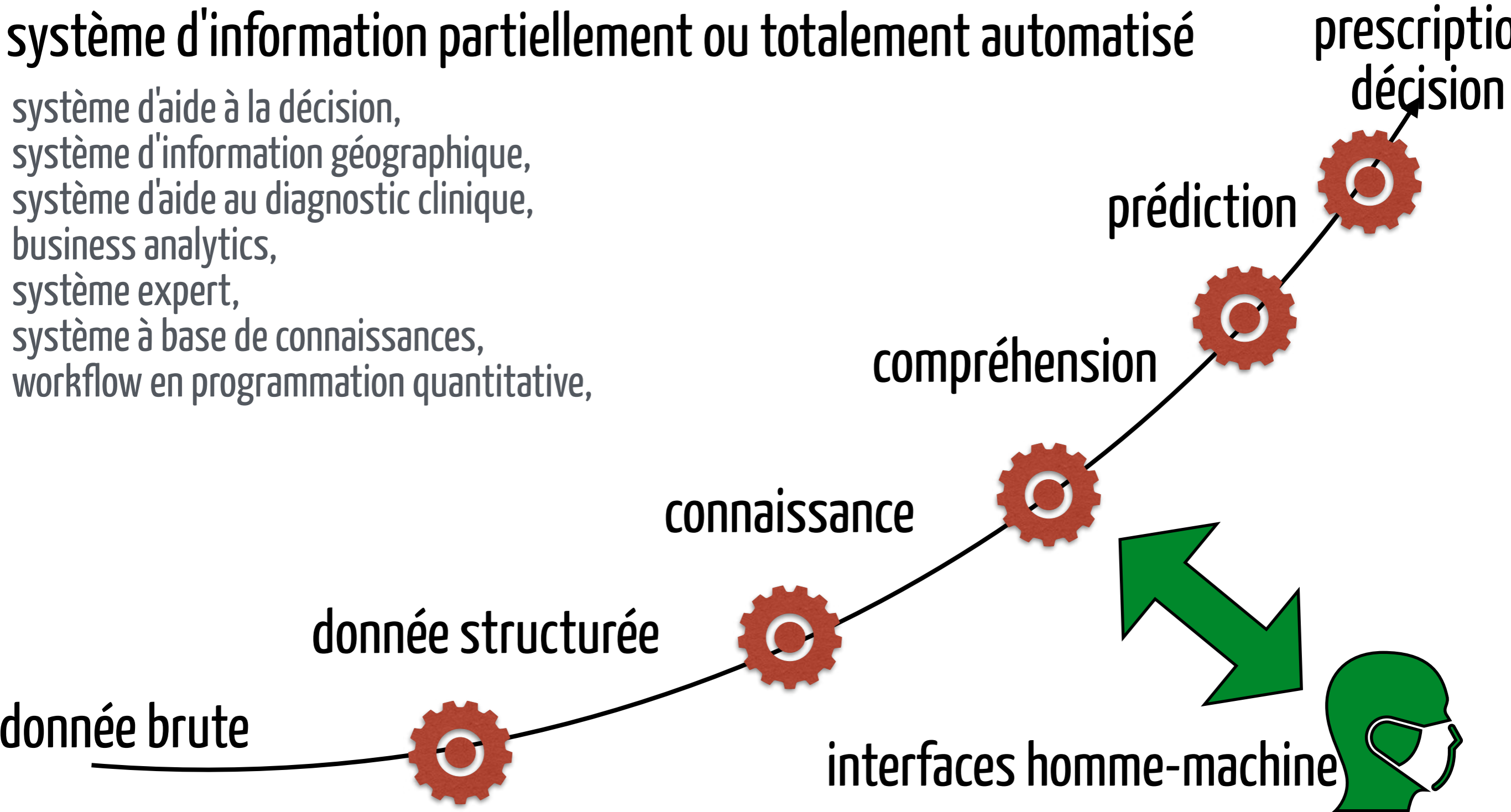
compréhension

connaissance

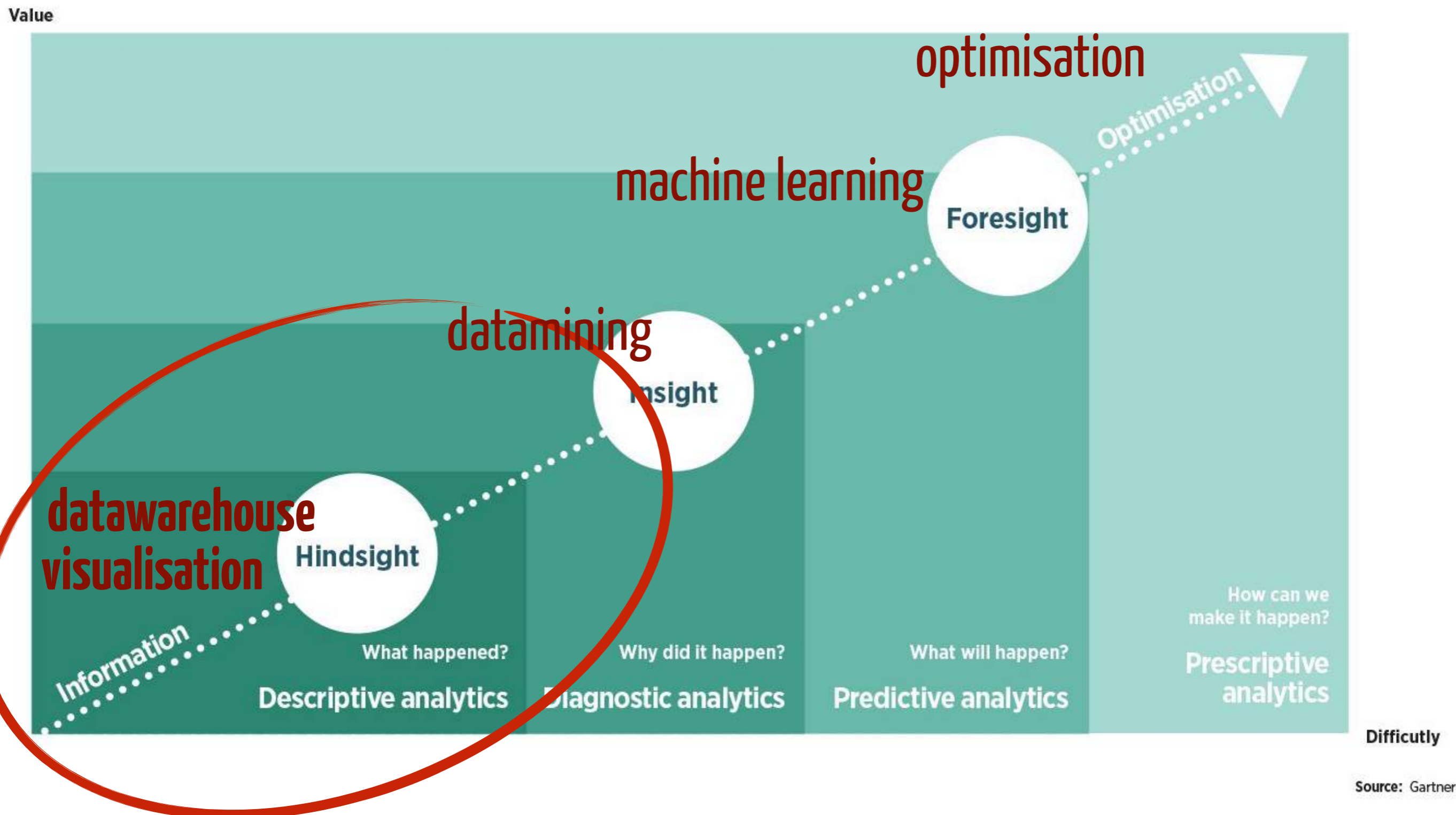
donnée structurée

donnée brute

interfaces homme-machine



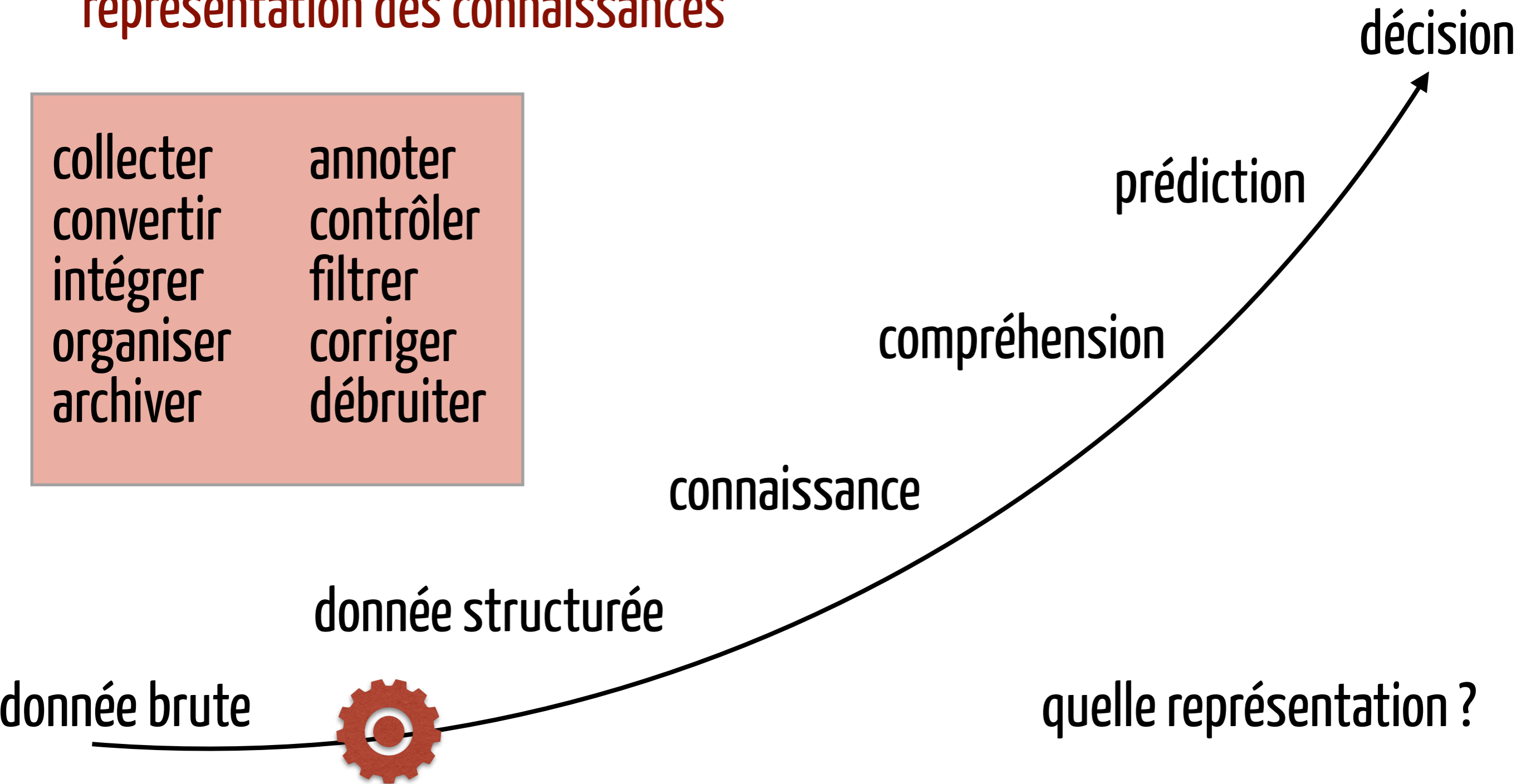
# ex: business analytics (~2006)



traitement du signal  
reconnaissance de forme/langage  
bases de données  
"big data"  
représentation des connaissances

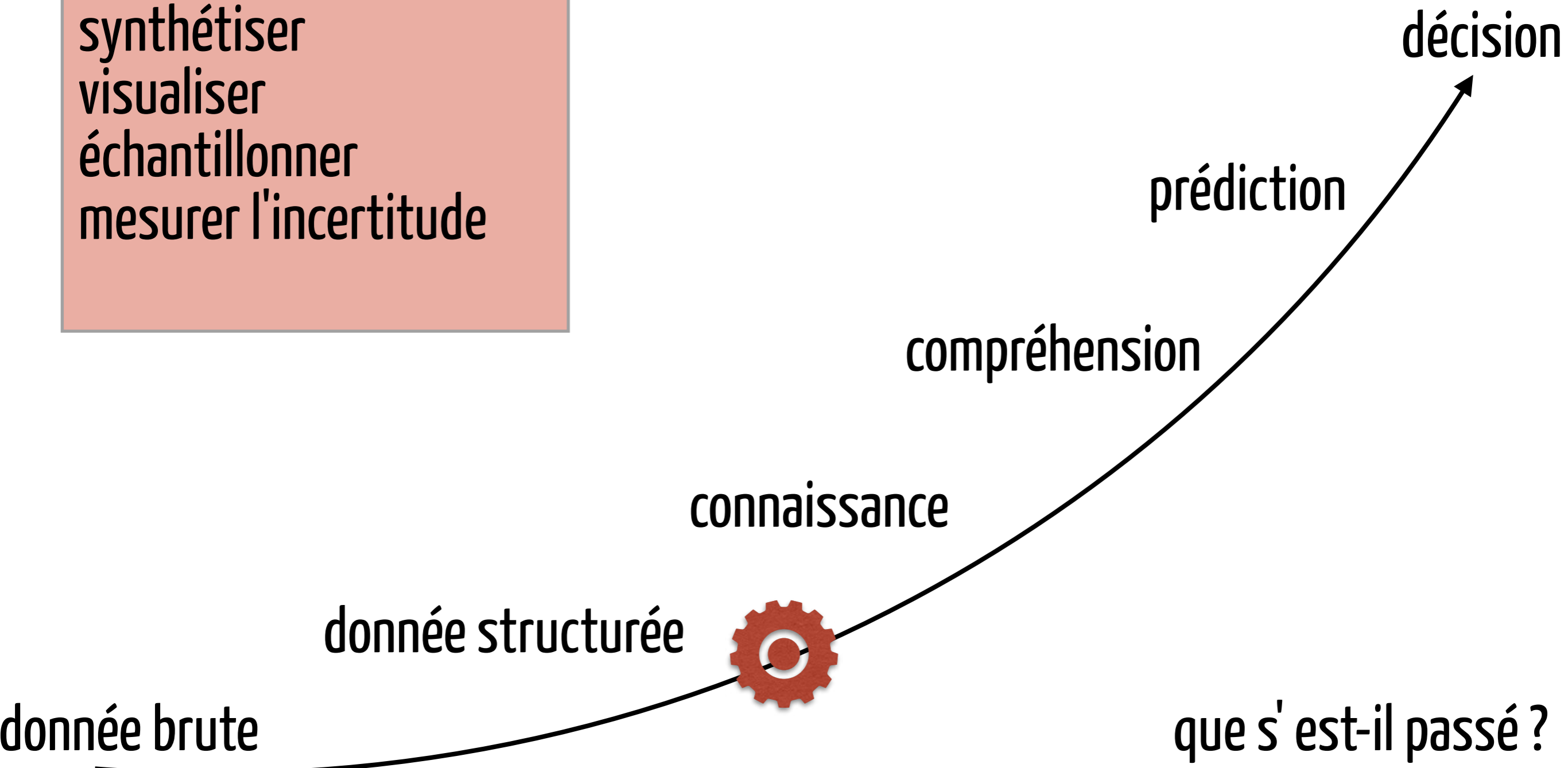
collecter  
convertir  
intégrer  
organiser  
archiver

annoter  
contrôler  
filtrer  
corriger  
débruiter



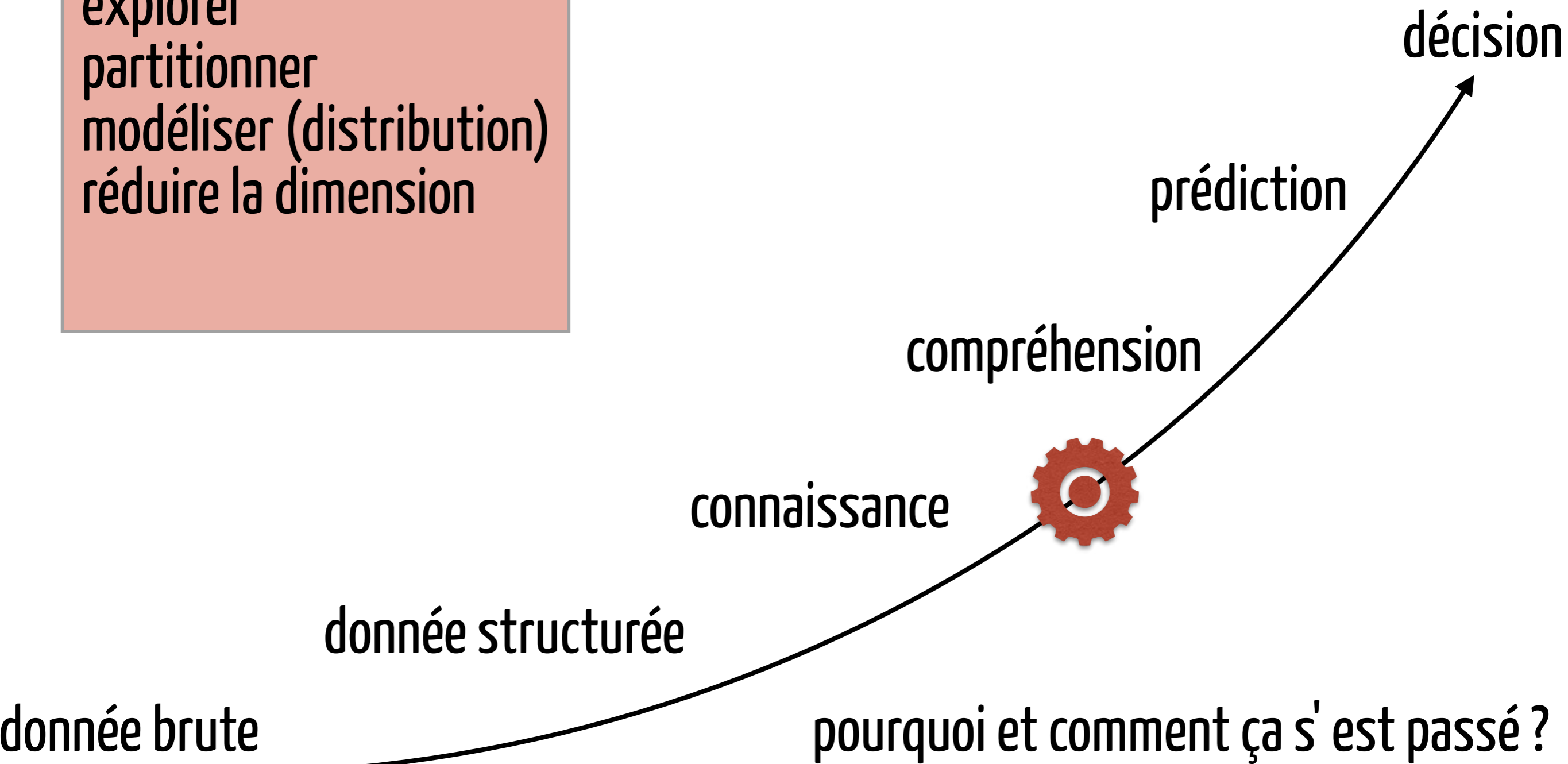
# statistique descriptive analyse de données

synthétiser  
visualiser  
échantillonner  
mesurer l'incertitude



statistique inférentielle  
fouille de données  
apprentissage non-supervisé

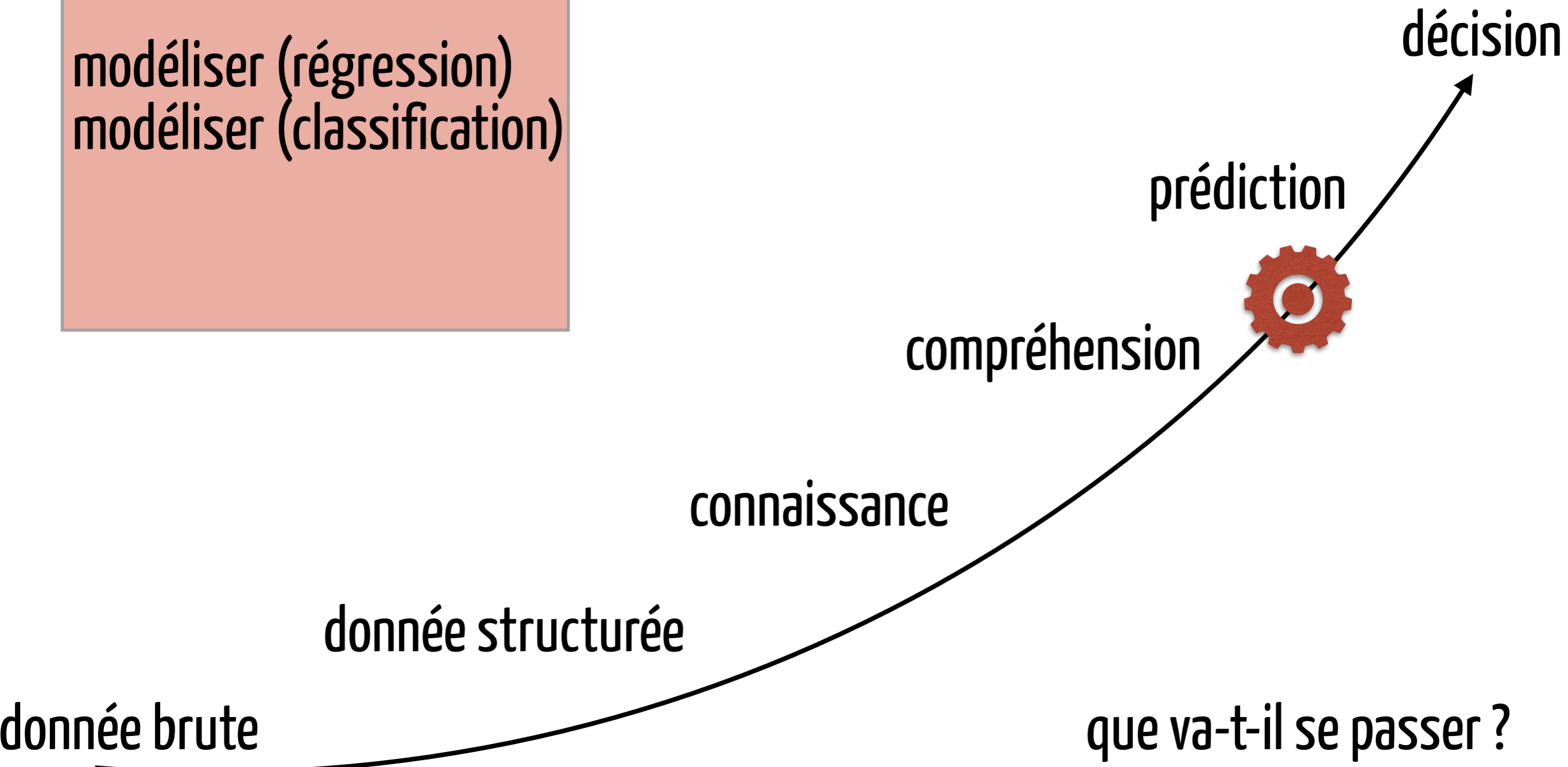
explorer  
partitionner  
modéliser (distribution)  
réduire la dimension





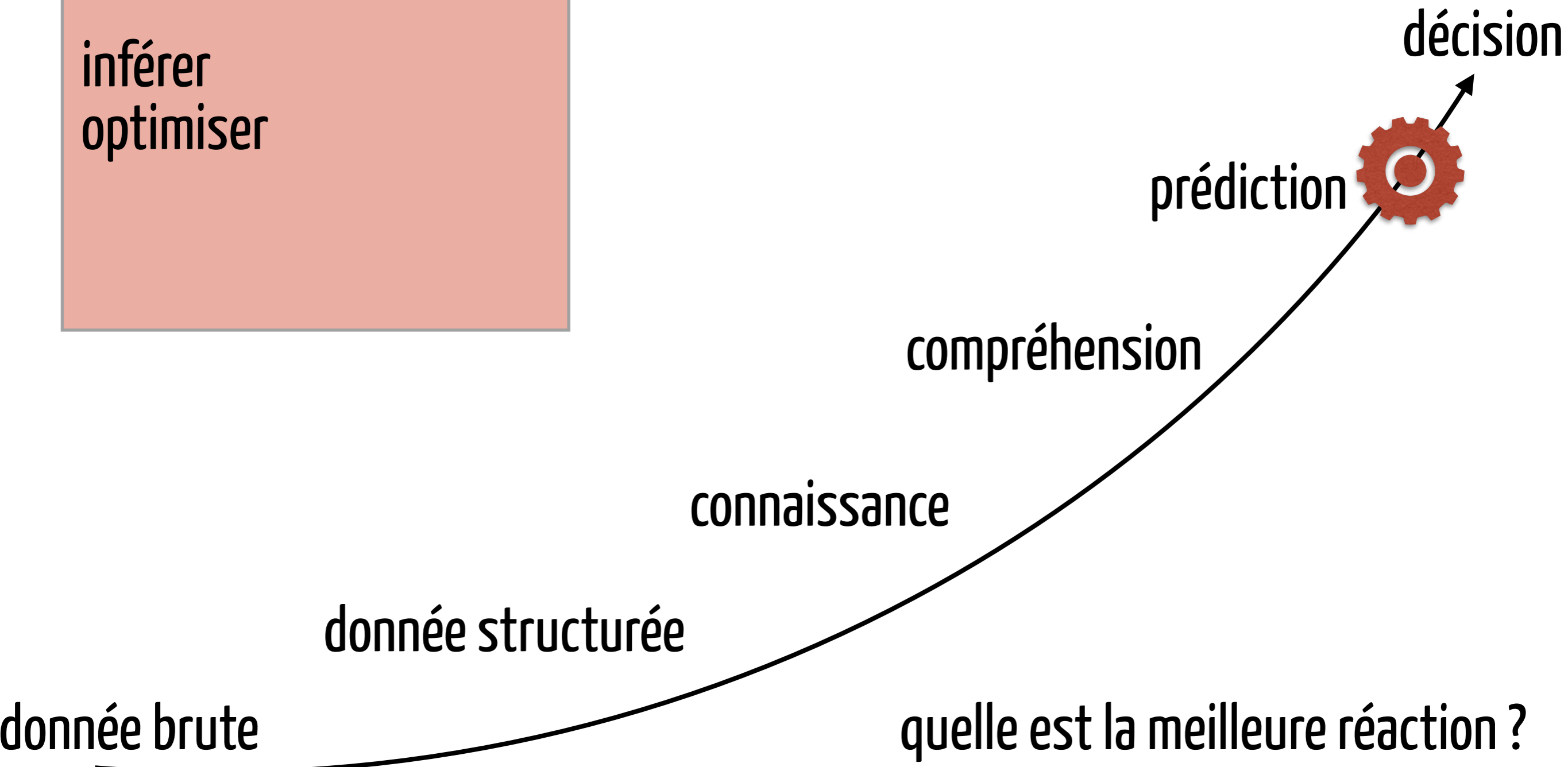
# apprentissage supervisé

modéliser (régression)  
modéliser (classification)



recherche opérationnelle  
raisonnement automatique

inférer  
optimiser

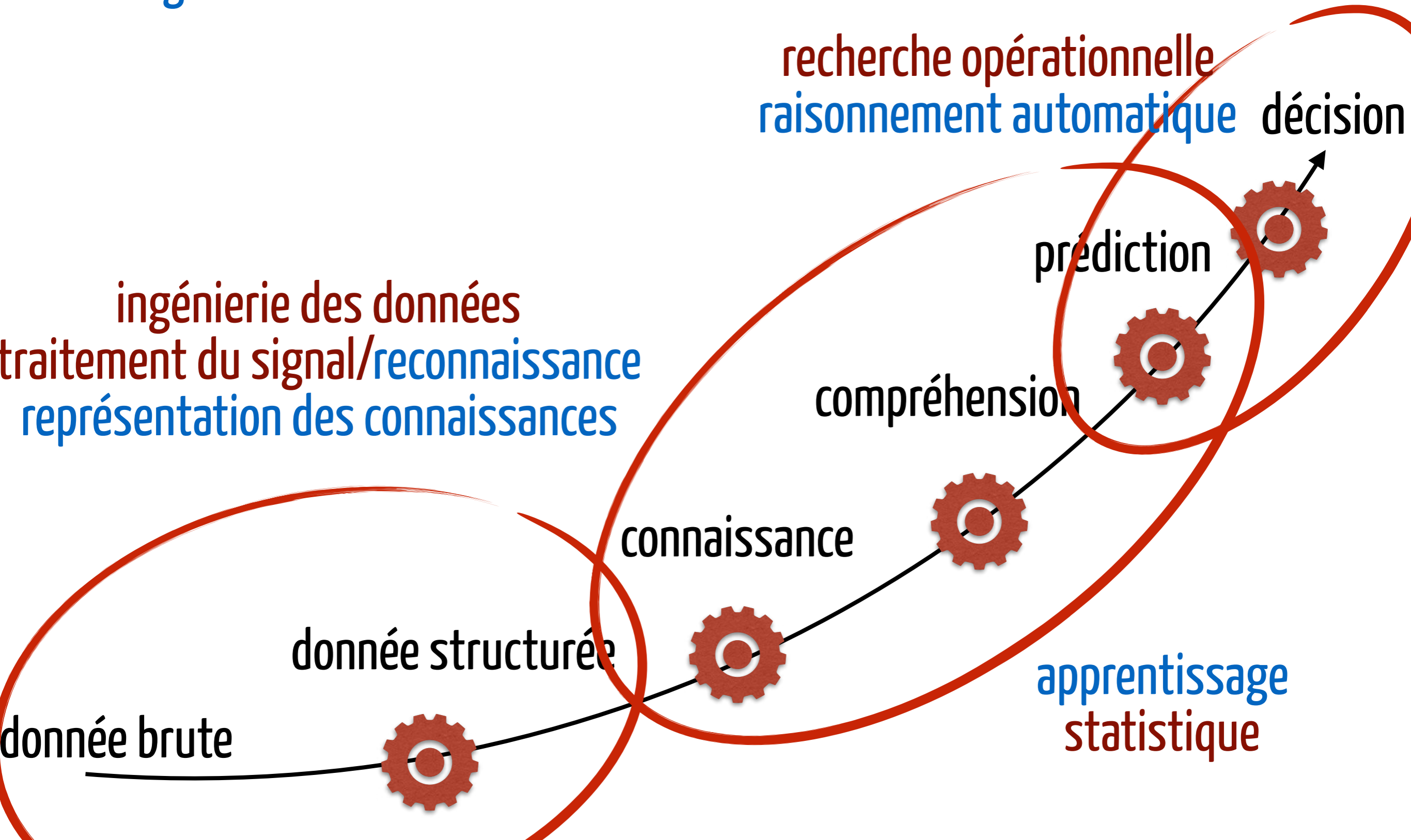


- ingénierie
- intelligence artificielle

# aide à la décision

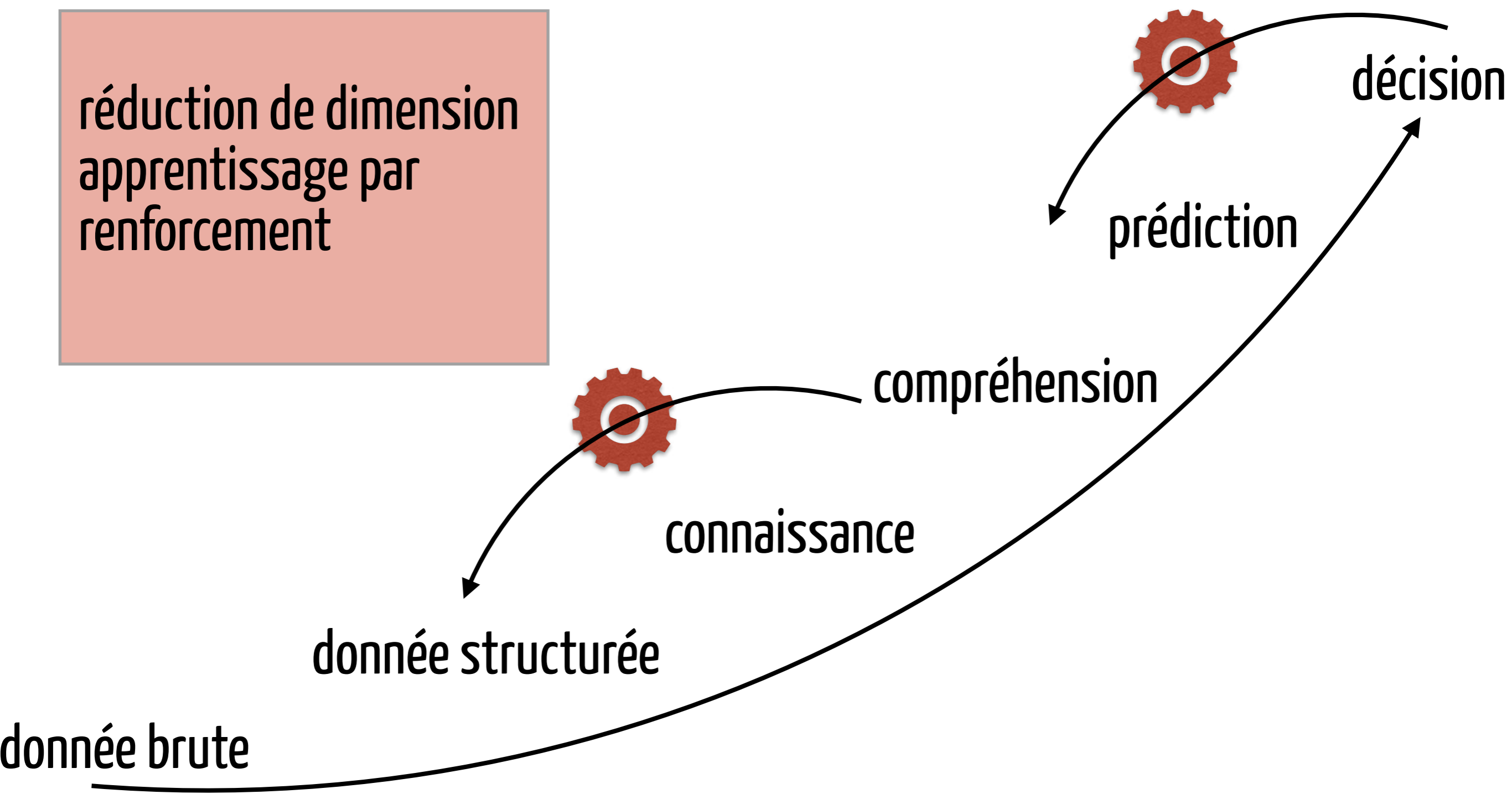
recherche opérationnelle  
raisonnement automatique décision

ingénierie des données  
traitement du signal/reconnaissance  
représentation des connaissances



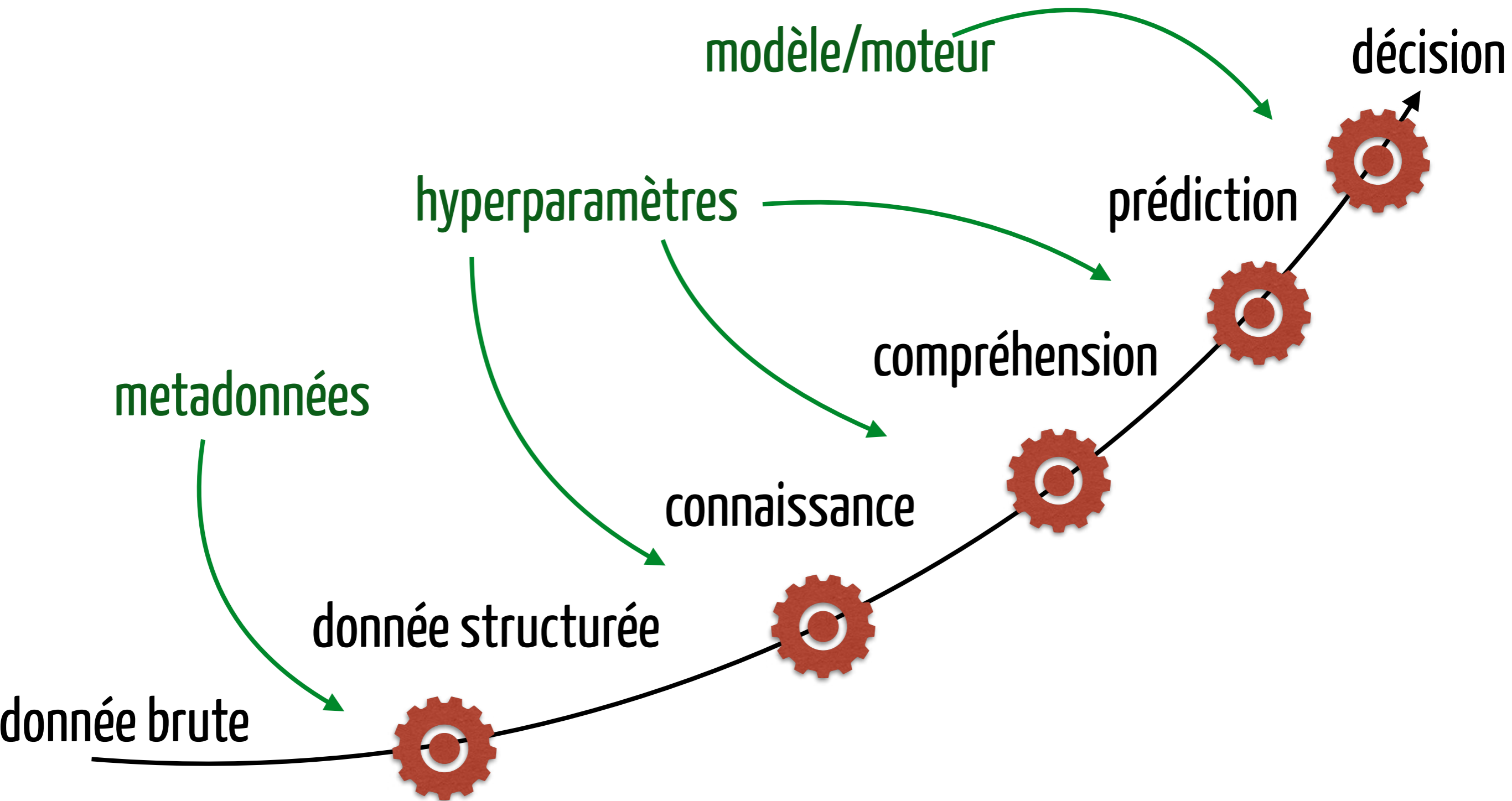
# pas si linéaire

réduction de dimension  
apprentissage par  
renforcement



# pas si automatique

certains éléments constitutifs n'évoluent pas sans intervention humaine



# 2 prédiction, prescription = optimisation

- l'apprentissage repose sur l'optimisation
- la recherche opérationnelle repose sur l'optimisation (différents enjeux)
- modèle: sortie de l'apprentissage, entrée de l'optimisation
- apprentissage pour l'optimisation

# optimiser pour apprendre (1)

- déterminer un modèle qui **minimise le risque empirique** en **apprentissage supervisé**:
  - risque = fonction de perte/erreur (distance entre prédiction  $f(x)$  et cible  $y$ ) sur une observation ou son espérance sur le jeu d'entraînement + régularisateur
  - ex 1 (régression): déterminer les coefficients de la fonction  $f$  qui minimisent la distance quadratique moyenne 
$$\min_f \frac{1}{2n} \sum_{i=1}^n |f(x_i) - y_i|^2$$
  - ex 2 (lasso): idem avec régularisateur  $\lambda \|a\|_1$  pour  $f(x) = ax$  linéaire
  - ex 3 (classification): coût binaire, logistique ou entropie croisée
- action qui **maximise la récompense** en **apprentissage par renforcement**
- calcul des **voisins les plus proches** en **partitionnement**
- **maximisation de l'homogénéité** dans les **arbres de décisions**
- plan séparateur de **marge maximale** dans les **SVM**
- **maximisation de la vraisemblance** dans les **graphes de Markov,...**

# optimiser pour apprendre (2)

- tous les problèmes d'optimisation ne sont pas égaux en terme de **complexité**: convexe/non-convexe, contraint/non-contraint, déterministe/stochastique
- optimisation en (très) **grande dimension**: approches spécifiques (non optimales)
- les problèmes les plus simples admettent des solutions analytiques, ex: régression linéaire et moindres carrés avec régularisation ridge

$$a^* \in \operatorname{argmin}_a \|aX - y\|_2^2 + \lambda \|a\|_2^2 \iff a^* = (\lambda I + X^T X)^{-1} X^T y$$

- d'autres (convexes et différentiables) peuvent être résolus à l'optimum par les algorithmes de descente du 1er (gradient) ou 2nd (Newton) ordre, éventuellement après dualisation dans le cas contraint
- d'autres (non-convexes ou non-différentiables) seront estimés par des **optimas locaux**, (rétropropagation du gradient dans les perceptrons, gradient stochastique,...)
- optimum global  $\equiv$  **sur-apprentissage** / extra-complexité (exécution et mémoire)
- besoin d'algorithmes spécialisés, simples mais **scalables**, à convergence rapide vers des solutions vaguement approchées



# optimiser pour décider



# optimiser pour décider

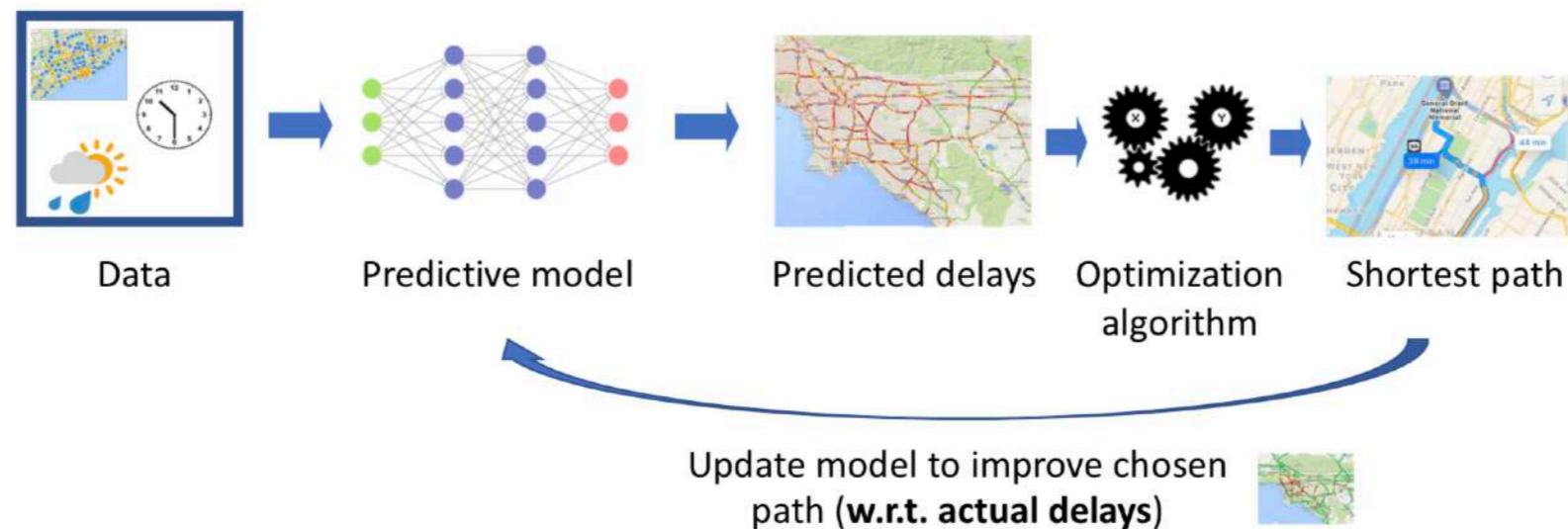
- les décisions sont le plus souvent **discrètes**: pour exprimer des conditions logiques ou des effets de seuil
- l'optimisation combinatoire repose sur des algorithmes différents de l'optimisation numérique, principalement basés sur la séparation et l'évaluation progressive de l'espace de recherche et l'inférence logique:
  - **recherche opérationnelle**: programmation mathématique, programmation dynamique, algorithmes de graphe
  - **intelligence artificielle**: règles, programmation logique/SAT/par contraintes
- approches déclaratives: données (instance) + modèle (spécifique) + solveur (générique)
- alternative (sans certificat d'optimalité):
  - recherche locale (optimisation dans des sous-espaces)
  - métaheuristiques (RO et IA: algorithmes génétiques, recuit simulé, colonies de fourmis, etc.) + simulation
- besoin de modèles **robustes** et d'algorithmes flexibles voire génériques avec certificat de réalisabilité et certificat ou **estimation d'optimalité**, taille des problèmes plus réduite

# apprendre pour décider


- apprendre les solutions (ex: arbres de décision) ou à résoudre [Vinyals 2015]
- metaheuristique + simulation par évaluation du score par un modèle d'apprentissage
- les données d'instances incluent généralement des données prévisionnelles (planification, transport,...) qui peuvent résulter, ainsi que leurs incertitudes, de modèles d'apprentissage. Elles peuvent être aussi échantillonnées a priori.
- le modèle peut être le résultat d'un apprentissage supervisé, ex: [Lombardi 2017]
- le solveur peut intégrer les résultats d'apprentissage supervisé dans ses composantes heuristiques: configuration hors ou en-ligne, ex: [Bonami 2018, Shao 2017]

# optimiser pour apprendre (3)

- **apprentissage par renforcement**: l'agent décide d'une action pour maximiser le score prédit en fonction de son état; l'interpréteur lui retourne le score de la décision après l'exécution de l'action; l'agent intègre ce score dans son modèle de score
- **apprentissage orientée décision**: maximiser la qualité de la décision plutôt que la précision de la prédiction [Ferber 2019]



- optimisation mathématique pour expliquer les modèles [Bertsimas 2017]
- optimisation des hyperparamètres [Audet 2007]



**Bennet, Parrado-Hernández (2006)** The Interplay of Optimization and Machine Learning Research.  
**Sra, Nowozin, Wright (2011)** Optimization for Machine Learning.  
**Bengio, Lodi, Prouvost (2020)** Machine learning for combinatorial optimization: a methodological tour d'horizon.  
**Donoho (2017)** 50 Years of Data Science