



3. Les cadres conceptuels

Les cadres conceptuels

Nous utilisons des données pour représenter le monde, mais aussi afin de :

- décrire le monde à l'aide du **langage**
- le représenter en construisant des **modèles physiques**

Fil conducteur : la **représentation** (un objet qui en remplace un autre, qui est utilisé à sa place afin de s'engager indirectement avec l'objet représenté).

“La carte n'est pas le territoire”, c'est vrai, mais nous n'avons pas besoin de beaucoup d'efforts pour utiliser la carte afin de naviguer le territoire.

La transition entre la **représentation** et le **représenté** peut se faire sans heurts, ce qui pose un risque : **confondre données/résultats analytiques** et le **monde réel**.

Les cadres conceptuels

Meilleure protection : **cadre conceptuel** réfléchi et décrit de manière explicite

- une **spécification des** parties du monde qui sont représentées
- **comment** ils sont représentés
- la **nature de la relation** entre le représenté et le représentant
- **des stratégies appropriées et rigoureuses** pour appliquer les résultats de l'analyse qui est effectuée dans ce cadre de représentation

On pourrait repartir à zéro pour chaque nouveau projet, mais il existe des **cadres de modélisation** qui sont largement applicables à de nombreux phénomènes différents, qui peuvent s'adapter à des cas spécifiques.

Trois stratégies de modélisation

Il y a 3 **stratégies de modélisation** principales (non exclusives) qui peuvent être utilisées pour guider la spécification d'un phénomène ou d'un domaine :

- modélisation **mathématique**
- modélisation **informatique**
- modélisation de **systems**

Les deux premiers ont leur propre monde mathématique/numérique, distinct du monde tangible physique étudié par les chimistes, les biologistes, etc :

- utilisés pour décrire des phénomènes du monde réel en **établissant des parallèles** entre les propriétés des objets et en raisonnant par le biais de ces parallèles.

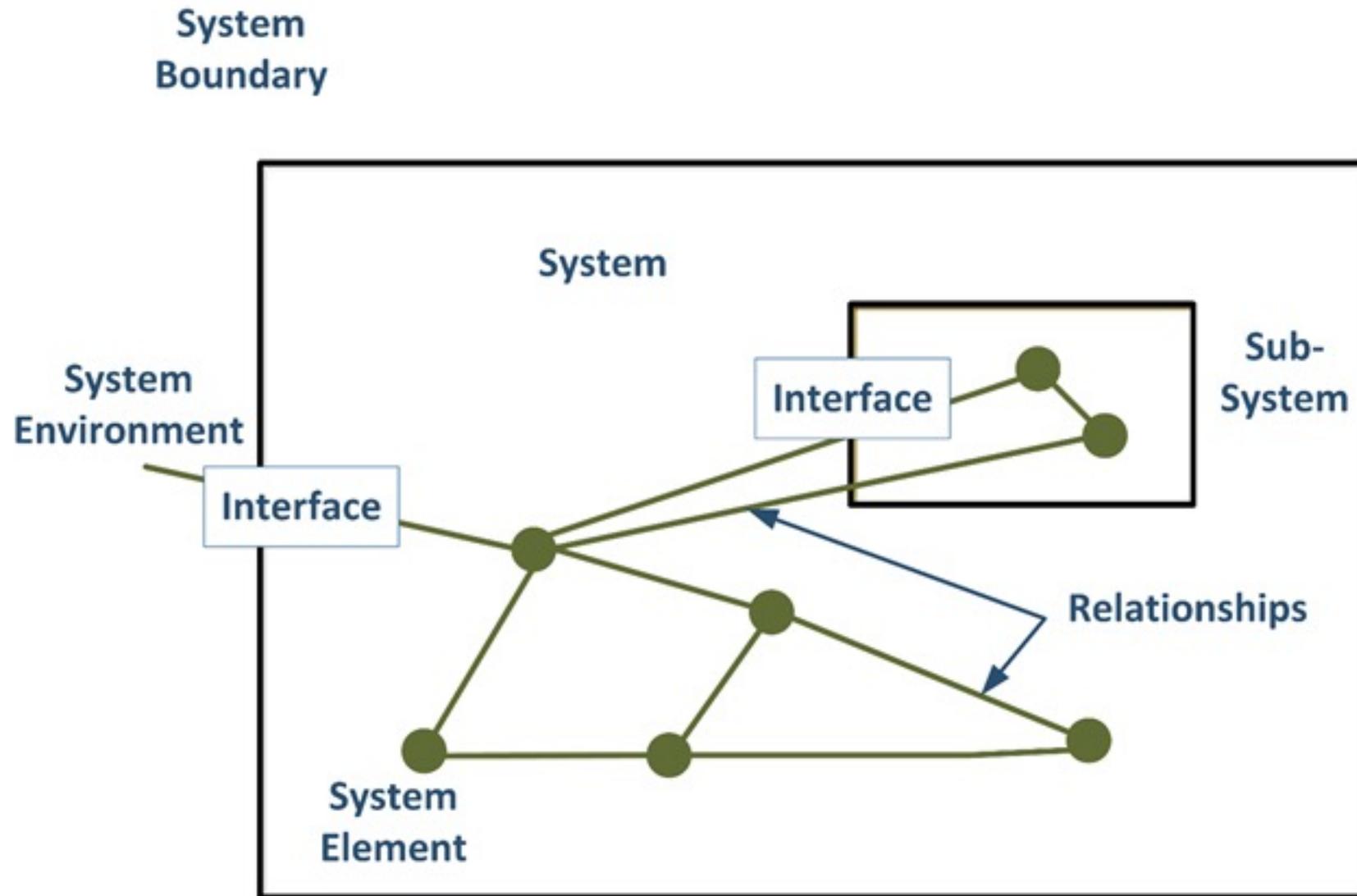
Trois stratégies de modélisation

La théorie générale des systèmes décrit des phénomènes naturels à l'aide d'un **cadre conceptuel commun**, tous étant des systèmes d'objets en interaction.

Lorsque nous sommes confrontés à une nouvelle situation, nous nous demandons :

- quels sont les objets qui semblent les plus pertinents dans les comportements du système ?
- quelles sont les propriétés de ces objets ?
- quels sont les comportements (ou actions) de ces objets ?
- quelles sont les relations entre ces objets ?
- comment les relations entre les objets influencent-elles leurs propriétés et leurs comportements ?

Objectifs : **comprendre le système**, développer une **compréhension commune** cohérente, informer la **collecte de données**, **guider l'interprétation des données**.



La collecte d'information

Il est crucial de parvenir à une **compréhension contextuelle** des données.

Concrètement, comment cette compréhension s'opère-t-elle ?

On peut l'obtenir par le biais :

- d'**excursions sur le terrain**
- des entretiens avec des **experts en la matière**
- de **lectures/visites**
- d'**exploration des données** (le simple fait d'**essayer d'obtenir** ou d'**accéder** aux données peut s'avérer très pénible), etc.

La collecte d'information

Les clients ou les parties prenantes **ne** sont **pas** des entités **uniformes** – les spécialistes des données des clients et les experts peuvent **ne pas apprécier l'implication** des analystes (externes et/ou internes).

La collecte d'informations donne aux analystes l'occasion de montrer que tout le monde tire dans la même direction, en :

- posant des questions **significatives**
- **s'intéressant véritablement** aux expériences des experts/clients
- reconnaissant la capacité de chacun à contribuer

Un peu de tact peut s'avérer utile lorsqu'il s'agit de recueillir des informations.

Penser en termes de systèmes

Un **système** est composé d'**objets** dont les **propriétés** peuvent changer au fil du temps.

Au sein du système, il y a des **actions/propriétés évolutives**, c-à-d des **processus**.

On comprend comment les différents aspects du monde interagissent ensemble en **découpant des morceaux** correspondant aux aspects et en définissant leurs limites.

Le travail avec d'autres intelligences requiert une **compréhension partagée** de ce qui est étudié.

Les objets eux-mêmes ont diverses propriétés.

Penser en termes de systèmes

Les processus naturels génèrent/détruisent des objets, et modifient les propriétés de ces objets au fil du temps.

Nous **observons**, **quantifions**, et **enregistrons** les valeurs de ces propriétés à des moments précis.

Les observations permettent de **saisir la réalité sous-jacente** avec un degré acceptable de **précision** et d'**erreur**, mais ... **même le meilleur modèle de système ne fournit jamais qu'une approximation de la situation analysée.**

Avec de la chance, de l'expérience, de la prévoyance, ces approximations peuvent être **valables**.

Identifier les lacunes de compréhension

Une **lacune dans les connaissances** est identifiée lorsque nous nous rendons compte que ce que nous pensions savoir sur un système s'avère **incomplet** (ou manifestement faux).

Causes :

- naïveté vis-à-vis de la situation modélisée
- la nature du projet envisagé

Avec **trop de parties mobiles**, des **objectifs irréalistes**, une **distance par rapport au pipeline**, les lacunes en matière de connaissances ne peuvent être évitées (même avec de petits projets bien organisés et faciles à contenir).

Identifier les lacunes de compréhension

Les lacunes en matière de connaissances peuvent survenir à **plusieurs reprises** :

- **nettoyage des** données
- **consolidation des** données
- **analyse des** données
- même pendant la **communication des résultats** (!)

Lorsque vous êtes confronté à un manque de connaissances, **soyez flexible** :

- **revenez en arrière**
- **posez des questions**
- **modifiez la représentation du système** aussi souvent que nécessaire

Il est préférable de combler ces lacunes dès le début du processus (évidemment).

Les modèles conceptuels

Les modèles conceptuels sont construits à l'aide d'outils d'investigation méthodiques :

- **diagrammes**
- **entretiens** structurés
- **des descriptions** structurées, etc.

Les scientifiques des données doivent se méfier des **modèles conceptuels implicites** (lacunes dans les connaissances).

Il est préférable de privilégier le côté du "trop de modélisation conceptuelle", mais n'oubliez pas que "tout modèle est faux ; certains modèles sont utiles" [G.E. Box].

Il est acceptable de construire de meilleurs modèles, de manière itérative.

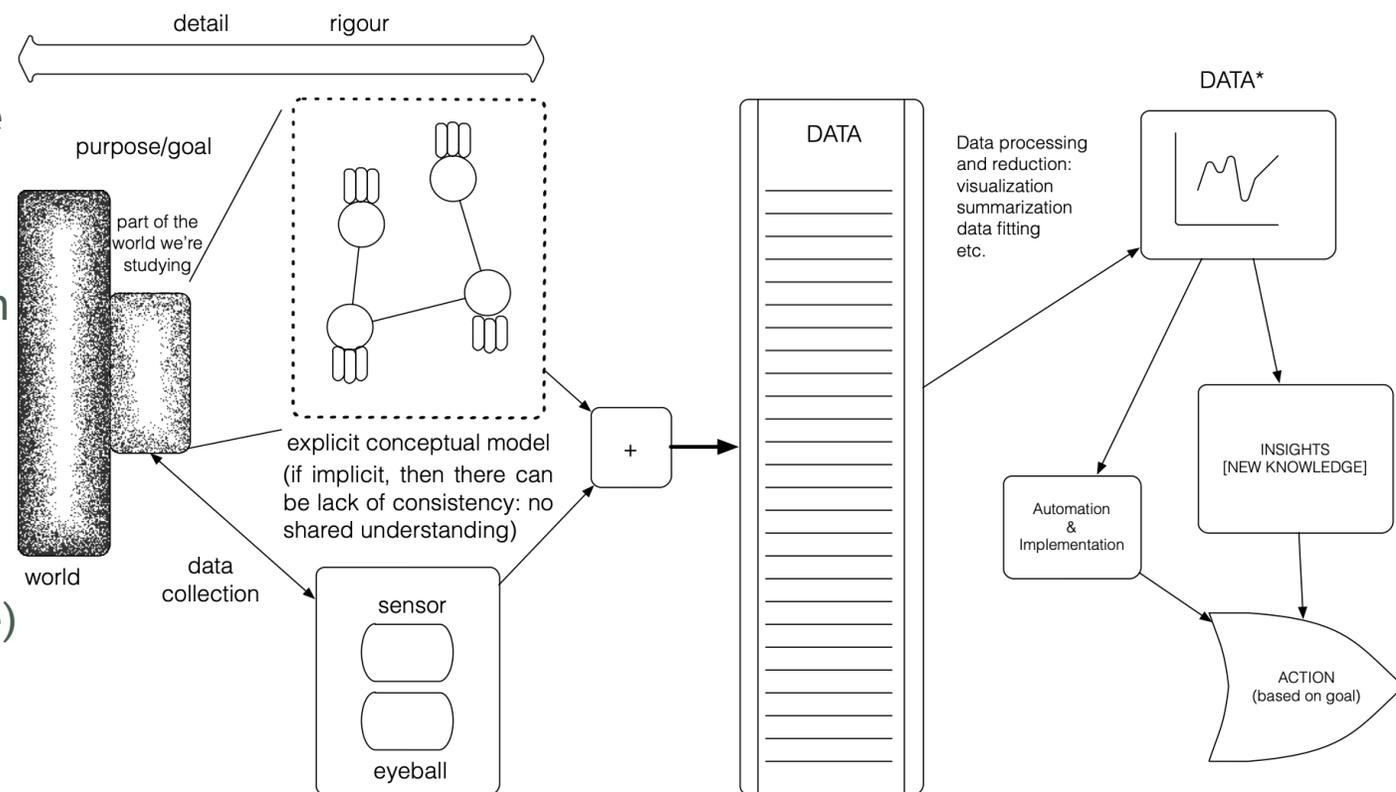
Les modèles conceptuels

Modèles conceptuels :

- ne sont pas mises en œuvre sous forme de modèle d'échelle ou de code informatique
- n'existent que de manière conceptuelle, souvent sous la forme d'un diagramme ou d'une description verbale d'un système – boîtes et flèches, cartes mentales, listes, définitions, etc.

L'accent est mis sur :

- les **états possibles** (pas de comportement spécifique)
- des types d'objets, et non des instances spécifiques ; l'objectif est l'**abstraction**

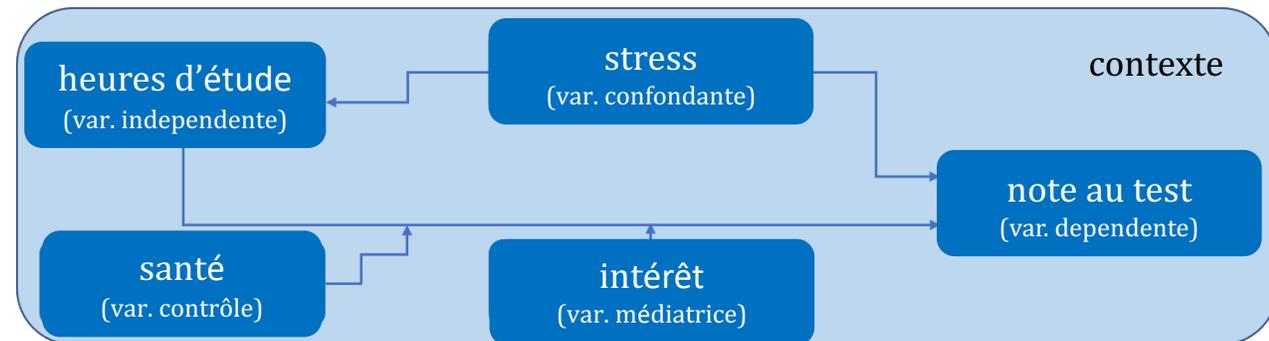


Les modèles conceptuels

En pratique, nous devons d'abord sélectionner un système pour la tâche à accomplir, puis générer un modèle conceptuel qui englobe :

- des **objets pertinents** et **clés** (abstraites ou concrets) ;
- les **propriétés** de ces objets, et leurs valeurs ;
- les **relations entre les objets** (partie-tout, est-un, 1-à-plusieurs, etc.), et
- les **relations entre les propriétés** à travers les instances d'un type d'objet.

Voici un exemple simpliste décrivant une relation supposée entre une **cause présumée** (heures d'étude) et un **effet présumé** (note au test).



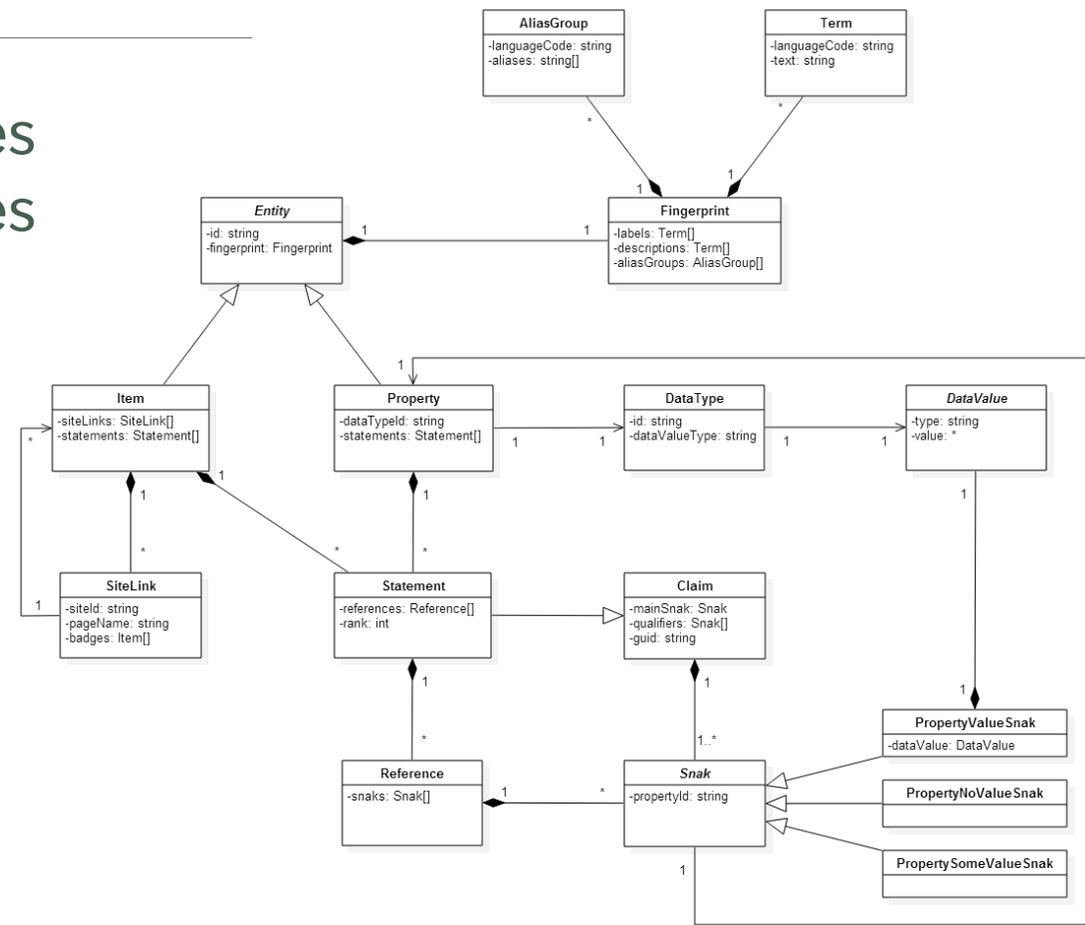
Les modèles conceptuels formels

La modélisation conceptuelle transforme les modèles conceptuels implicites en modèles **explicites** et **tangibles**.

Elle offre la possibilité d'**examiner** et d'**explorer** les idées et les hypothèses.

Divers efforts ont été déployés pour **formaliser** la modélisation conceptuelle :

- UML (langage universel de modélisation)
- modèles de relations entre entités (ER)



Relier les données au système

Les données collectées et analysées sont-elles **utiles pour comprendre le système** ? On peut mieux répondre à cette question si l'on comprend :

- **comment les** données sont collectées
- la **nature approximative** des données et du système
- ce que les données représentent (observations et caractéristiques)

La combinaison du système et des données est-elle **suffisante** pour comprendre la situation considérée ? Il est difficile de répondre en pratique.

Si les données, le système, et le monde réel ne sont **pas alignés**, tout aperçu des données tiré de la modélisation et de l'analyse pourrait s'avérer inutile.

Les biais cognitifs

Les biais cognitifs ont un impact sur la façon de construire des modèles et de rechercher des schémas dans les données :

- le **biais d'ancrage** nous amène à nous fier trop fortement à la première information que l'on nous donne sur un sujet
- l'**heuristique de disponibilité** décrit notre tendance à utiliser les informations qui nous viennent rapidement et facilement à l'esprit lorsque nous prenons des décisions
- l'**effet "bandwagon"** désigne notre habitude d'adopter des comportements ou des croyances parce que bcp d'autres personnes font de même
- le **biais d'appui du choix** nous mène à considérer nos actions sous un jour positif
- l'**illusion du regroupement** fait référence à notre tendance à voir des schémas dans l'aléatoire
- le **biais de confirmation** décrit notre tendance à remarquer et à accorder plus de crédit aux preuves qui appui nos croyances existantes
- le **biais de conservation** se produit lorsque nous privilégions les preuves antérieures par rapport aux nouvelles informations
- l'**effet de l'autruche** décrit la façon dont les gens évitent souvent les informations négatives, y compris les commentaires qui les aident à suivre la progression de leurs objectifs

Les biais cognitifs

- le **biais lié aux résultats** consiste à juger une décision en fonction du résultat, plutôt que de la raison pour laquelle elle a été prise
- l'**excès de confiance** nous pousse à prendre plus de risques dans notre vie quotidienne
- le **biais pro-innovation** se produit lorsque les partisans d'une technologie sur-évaluent son utilité et sous-évaluent ses limites
- le **biais de récence** se produit lorsque nous favorisons les nouvelles informations par rapport aux preuves antérieures
- le **biais du risque zéro** est lié à notre préférence pour la certitude absolue
- le **biais de survie** est un raccourci cognitif qui se produit lorsqu'un sous-groupe visible ayant réussi est pris pour un groupe entier
- le **biais de saillance** décrit notre tendance à nous concentrer sur les éléments ou les informations les plus remarquables et à ignorer ceux qui n'attirent pas notre attention.

Autres biais :

- sophisme du taux de base, biais de la rationalité limitée, biais de la taille des catégories, effet Dunning-Kruger, effet de cadrage, sophisme de la main chaude, effet IKEA, illusion de validité, corrélations illusoire, etc.

Lectures suggérées

Les cadres conceptuels

Data Understanding, Data Analysis, Data Science
Data Science Basics

Conceptual Frameworks for Data Work

- Three Modeling Strategies
- Information Gathering
- Cognitive Biases

Exercices

Les cadres conceptuels

1. Considérez la situation suivante : vous êtes en voyage d'affaires et vous avez oublié de remettre un dessin d'architecture très important (et requis de toute urgence) à votre superviseur avant de partir. Votre bureau enverra un stagiaire pour le récupérer dans votre espace de vie. Comment allez-vous lui expliquer, par téléphone, comment trouver le document ? Si le stagiaire est déjà venu dans votre espace de vie, si son espace de vie est comparable au vôtre, ou si votre conjoint est à la maison, le processus peut être considérablement accéléré, mais avec quelqu'un pour qui l'espace est nouveau (ou une personne ayant une déficience visuelle, par exemple), il est facile de voir comment les choses pourraient se compliquer. Le temps est un facteur essentiel - vous et le stagiaire devez faire le travail **correctement** et le plus **rapidement possible**. Quelle est votre stratégie ?
2. Traduisez les biais cognitifs en contextes analytiques. Quels sont les biais cognitifs auxquels vous, votre équipe et votre organisation êtes les plus sensibles ? Le moins ?