



HAL
open science

Evaluation de la valeur des données - Modèle et méthode

Jacky Akoka, Isabelle Comyn-Wattiau

► **To cite this version:**

Jacky Akoka, Isabelle Comyn-Wattiau. Evaluation de la valeur des données - Modèle et méthode. Actes du 40ème congrès INFORSID, Association Inforsid, May 2022, Dijon, France. pp.163-178. hal-04026275

HAL Id: hal-04026275

<https://hal.science/hal-04026275>

Submitted on 11 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluation de la valeur des données

Modèle et méthode

Jacky Akoka¹, Isabelle Comyn-Wattiau²

1. CEDRIC-CNAM & IMT Business School

2 rue Conté, Paris, FRANCE

jacky.akoka@lecnam.net

2. ESSEC Business School

RÉSUMÉ. Les données sont un actif stratégique pour les organisations. Évaluer leur valeur permet d'identifier les stratégies offrant un avantage concurrentiel et de mesurer le capital informationnel d'une entreprise. Il n'existe pas d'approche ni de normes et règles à appliquer dans cette évaluation du fait du caractère immatériel des données. Dans cet article, nous proposons une approche combinant un concept enrichi de valeur de la donnée, un modèle conceptuel incluant entre autres ce concept et la notion de contexte et une méthode d'évaluation de la valeur s'appuyant sur ce modèle. L'originalité est de combiner les approches fondées sur les risques avec les approches plus classiques d'évaluation des actifs tangibles (approche par les coûts, par le marché ou par le revenu). L'approche est confrontée à plusieurs cas réels issus de la littérature permettant une validation de son utilité et de sa valeur ajoutée.

ABSTRACT. Data is a strategic asset for organizations. Assessing its value provides companies with a mean to identify strategies that offer a competitive advantage and to measure their informational capital. There is neither an approach nor a set of standards and rules to apply for this valuation due to the intangible nature of data. In this paper, we propose an approach combining an enriched concept of data value, a conceptual model including this concept and the notion of context, and a value assessment method based on this model. The originality is to combine risk-based approaches with the more classical approaches to the valuation of tangible assets (cost, market, or income approaches). The approach is confronted with several real cases from the literature allowing a validation of its usefulness and added-value.

MOTS-CLES : Valeur des données, évaluation, modèle conceptuel, approche hiérarchique multicritères, contexte, risque, coût.

KEYWORDS: Data value, data valuation, conceptual model, analytical hierarchy process (AHP), context, risk, cost.

1. Introduction

Les données jouent un rôle central dans le fonctionnement des organisations. Elles sont les actifs intangibles qui sous-tendent l'économie numérique. Les entreprises s'appuient sur les données pour prendre des décisions, notamment pour leurs investissements, ou pour générer des indicateurs de performance. Selon McKinsey, plus d'un tiers des ventes d'Amazon proviennent de son moteur de recommandation. Netflix affirme que ses recommandations fondées sur des algorithmes de traitement des données lui permettent d'économiser un milliard de dollars par an. La capitalisation boursière de Facebook est d'environ 1 000 milliards de dollars, mais la valeur nette de l'entreprise fondée sur l'actif et le passif n'est que de 138 milliards de dollars. La différence en termes de valeur s'explique par les données que Facebook collecte auprès des utilisateurs et qu'elle utilise à son tour pour alimenter ses algorithmes publicitaires. Les données sont donc un actif essentiel de nombreuses entreprises.

L'évaluation de la valeur des données est le processus qui consiste à attribuer une valeur monétaire qui représente l'importance économique des données pour l'organisation et ses différentes parties prenantes (employés, actionnaires, clients, fournisseurs, etc.). Toutefois, ce concept de valeur est multi facettes. Il s'agit bien sûr de valeur monétaire, mais aussi d'un concept économique relatif au bien-être social et sociétal. Dans ce cas, les données génèrent de la valeur lorsqu'elles permettent aux entreprises de créer des emplois, de devenir plus productives ou encore aux gouvernements de fournir des services publics plus efficaces.

L'évaluation de la valeur n'est pas simple. D'une part, les mêmes données peuvent être très importantes pour une organisation et n'avoir aucune valeur pour une autre. D'autre part, la maturité de certaines organisations ne leur permet pas de comprendre la valeur de leurs données et, encore moins, de l'évaluer. Enfin, le choix d'une méthode d'évaluation n'est pas aisé.

L'objectif de cet article est de proposer un modèle et une méthode permettant d'assister les entreprises dans le processus d'évaluation de la valeur de leurs données. Une contribution originale est la prise en compte combinée des dimensions *valeur, risque et coût* pour une meilleure appréhension de la valeur résultante. Nous proposons une approche d'évaluation de la valeur fondée sur trois piliers. Le premier est la définition du concept de *Valeur de la donnée*. Le second est un modèle conceptuel décrivant les informations nécessaires à l'évaluation de cette valeur des données. Le troisième pilier est une méthode d'aide à l'évaluation de la valeur de la donnée fondée sur le modèle conceptuel et structurée suivant l'approche hiérarchique multicritère.

Le reste de cet article est organisé comme suit. La deuxième section propose un état de l'art des principaux concepts et approches d'évaluation de la donnée. La

troisième section décrit successivement les trois piliers de l'approche et sa validation. Enfin, la quatrième section conclut et esquisse des voies de recherche future.

2. L'évaluation de la valeur des données : un état de l'art

De nombreux articles décrivent les données comme des actifs intangibles (Savona, 2019; Corrado, 2019; Otto, 2015; Wdowin et Diepeveen, 2020). Ces actifs sont soumis à des réglementations et évalués à l'aide de techniques spécifiques (Ciuriak, 2019). De nature intangible, les données n'obéissent pas à la même dynamique concurrentielle que les actifs tangibles traditionnels (équipements, stocks, immobilier, etc.). Toutefois, leur gestion suit ou devrait suivre une approche similaire. En d'autres termes, il est impératif de comprendre la valeur des données et de les gérer tout au long de leur cycle de vie. Il est tout aussi important d'évaluer les risques liés à leur exploitation et de mesurer le retour sur investissement des projets d'acquisition et de transformation de données (Short et Todd, 2017).

Les données ont un coût parfois considéré comme irrécupérable (Wang et Zhao, 2020) ou dont le retour sur investissement est difficile à évaluer. En effet, la manière dont elles créent de la valeur n'est pas toujours claire pour les entreprises (Zeiter *et al.*, 2021). Elles peuvent notamment créer de la valeur en éclairant les décisions. Par exemple, elles peuvent permettre de concentrer les ressources limitées, dont disposent les organisations, sur les domaines qui créeront le plus de valeur (Attard et Brennan, 2018). Certains auteurs considèrent que la valeur des données augmente avec leur utilisation notamment en exploitant les métadonnées. D'autre part, plus les données sont précises, plus elles ont de la valeur (PwC, 2019). Toutefois, la qualité et la quantité des données sont soumises à la loi des rendements décroissants (Moody et Walsh, 1999). Certaines données sont périssables, par exemple, les données sur les clients qui ne sont pas régulièrement mises à jour. Enfin, plusieurs facteurs, comme l'exactitude, l'intégrité, la disponibilité et la fraîcheur, affectent le coût de la donnée.

Les gouvernements et les organismes internationaux mettent en place des initiatives pour mieux maîtriser l'actif « donnée ». Ainsi, l'initiative Data for Common Purpose (World Economic Forum, 2021) vise à produire un cadre de gouvernance pour améliorer les avantages sociétaux des données. C'est aussi le cas de la loi européenne sur la gouvernance des données (European Commission, 2020) ou encore le plan d'action de l'US Federal Data Strategy (Federal Data Strategy, 2022) qui vise à mieux exploiter les données en tant qu'actif stratégique.

Plusieurs méthodes d'évaluation de la valeur des données ont été proposées (Moody et Walsh, 1999 ; Ciuriak, 2019 ; Wang et Zhao, 2020 ; Otto, 2015). Une première famille est composée de techniques d'estimation de la valeur au travers des coûts que la gestion des données requiert (coût de production et de stockage des données, coût de remplacement des données obsolètes, etc.) et de leur impact sur le flux de trésorerie. L'un des principaux avantages de cette famille de méthodes est sa facilité

d'utilisation. La deuxième famille cible l'évaluation de la valeur marchande des données, fondée sur l'estimation du prix que les entreprises paient pour des données comparables sur le marché. Simples à mettre en œuvre, ces méthodes ne s'appliquent pas à toutes les données dans la mesure où certaines données ne sont tout simplement pas échangeables, parce qu'elles représentent pour les entreprises un avantage concurrentiel. De plus, pour obtenir un prix réel des données, il faut qu'il existe un marché efficace, ce qui n'est pas toujours le cas. En outre, il convient de rappeler que le prix n'est pas synonyme de valeur. La troisième famille est caractérisée par l'approche de la valeur économique (Garifova, 2015) qui peut être utilisée pour identifier la valeur ajoutée de la donnée, par exemple à des fins commerciales ou pour des cas d'usage spécifiques. Comme pour les valeurs obtenues via les autres approches, une grande partie de cette valeur est subjective.

Les approches précédentes appliquent les techniques classiques d'évaluation des actifs tangibles à la donnée, qui est par nature intangible. D'autres approches, fondées sur les externalités, proposent de considérer la donnée à un niveau plus large, par exemple en considérant les bénéfices tirés par la société de l'utilisation de ces données (Coyle et Diepeveen, 2021 ; Antuca & Noble, 2021). C'est le cas notamment des données ouvertes. A notre connaissance, aucune approche d'évaluation de la valeur de la donnée proprement dite ne s'appuie explicitement sur les risques associés. A noter toutefois qu'une étude comparative de ces approches y insère aussi les approches d'évaluation du risque (Bodendorf *et al.*, 2022). Ainsi, il n'existe pas d'approche holistique qui combine tous ces points de vue pour produire une évaluation de la valeur de la donnée. Cette combinaison suppose toutefois que les trois dimensions soient ramenées à une même unité qui est financière. C'est précisément cette évaluation holistique que nous proposons de faire dans cet article.

3. Notre approche

Nous proposons ci-dessous une approche facilitant le processus d'évaluation de la valeur des données. Elle s'appuie sur trois artefacts. Le premier est un nouveau construit représentant la valeur de la donnée. Le second consiste en un modèle conceptuel regroupant l'ensemble des informations qui interviennent dans le processus d'évaluation de la donnée. Le troisième est la méthode proprement dite qui est fondée sur le modèle conceptuel des données et structurée suivant l'approche hiérarchique multicritère. La validation de notre approche est décrite ensuite.

3.1. La valeur de la donnée

La donnée, actif de l'entreprise, nécessite la mise en place d'une gouvernance au même titre que les ressources humaines, le patrimoine immobilier ou tout autre actif stratégique. La gouvernance de la donnée peut ainsi être définie par un triple objectif : maximiser sa valeur en minimisant les risques et les coûts associés (Tallon, 2013). La valeur est ainsi définie selon trois dimensions : juridique (obligation de conformité et gestion des risques), métier (pertinence pour le fonctionnement et les revenus de l'entreprise) et historique (coûts et impact sur les flux de trésorerie) alors

que les méthodes d'évaluation de la littérature n'intègrent qu'une seule de ces dimensions : la valeur métier ou la valeur historique fondée sur les coûts. Les méthodes d'évaluation des risques sont cantonnées dans le domaine de la sécurité des systèmes d'information mais ne sont jamais agrégées aux approches d'évaluation de la valeur. C'est pour y remédier que nous proposons de définir la *valeur de la donnée* comme un concept à trois dimensions :

- la *valeur* produite par la donnée pour l'organisation qu'il s'agisse de sa valeur d'usage (l'utilisation de la donnée produit une valeur) ou de sa valeur d'échange (la donnée peut être vendue en tant que telle ou associée à un produit ou un service) ;
- le *coût* engendré par la donnée qui ne se limite pas au coût d'acquisition mais inclut tous les aspects de son cycle de vie (acquisition, production, stockage ou archivage, utilisation, partage, destruction). Ce coût vient en diminution de la valeur produite par la donnée ;
- enfin, le *risque* lié à la donnée. Il est légal ou réglementaire (un non-respect des lois ou réglementations liées aux données peut générer une perte financière). Il peut aussi être stratégique (une mauvaise décision prise à cause de données erronées) ou entacher la réputation de l'organisation (un site web de mauvaise qualité impacte négativement l'image). Enfin, les risques opérationnels incluent des aspects liés aux données : de nombreuses activités des entreprises sont dépendantes de la disponibilité et de la qualité des données qu'elles utilisent.

Pour faciliter le processus d'évaluation de la valeur des données, nous présentons ci-dessous un modèle conceptuel qui s'appuie ces trois dimensions de la valeur.

3.2. Le modèle conceptuel de données

Le modèle conceptuel a pour objectif de décrire et relier toutes les informations nécessaires à l'évaluation de la valeur de la donnée (Figure 1). Il décrit donc les trois dimensions de la valeur présentées précédemment. La *donnée* elle-même est décrite. On distingue la donnée interne de la donnée externe, dont les méthodes d'évaluation sont différentes. D'autres attributs qui interviennent dans le processus d'évaluation sont représentés tels que le volume, l'accessibilité, la fréquence d'usage et l'exclusivité. La donnée est relative à un domaine (par exemple le domaine client, le domaine RH, etc.). Elle est caractérisée par des facteurs de qualité (fraîcheur, précision, etc.) qui impactent sa valeur. Sa nature (donnée numérique, image, donnée capteur, etc.) est aussi un élément important.

Le deuxième concept est celui de *dimension*. Une donnée est évaluée selon trois dimensions : le coût, la valeur et le risque. Sa valeur nette résulte de l'agrégation de ces trois dimensions. La valeur d'usage représente la capacité de l'entreprise à améliorer son efficacité interne ainsi que le potentiel de développement du business qu'elle peut générer ou encore la capacité de croissance externe (alliances, fusions,

dans l'évaluation de ses actifs qu'une entreprise privée. Le contexte externe joue un rôle tout aussi important. Nous le caractérisons par les attributs de type PESTEL (Politique, Economique, Social, Technologique, Environnemental et Légal) (Aguilar, 1967). Le contexte détermine la pertinence et l'importance relatives des différentes dimensions de l'évaluation. Ainsi, dans certains contextes, par exemple un établissement public, la vente de données n'est pas envisageable. Dans le cadre bancaire, les aspects liés aux risques de la donnée sont plus prégnants que dans le secteur de la distribution par exemple. C'est la raison pour laquelle le modèle contient une relation binaire entre la dimension et le contexte caractérisée par un poids qui peut être nul quand la dimension n'est pas mobilisable dans le contexte et qui permet, le cas échéant, de combiner différentes dimensions pour une évaluation globale. Ainsi ces poids permettent d'annuler l'impact d'un critère ou sous-critère ou de le nuancer quand, par exemple, un risque est moindre qu'un autre dans un contexte donné.

L'*estimation* de la valeur est une relation ternaire entre la donnée, la dimension et la *méthode d'évaluation*. Ses attributs principaux sont la procédure de calcul et la valeur calculée. Comme nous l'avons remarqué dans l'état de l'art, il existe un nombre important de méthodes d'évaluation de la donnée. Les méthodes d'évaluation appartiennent à des catégories complémentaires. Cinq catégories sont mobilisées dans notre processus d'évaluation : les approches fondées sur le revenu, celles fondées sur les coûts, les approches de type marché, les méthodes d'analyse de risque et les approches par externalités (Coyle et Diepeveen, 2021).

Une des contributions principales de ce modèle, au-delà de la prise en compte du risque combiné avec le coût et la valeur, est l'intégration du contexte dans le processus d'évaluation de la valeur.

Fondé sur ce modèle conceptuel, nous présentons dans le paragraphe suivant l'arborescence multicritères qui guide le processus d'évaluation. Elle permet au décideur une vision plus holistique de l'évaluation de la valeur et la possibilité de tester voire de combiner plusieurs scénarii.

3.3. La méthode d'évaluation

L'évaluation de la valeur des données permet aux entreprises d'avoir une vue d'ensemble de leur patrimoine informationnel. Elle facilite la comparaison de la valeur des données (actifs intangibles) avec celle des actifs tangibles. Elle permet également de comparer la valeur des données pour différents cas d'usage. Nous avons choisi de représenter le modèle d'aide à la décision sous la forme d'une hiérarchie de critères (Saaty, 1994). Les raisons principales de ce choix résident dans la capacité de la démarche à structurer de manière hiérarchique le problème d'évaluation de la valeur de la donnée, qui est complexe du fait des nombreux critères et sous critères qui interviennent dans cette évaluation. De plus, la démarche permet une comparaison binaire des différentes alternatives en structurant les

priorités représentées par des poids. Enfin, la démarche permet une analyse de sensibilité plus aisée. En effet, les critères et les sous-critères peuvent avoir des poids variables et leur nombre n'est pas limité. Il est donc possible au décideur de modifier la valeur d'un critère ou d'ajouter ou de supprimer des critères. Pour des raisons d'espace nous ne présentons que les deux premiers niveaux de la hiérarchie (Fig. 2). Les autres niveaux sont détaillés dans les tableaux 1, 2 et 3.

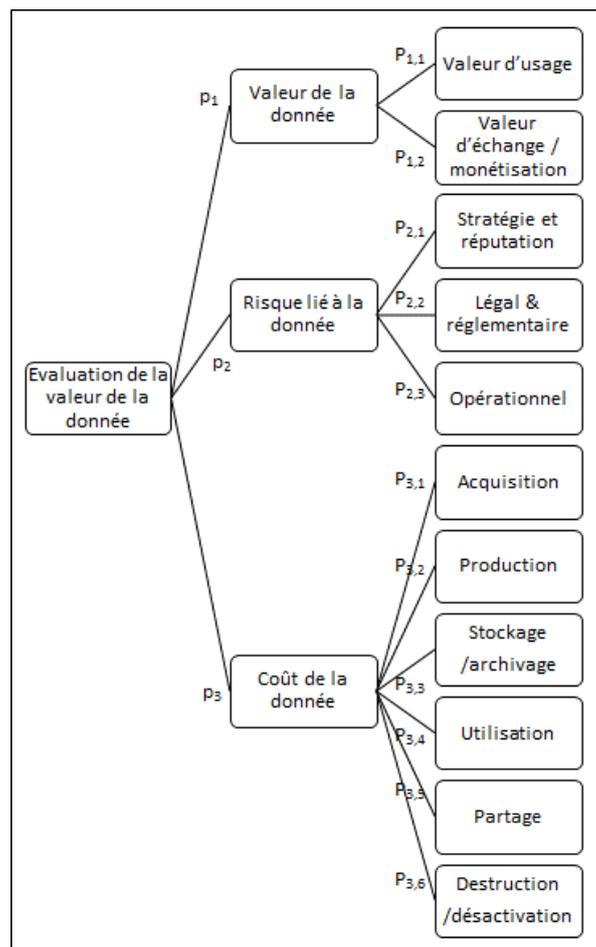


Figure 2. La hiérarchie de critères

Chaque nœud terminal de la hiérarchie de la figure 2 est lui-même décomposé en sous-critères. Puis plusieurs scénarii sont proposés pour faciliter l'évaluation de ces critères. Enfin, un exemple de métrique est associé à chaque scénario.

Tableau 1. Décomposition de la hiérarchie (dimension Valeur)

Sous-critère	Décomposition du sous-critère	Exemple de scénario d'illustration	Exemple de métrique	
Valeur d'usage (nb d'utilisations, utilisations prévues, utilisations effectives, etc.)	améliorer l'efficacité interne (différentiel d'efficacité)	améliorer les processus	temps gagné dans la livraison d'un produit	
		optimiser les tâches	temps gagné dans l'exécution de la tâche	
		réduire les coûts fournisseurs	augmentation du ratio de "gros" fournisseurs	
		réduire les coûts de production	augmentation juste-à-temps	
		optimiser les RH	diminution du délai de recrutement	
	se développer par croissance interne (augmentation parts de marché)	pénétrer le marché	augmentation chiffre d'affaires produits	
		étendre le marché	chiffre d'affaires nouveaux marchés	
		étendre les produits/services	chiffre d'affaires nouveaux produits/services	
		diversifier les produits/services	vente d'études fondées sur les données	
	développer par croissance externe	nouer des alliances	part de marché additionnelle	
		procéder à des fusions/acquisitions	économie d'échelle	
	Valeur d'échange / monétisation (chiffre d'affaires)	vendre/louer des données (chiffre d'affaires)	vendre données de consommation	chiffre d'affaires réalisé
			vendre données métier	
vendre données de déplacement				
vendre données de navigation Internet				
donner des données (externalité)		mettre à disposition données ouvertes	amélioration du bien commun	
		mettre à disposition données ESG		

Les tableaux 1, 2 et 3 décrivent successivement les branches Valeur, Risque et Coût ainsi décomposées. La première colonne de chaque tableau reprend les sous-critères déjà décrits. Ainsi, par exemple, la dimension Valeur (Tableau 1) se décline en deux familles liées respectivement à l'utilisation de la donnée (valeur d'usage) et sa monétisation (valeur d'échange). La seconde colonne décompose chaque famille à un niveau plus fin. Ainsi, l'usage de la donnée peut améliorer l'efficacité interne et/ou permettre la croissance interne de l'entreprise par développement de ses produits/services et marchés et/ou encore contribuer à sa croissance externe.

Le tableau 2 décrit la dimension Risque qui se décline en trois familles liées respectivement à l'impact négatif de la donnée sur la stratégie et la réputation, à l'aspect légal et réglementaire et en termes de risques opérationnels.

Tableau 2. Décomposition de la hiérarchie (dimension Risque)

Sous-critère	Décomposition du sous-critère	Exemple de scénario d'illustration	Exemple de métrique
Stratégie et réputation	ESG (Environnement-Social-Gouvernance)	augmenter l'empreinte carbone	variation du taux de respect des engagements carbone
		faire de la discrimination via les données	mesure du risque psychosocial par enquête
		diffuser des données erronées sur la gouvernance	variation de l'indice de satisfaction des actionnaires
	Ethique	diffuser des informations trompeuses	mesure de l'évolution de l'image de l'entreprise
		informer de manière non objective sur les produits	mesure de l'évolution de la perception des produits
	Confiance	diffuser des informations erronées	variation de l'indice de confiance client
		subir des fuites d'informations sensibles	mesure de l'évolution de l'image de l'entreprise
	Prise de décision	décider avec des données erronées	différentiel de performance globale
		décider avec des données obsolètes	
		prendre des décisions en l'absence de données	
Légal & réglementaire	Conformité à la loi	diffuser des informations discriminatoires	mesure financière du risque juridique encouru
		diffuser des informations confidentielles	
	Conformité aux réglementations	enfreindre le RGPD	mesure de la conséquence financière
		enfreindre le Protection of Personal Information Act (APPI)	
Propriété intellectuelle	brevet	mesure de la conséquence financière	
	enfreindre les droits d'auteur		
Opérationnel	Personnes	commettre des malveillances dans les données	évaluation des pertes attendues et des pertes exceptionnelles
		commettre des erreurs dans les données	
	Processus	ralentir les processus à cause de données manquantes	
		perturber les processus à cause de données erronées	
	Systèmes et technologies	diffuser des informations non autorisées	
interrompre des systèmes à cause de données manquantes			

A titre d'exemple, l'aspect légal et réglementaire est lui-même décomposé en la conformité aux lois, à la réglementation et à la propriété intellectuelle. Chaque composant est lui-même illustré par plusieurs scénarii. Ainsi la conformité aux

réglementations est illustrée par les scénarii « enfreindre le Règlement Général sur la Protection des Données (RGPD) » et « enfreindre le règlement japonais Protection of Personal Information Act (APPI) ».

Tableau 3. Décomposition de la hiérarchie (dimension Coût)

Sous-critère	Décomposition du sous-critère	Exemple de scénario d'illustration	Exemple de métrique
Acquisition	Données externes	payer les données	prix d'achat
		collecter des données (enquête)	coût de l'enquête externe
	Données internes	saisir des données	nombre d'heures*coût heure saisie
		collecter des données (enquête interne)	coût de l'enquête interne
Production	Données opérationnelles	contrôler les données	effort de vérification
		corriger les données erronées	effort de correction
	Données décisionnelles	nettoyer les données pour l'analytique	effort de nettoyage
		transférer les données de l'opérationnel vers le décisionnel	coût de maintenance des processus de flux
Stockage/archivage	Stockage des données	stocker les données sur site	coût total * part volume
		stocker les données sur le cloud	montant contrat
	Sécurisation des données	sauvegarder les données	coût des processus de sauvegarde
		anonymiser les données sensibles	effort d'anonymisation
Utilisation	Données opérationnelles	traiter les données opérationnelles	quote-part des frais IT
		vérifier les données traitées	effort de vérification
	Données décisionnelles	visualiser les données	coût des processus de visualisation
		calculer les indicateurs	quote-part des efforts de business analytics
Partage	Acheminement par courrier	acheminer l'information par voie postale	frais d'affranchissement
		acheminer l'information par email	quote-part coût serveur de messagerie
	Mise à disposition	virtualiser les données	coût architecture de virtualisation
		maintenir l'architecture de partage des données	coût architecture de données
	Sites web et plateformes électroniques	mettre à jour les contenus électroniques	coût humain et technique gestion des contenus
		mettre à disposition l'information sous forme accessible	coût de mise en conformité et accessibilité
Destruction/désactivation	Données sensibles	écraser les données	coût destruction "forte"
	Données non sensibles	effacer les données	coût standard destruction

Puis, la perte de confiance liée à un incident relatif aux données se traduit par une métrique de variation de l'indice de confiance client. L'importance du risque n'est

pas nécessairement la même selon le domaine d'activité de l'organisation. C'est pourquoi toutes ces mesures doivent être pondérées en fonction du contexte, comme expliqué plus loin.

Enfin, le tableau 3 détaille la dimension Coût de la donnée selon son cycle de vie. La colonne suivante décrit les caractérisations de la donnée permettant ensuite de décliner les différents scénarii, composants du coût, puis d'associer des métriques (dernière colonne). Ainsi, la destruction des données se décline en scénarii qui diffèrent en fonction de la sensibilité des données (effacer les données ou écraser les données). Les métriques de coût associées (coût destruction forte et coût standard) sont spécifiques.

Une des originalités de l'approche est de combiner cinq familles de méthodes d'évaluation de la valeur de la donnée. Ces familles sont matérialisées par la nuance de couleur de la deuxième colonne des trois tableaux : blanc pour les approches fondées sur le revenu, gris clair pour celles fondées sur les coûts, tacheté pour les approches de type marché, gris foncé pour les méthodes d'analyse de risque et hachuré pour les approches par externalités. Même si les approches utilisées proviennent de familles différentes, toutes les métriques sont financières, ce qui permet de les comparer, de les combiner avec des pondérations choisies et de les agréger pour construire un indicateur composite de la valeur.

La procédure d'évaluation (Saaty 1994) comprend donc les phases suivantes :

- Pour chaque nœud de la hiérarchie, à l'exclusion de la racine, le contexte détermine l'importance de chaque critère.
- Nous procédons ensuite à l'évaluation proprement dite de chaque critère se trouvant sur les nœuds terminaux de la hiérarchie.
- Puis nous évaluons la valeur de tous les autres nœuds inclus la racine de l'arborescence en procédant au calcul de la somme pondérée.

A noter qu'à certains nœuds de l'arborescence, peuvent être associées plusieurs métriques. Il peut s'agir ainsi de mesures résultant de l'agrégation des mesures des niveaux inférieurs de l'arborescence ou de mesures alternatives. Dans ce cas, il faut soit choisir une des mesures, soit les combiner. A titre d'exemple, la valeur d'usage des données peut être mesurée par le nombre d'utilisations effectives de celles-ci ou par l'évaluation de l'efficacité accrue de l'organisation grâce à ces utilisations.

3.4. La validation

La contribution de cet article consiste en une approche fondée sur trois artefacts : un construit (concept de valeur de la donnée), un modèle conceptuel et une méthode d'évaluation de la valeur. Nous proposons de valider l'approche par quelques cas d'usage décrits ci-après.

Une chaîne de salons d'esthétique a acheté des données décrivant les comportements d'achat des consommateurs dans le secteur de la beauté (Pwc, 2019). En étudiant ces

données, l'entreprise a pu corréler l'âge des clients et le type de campagnes marketing auxquelles ceux-ci sont les plus réceptifs. Ainsi, elle a recruté des influenceurs sur les réseaux sociaux pour atteindre les femmes de moins de 30 ans et faire la promotion des produits cosmétiques et de maquillage. Pour cibler les femmes de plus de 50 ans, elle a fait appel à des actrices célèbres et mis en avant des traitements de raffermissement de la peau. Par cette segmentation, l'entreprise a augmenté de 20% son chiffre d'affaires annuel. Cette évaluation de la valeur correspond à la mesure du différentiel en parts de marché obtenu par une meilleure pénétration de celui-ci (composant *pénétrer le marché* de l'action de *croissance interne* via l'utilisation de données – *valeur d'usage*). L'utilisation de notre approche consiste à enrichir cette dimension Valeur en intégrant d'une part les risques potentiels (par exemple, l'utilisation de données non conformes au RGPD) et d'autre part les coûts associés (acquisition des données, production des informations incluant la préparation, analyse, destruction, etc.). Cet exemple valide ainsi partiellement notre approche et en montre aussi la valeur ajoutée par la prise en compte d'une évaluation holistique.

Une société californienne de tests ADN a vendu un accès exclusif à sa base de données, comprenant les génomes de plus de 5 millions de personnes à un géant pharmaceutique pour 300 millions de dollars (PwC, 2019). La base de données peut être utilisée pour la recherche et le développement de nouveaux médicaments. Grâce à cette transaction, on peut ainsi estimer qu'une base de données génétiques comparable peut être vendue à 60 dollars par ligne de données. Cet exemple corrobore la métrique utilisant le chiffre d'affaires réalisé grâce à la vente d'un jeu de données (composant *vendre des données métier* de l'action de *vendre/louer des données* via la monétisation des données – *valeur d'échange*). Toutefois, notre approche permet de mettre l'accent sur le risque important tant en conformité à la réglementation RGPD qu'à l'éthique ou la confiance qui pourraient être compromises par cette vente si l'anonymisation complète des données n'était pas garantie. Cette anonymisation est particulièrement coûteuse quand elle concerne des données génétiques quasi-identifiantes. Ainsi notre approche est validée en ce qui concerne la monétisation et l'exemple montre sa contribution supplémentaire dans l'évaluation des coûts et des risques induits.

Ces deux cas illustrent l'approche et constituent les premiers pas d'une validation de son utilité. L'arborescence guide l'évaluation en parcourant depuis la feuille jusqu'à la racine. Puis le parcours de haut en bas permet d'amplifier le raisonnement de l'évaluateur en sollicitant les branches complémentaires pour une évaluation plus holistique. D'autre part, ils nous ont permis de valider la complétude du modèle conceptuel et, par-là, son utilité.

La validation de notre approche auprès d'experts nous a permis à la fois de souligner l'intérêt de combiner l'approche par les risques et celles par les coûts et la valeur mais aussi de mettre le doigt sur la difficulté d'assurer une cohérence dans l'évaluation quand celle-ci est effectuée par des parties prenantes de culture différente. Ainsi, l'appréhension du risque est différente par les juristes et par les

experts du marketing. Le cabinet de conseil que nous avons interrogé est surtout intéressé par l'approche qui structure sa pensée et lui permet d'améliorer sa proposition de valeur pour des clients ou prospects. Le directeur des données d'une grande entreprise d'assurance que nous avons aussi confronté à l'approche a, quant à lui, vu un moyen convaincant de présenter à ses dirigeants ses demandes de budgets tant pour valoriser les données que pour gérer les risques associés.

4. Conclusions et recherches futures

Les principales contributions de cet article sont :

- la définition d'un concept de valeur de la donnée intégrant les trois dimensions du risque, du coût et de la valeur,
- la proposition d'un modèle conceptuel qui relie l'ensemble des concepts participant au processus d'évaluation de la valeur des données, et notamment le contexte interne et externe de l'organisation concernée,
- le développement d'une arborescence hiérarchique multicritère qui sert de base à un système d'aide à l'évaluation holistique de la valeur des données d'une organisation,
- la validation de l'approche à l'aide de cas d'usage fondés sur des cas réels décrits dans la presse professionnelle. Cela ne constitue qu'une première étape dans la validation de l'approche proposée.

Il est bien entendu qu'on ne peut prétendre à l'exhaustivité dans ce type d'approche d'évaluation d'un concept très évolutif. Toutefois, le modèle proposé est extensible sans contrainte. L'introduction d'une nouvelle dimension, d'un nouveau critère ou sous-critère n'impacte que l'éventuelle adaptation des poids associés aux autres nœuds du même niveau de l'arborescence. Cet arbre pondéré permet aussi de garantir l'adaptabilité de l'approche à toutes les organisations, publiques ou privées, quel que soit le secteur, au prix de la redéfinition des poids.

Nous prévoyons notamment des ateliers pour compléter les scénarii et l'application de la démarche à plusieurs exemples réels et le développement d'un outil d'automatisation de la démarche facilitant le processus de décision des parties prenantes de l'entreprise. Ces étapes permettront une validation conjointe et un raffinement des différents artefacts de l'approche (concept, modèle, méthode). Un autre axe de recherche future sera d'affiner la description des données (domaine, nature, qualité) en s'appuyant, le cas échéant, sur les normes existantes. Cette approche ainsi confortée pourra utilement alimenter l'évaluation des entreprises en cas de rachat et contribuer à la base de connaissances sur les stratégies de valorisation des actifs informationnels des entreprises.

Remerciements. Les auteurs remercient les partenaires de la Chaire Stratégie et Gouvernance de l'Information de l'ESSEC, au sein de laquelle cette recherche a été réalisée.

Bibliographie

Aguilar, F. J. (1967). *Scanning the business environment*. Macmillan.

Antuca, A., Noble, R. (2021). Data: how it affects competitive dynamics, how to value it, and whether to provide third-party access to it. *Competition Law Journal*, 20(2), 102-110.

Attard, J., Brennan, R. (2018). A Semantic Data Value Vocabulary Supporting Data Value Assessment and Measurement Integration. *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS)*, pages 133-144.

Bodendorf, F., Dehmel, K., Franke, J (2022). Scientific Approaches and Methodology to Determine the Value of Data as an Asset and Use Case in the Automotive Industry, *Proceedings of the 55th Hawaii International Conference on System Sciences*.

Ciuriak, D. (2019). Unpacking the Valuation of Data in the Data-Driven Economy, *Conference on Global Data Law*, New York, 26-27.

Corrado, C. (2019) Data as an Asset: Expanding the Intangible Framework, *Conference on the Economics, Governance and Management of AI, Robots and Digital Transformation (EMAAE)*.

Coyle, D., Diepeveen, S. (2021). "Creating and governing social value from data." Available at SSRN 3973034.

European Commission (2020). *Data Governance Act, Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) — COM/2020/767 final*. Brussels, Belgium: European.

Federal Data Strategy (2022). *Data, accountability, and transparency: creating a data strategy and infrastructure for the future*, <https://strategy.data.gov>, accédé le 26.02.2022.

Garifova, L.F. (2015). Infonomics and The Value of Information in The Digital Economy, *Procedia Economics and Finance* 23 (pp. 738 – 743).

Moody, D., Walsh, P. (1999). Measuring The Value Of Information: An Asset Valuation Approach, *European Conference on Information Systems*, ECIS'99, pp.1-17.

Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises, *Information Systems Management*, 32 (pp. 234–251).

PwC (2019). *Putting a value on data*, <https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf>.

Saaty, T. L. (1994). How to make a decision: the analytic hierarchy process. *Interfaces*, 24(6), (pp. 19-43).

Savona, M. (2019). *The Value of Data: Towards a Framework to Redistribute It*, SPRU Working Paper Series SWPS 2019-21 (Octobre 2019).

Short, J., Todd, S. (2017). What's your data worth? *MIT Sloan Management Review*, Spring.

Tallon, P. P. (2013). Corporate governance of big data: Perspectives on value, risk, and cost. *Computer*, 46(6), (pp. 32-38).

Wang, Y., Zhao, H. (2020). Data Asset Value Assessment Literature Review and Prospect, *Journal of Physics: Conference Series*, 1550.

Wdowin, J., Diepeveen, S. (2020). *The Value of Data – Literature review*, Bennett Institute for Public Policy, Cambridge.

World Economic Forum (2021). *Data for Common Purpose: Leveraging Consent to Build Trust*, White paper, Novembre 2021.

Zeiter, A., Hagel, J., Snyder, S (2021). *Articulating Value from Data*, White Paper, World Economic Forum, Novembre 2021.