

Une approche arborescente d'évaluation de la valeur des données

A hierarchical approach to assessing the value of data

Jacky Akoka¹, Isabelle Comyn-Wattiau²

¹ Laboratoire Cedric, Conservatoire National des Arts et Métiers, France, jacky.akoka@lecnam.net

² ESSEC Business School, France, wattiau@essec.edu

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-CLÉS. 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. L'approche proposée est générique, dans la mesure où elle permet de tester différents scénarios, tant au niveau de la stratégie d'exploitation des données qu'au niveau de la méthode d'évaluation de leur valeur. L'approche hiérarchique multicritère structure la connaissance sous forme arborescente et quantifie les scénarios à comparer.

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 l'approche fondée sur trois piliers, un concept, un modèle et une méthode. La quatrième section propose une validation de l'ensemble de l'approche. Enfin, la cinquième 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 [SAV 19, COR 19, OTT 15, WDO 20]. Ces actifs sont soumis à des réglementations et évalués à l'aide de techniques spécifiques [CIU 19]. 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 [SHO 17].

Les données ont un coût parfois considéré comme irrécupérable [WAN 20] 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 [ZEI 21]. 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 [ATT 18]. 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 19]. Toutefois, la qualité et la quantité des données sont soumises à la loi des rendements décroissants [MOO 99]. 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 [WOR 21] 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 [EUR 20] ou encore le plan d'action de l'US Federal Data Strategy [FED 22] qui vise à mieux exploiter les données en tant qu'actif stratégique.

Plusieurs méthodes d'évaluation de la valeur des données ont été proposées [MOO 99, CIU 19, WAN 20, OTT 15]. 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 [GAR 15] 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 [COY 21, ANT21]. 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 [BOD 22]. 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.

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 [TAL 13]. 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 sur 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. Tant pour la nature, pour la qualité que pour le domaine, des normes de représentation des données sont disponibles et pourraient faciliter la comparaison de différents jeux de données ou scénarios d'utilisation de celles-ci.

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, etc.) qui en résulte. La valeur d'échange traduit la possibilité de vendre, de louer voire de mettre à disposition gratuitement ses données. L'évaluation du risque et du coût sont nécessaires pour le calcul de la valeur nette.

Le troisième concept est celui du *contexte*. Ce dernier joue un rôle important dans le processus d'évaluation de la valeur des données. Il existe un lien et une certaine interdépendance entre le contexte et les processus d'évaluation. Le contexte peut être interne à l'entreprise qui évalue les données. La structure organisationnelle, la culture et la gouvernance de l'entreprise jouent alors un rôle important dans le processus d'évaluation. Ainsi, un organisme public n'obéit pas aux mêmes règles 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) [AGU 67]. 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.

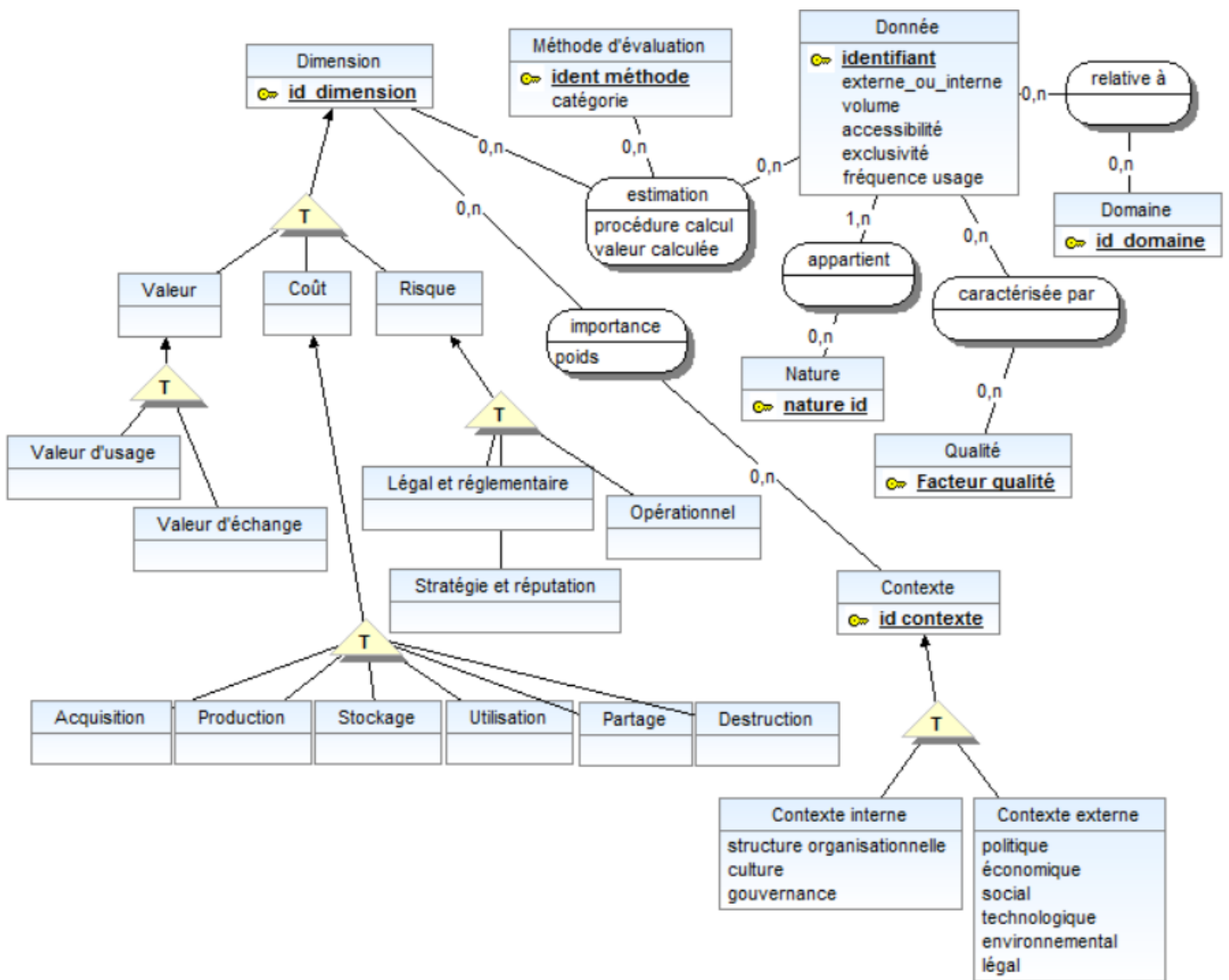


Figure 1. Le modèle conceptuel de données

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 [COY 21].

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 [SAA 94]. 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. La figure 2 décrit les deux niveaux de critères d'évaluation de la valeur. Les autres éléments de l'approche, dont les scénarios et les métriques sont décrits dans les tableaux 1, 2 et 3.

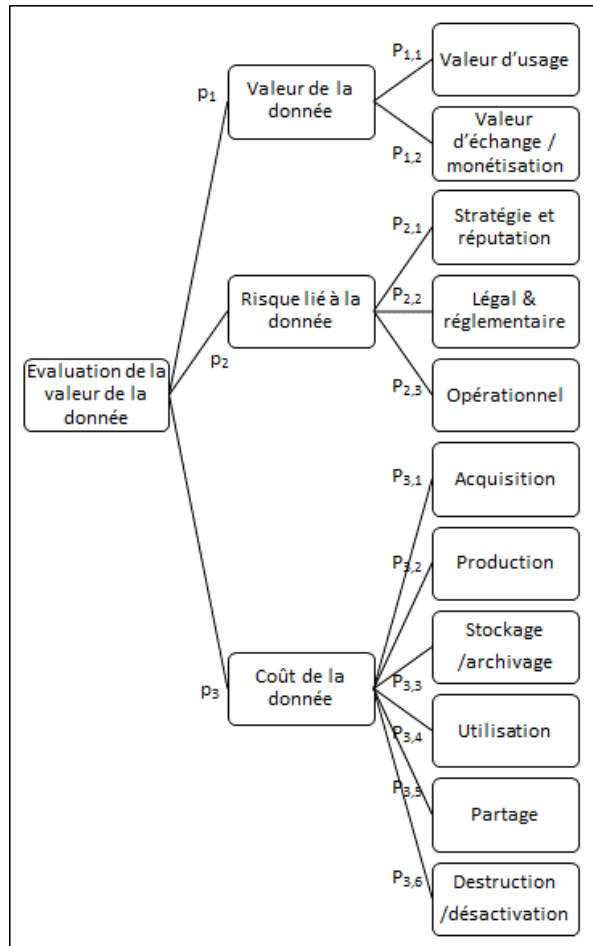


Figure 2. La hiérarchie de critères

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

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 présentés à la figure 2. 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.

Sous-critère	Décomposition du sous-critère	Exemples de scénarios d'illustration	Exemple de métriques	
Valeur d'usage	Amélioration de l'efficacité interne	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 du juste-à-temps	
		Optimiser les RH	Diminution du délai de recrutement	
	Développement par croissance interne	Pénétrer le marché	Augmentation du chiffre d'affaires produits/services	
		Etendre le marché	Chiffre d'affaires nouveau marché	
		Etendre les produits/services	Chiffre d'affaires nouveaux produits/services	
		Diversifier les produits/services	Chiffre d'affaires nouveaux produits/services, exemple : vente d'études fondées sur les données	
	Développement par croissance externe	Nouer des alliances	Part de marché additionnelle	
		Procéder à des fusions/acquisitions	Economie d'échelle	
	Valeur d'échange	Vente/location de données	Vendre des données métier	Chiffre d'affaires réalisé
			Vendre des données de consommation	
			Vendre des données de navigation Internet	
Vendre des données de déplacement				
Mise à disposition de données		Mettre à disposition des données ouvertes	Amélioration du bien commun	
		Mettre à disposition des données ESG		

Tableau 1. Décomposition de la hiérarchie (Dimension Valeur)

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. Ces risques peuvent être d'importance très différente, c'est ce que le poids associé à la branche permettra de prendre en compte. A titre d'exemple, les données personnelles sont, dans la plupart des pays, soumises à des réglementations ou des lois qui prévoient des sanctions importantes en cas de non-respect. Par voie de conséquence, le risque légal ou réglementaire lié aux données personnelles sera affecté d'un poids plus important que celui associé à des données météorologiques.

Sous-critère	Décomposition du sous-critère	Exemples de scénarios d'illustration	Exemple de métriques
Stratégie et réputation	Responsabilité sociale de l'entreprise (RSE)	Augmenter l'empreinte carbone	Variation du taux de respect des engagements carbone
		Faire de la discrimination via les données	Mesure du risque psycho-social
		Diffuser des informations erronées sur la gouvernance	Variation de l'indice de satisfaction des actionnaires
	Ethique	Diffuser des informations trompeuses	Evolution de l'image de l'entreprise
		Informers de manière non objective sur les produits	Evolution de la perception des produits
	Confiance	Diffuser des informations erronées	Variation de l'indice de confiance client
		Subir des fuites d'informations sensibles	Evolution de l'image de l'entreprise
	Prise de décision	Décider avec des informations erronées	Différentiel de performance globale
		Décider avec des informations obsolètes	
		Prendre des décisions en l'absence de données	
Légal et 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	
	Propriété intellectuelle	Enfreindre la propriété d'un brevet	
		Enfreindre les droits d'auteur	
Opérationnel	Personnes	Commencer des malveillances dans les données	Evaluation des pertes attendues et des pertes exceptionnelles
		Commencer 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	

Tableau 2. Décomposition de la hiérarchie (Dimension Risque)

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) ». 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, bleu pour les approches de type marché, gris foncé pour les méthodes d'analyse de risque et vert 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 combinaison des méthodes se produit lors de l'agrégation des notes puisque chaque famille correspond à une branche plus ou moins profonde de l'arbre.

La procédure d'évaluation [SAA 94] 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.

Sous-critère	Décomposition du sous-critère	Exemples de scénarios d'illustration	Exemple de métriques
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
		Restaurer les données perdues	Coût humain et technique de la restauration
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
		Appliquer des algos (ML, data mining, etc.)	Coût de la mise en œuvre des algos
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	Données sensibles	Ecraser les données	Coût destruction « forte »
	Données non sensibles	Effacer les données	Coût standard destruction

Tableau 3. Décomposition de la hiérarchie (Dimension Coût)

Nous présentons ci-dessous la procédure d'évaluation de la valeur d'un ensemble de données.

```

Procédure d'évaluation d'un ensemble de données
/* pij est le poids du sous-critère j associé au critère i
/* valeurij est l'estimation de la valeur de l'ensemble de données
/* pour le sous-critère j associé au critère i
  Valeur(ensemble_données) : réel ;
  Valeur :=0 ;
  Pour i :=1 à n
  Pour j :=1 à m
  Valeur :=Valeur+pij*valeurij
  Fin pour
  Fin pour
  Renvoyer Valeur

```

Dans cet algorithme, on considère qu'on a les estimations liées à chaque sous-critère. On peut avoir à décomposer à un niveau plus fin en décomposant à nouveau le sous-critère et en évaluant l'ensemble des scénarii associés à chacun des nœuds ainsi obtenus, comme le montrent certaines branches de la hiérarchie.

4. La validation de l'approche

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.

4.1. Pertinence de l'approche d'évaluation

Considérons le cas d'une chaîne de salons d'esthétique qui a acheté des données décrivant les comportements d'achat des consommateurs dans le secteur de la beauté [PWC 19]. En étudiant ces données, l'entreprise a pu corréliser 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 raffermisssement 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 (scénario « pénétrer le marché » de la branche « développement par croissance interne » via la 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.

Considérons à présent le cas d'une société californienne de tests ADN qui 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 19]. 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 (scénario « vendre des données métier » de la branche « vendre/louer des données » via la 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.

4.2. Utilité de l'approche d'évaluation

La section précédente avait pour but de confronter certains scénarii de notre hiérarchie à des exemples publiés d'évaluation de la valeur des données. Pour des raisons d'espace, nous ne pouvons pas illustrer toutes les branches de la hiérarchie. Dans cette section, nous décrivons comment l'approche peut être utilisée *a posteriori* pour déduire les poids en fonction d'un contexte, puis *a priori* pour estimer à l'avance la valeur qu'on peut anticiper sur la base de ces poids.

4.2.1. Evaluation *a posteriori*

Reprenons l'exemple réel mentionné plus haut du directeur marketing qui doit entreprendre les actions permettant, à budget constant, d'améliorer le chiffre d'affaires de son entreprise. A cet effet, il demande au Chief Data Officer si ce dernier peut l'aider à imaginer des moyens fondés sur les données pour atteindre son objectif. Ensemble, ils élaborent un scénario consistant à acheter des données décrivant l'exposition aux différents media des clients potentiels. Après mise en œuvre du scénario, une amélioration de 50000 euros du chiffre d'affaires est constatée. Ils utilisent l'arborescence pour décliner les coûts visibles et cachés afin de déduire la valeur ajoutée réelle de l'opération d'achat des données.

L'utilisation de l'approche dans ce cadre consiste à décliner les différents éléments contribuant à la valeur sans leur affecter de poids, avec les bénéfices tirés de l'exploitation des données, diminués des coûts associés et tenant compte des risques (Tableau 4).

Ainsi, en matière de risque, la réputation de l'entreprise peut être impactée par l'augmentation de l'empreinte carbone suspectée par ce traitement de données. On estime à 1000 euros la perte de chiffre d'affaires du fait des clients qui se détournent pour cette raison. Le fait que l'entreprise ait discriminé, dans sa campagne marketing, les femmes selon leur âge peut déclencher, en interne, une réaction des salariées. La perte d'attractivité des "talents" découlant de cette image négative est évaluée à 1500 euros. De la même façon, cette image négative peut être perçue par les clients. On évalue à 3000 euros la conséquence sur le chiffre d'affaires. On suppose aussi que les données ainsi achetées et insuffisamment anonymisées peuvent générer une dégradation de l'image dont l'impact sur le chiffre d'affaires est évalué à 1000 euros. On considère qu'il n'y a pas de risque légal ni réglementaire puisque les données sont achetées contractuellement, dégageant la responsabilité de l'entreprise dans ce domaine. En ce qui concerne un éventuel risque opérationnel lié à des erreurs ou malveillances dans les données, on l'estime nul puisque le manque à gagner potentiel n'est pas intégré dans l'évolution du chiffre d'affaires constaté. Les données manquantes ou erronées ont pu générer une activité supplémentaire ayant par rebond pénalisé les autres processus, ce qui est estimé à 2000 euros.

Dimension	Sous-critère	Scénario d'illustration	Evaluation
Valeur	Valeur d'usage	Pénétrer le marché	50000 €
Risque	Stratégie et réputation	Augmenter l'empreinte carbone	-1000 €
		Faire de la discrimination via les données	-1500 €
		Discriminer les clients par l'utilisation d'information sur eux	-3000 €
		Subir des fuites d'informations sensibles	-1000 €
Coût	Opérationnel	Ralentir les processus à cause de données manquantes	-1000 €
		Perturber les processus à cause de données erronées	-1000 €
	Acquisition	Payer les données	-3000 €
	Production	Nettoyer les données pour l'analytique	-500 €
	Stockage	Stocker les données sur le cloud	-500 €
	Utilisation	Visualiser les données	-1000 €
		Calculer les indicateurs	-1000 €
		Appliquer des algos	-1000 €
	Destruction	Effacer les données	-100 €
RESULTAT			34400 €

Tableau 4. Application de l'approche pour l'évaluation a posteriori

Les coûts liés à ce jeu de données se répartissent en :

- un coût d'acquisition estimé à 3000 euros,
- un effort de nettoyage de ces données entraînant un coût de 500 euros,
- un coût de stockage sur le cloud de 500 euros, incluant le coût de sauvegarde,
- un coût d'utilisation : grâce aux données achetées, l'entreprise peut mettre en œuvre un ensemble d'algorithmes, sophistiqués ou non, pour détecter les segments de clientèle, ici la découverte de règles d'association entre l'âge de la cliente et les types de produits consommés, la visualisation des résultats pour l'équipe marketing qui prépare le plan média et le calcul d'indicateurs pour répartir le budget par média, l'ensemble de ces trois postes est évalué à 3000 euros,
- enfin le coût de destruction standard du fichier après traitement est estimé à 100 euros.

Il y a plusieurs lectures possibles de ce tableau. La première consiste à ramener le chiffre d'affaires additionnel de 50000 euros à 34400 euros, une fois déduits les coûts de gestion des données et l'impact financier des risques encourus. La seconde consiste à adopter le point de vue de la Direction des Systèmes d'Information (DSI), considérée comme un centre de profits et non un centre de coûts. Dans ce cas, la même estimation a posteriori contribue à l'obtention d'un Retour sur investissement (ROI) de 50000 euros grâce à l'effort de 7100 euros pour acheter et gérer ces données. Enfin, dans les deux lectures, les risques encourus sont estimés à 8500 euros s'ils sont avérés. Ces risques sont de nature très différente et la probabilité que tous soient avérés n'est pas significative. Néanmoins leur prise en compte globale permet une analyse exhaustive.

Cette utilisation de l'arborescence pour cette analyse *a posteriori* n'est pas à proprement parler une analyse hiérarchique multicritères classique puisqu'elle n'associe pas de poids à chaque branche pour comparer plusieurs scénarios *a priori*. En revanche, cette analyse *a posteriori* présente l'avantage supplémentaire de générer de possibles poids pour une analyse similaire *a priori*, comme illustré dans la suite.

4.2.2. Evaluation *a priori*

En appliquant la hiérarchie *a posteriori* à différentes reprises sur différents scénarios dans différents contextes, on peut capitaliser la connaissance ainsi acquise sous la forme de poids, à chacun des niveaux de l'arborescence. Ainsi, l'estimation décrite auparavant conduit à trois montants : 7100 euros de coûts, 8500 euros en termes de risques et, par voie de conséquence, 34400 euros de valeur ajoutée constituent les 50000 euros mesurés sur le chiffre d'affaires. Ces trois montants représentent respectivement un poids de 14%, 17% et 69% dans la contribution à la valeur de la donnée. En appliquant le même raisonnement aux différents niveaux de la hiérarchie, on peut décliner un ensemble de poids pour toute l'arborescence. Cela peut se faire par induction en généralisant des estimations *a posteriori* ou en faisant appel à des experts (Tableau 5).

Dimension (poids)	Sous-critère (poids)	Achat de données	Exploitation des données internes
Valeur (69%)	Valeur d'usage (80%)	50000 €	40000 €
	Valeur d'échange (20%)	0 €	0 €
	<i>Evaluation dimension Valeur</i>	40000 €	32000 €
Risque (17%)	Stratégie et réputation (70%)	6500 €	6500 €
	Légal et réglementaire (10%)	0 €	1000 €
	Opérationnel (20%)	2000 €	2000 €
	<i>Evaluation dimension Risque</i>	4900 €	5050 €
Coût (14%)	Acquisition (34%)	3000 €	2000 €
	Production (5%)	1000 €	500 €
	Stockage (5%)	500 €	500 €
	Utilisation (34%)	3000 €	2000 €
	Partage (20%)	0 €	0 €
	Destruction (2%)	100 €	100 €
	<i>Evaluation dimension Coût</i>	2092 €	1437 €
	<i>Evaluation globale</i>	26467 €	21020 €

Tableau 5. Utilisation de l'approche pour l'évaluation *a priori*

Cette arborescence pondérée peut ensuite être utilisée pour comparer deux scénarios, par exemple celui relatif à l'achat d'un ensemble de données et celui consistant à exploiter les données disponibles en interne. Dans le même tableau 5, on a comparé le scénario précédent d'achat de données externes à celui qui consisterait à exploiter des données internes dans le même but de développer le chiffre d'affaires. La comparaison conduit à une valeur de 26467 euros en cas d'achat et à 21020 euros en cas d'exploitation des données internes. Ces dernières ne représentent pas un coût vraiment moindre, au vu de tout le cycle de vie à gérer. Elles présentent un risque supérieur dans la mesure où l'aspect légal et

réglementaire n'est pas couvert pas un contrat d'achat. Enfin, leur exploitation ne mène pas à la même opportunité, ne s'appuyant que sur de la connaissance interne.

Cette comparaison illustre l'approche et constitue les premiers pas d'une validation de son utilité. L'arborescence parcourue depuis la feuille jusqu'à la racine guide l'évaluation. Plus généralement, la mise à disposition d'un ensemble riche mais ouvert de scénarios d'évaluation permet une prise en compte plus holistique de la valeur de la donnée, amplifiant le raisonnement de l'évaluateur. D'autre part, on a pu valider la complétude du modèle conceptuel qui peut servir de structure pour une base de partage de tous les scénarios d'évaluation de la donnée, au-delà de ceux qui ont pu être décrits jusqu'ici.

4.3. Valeur ajoutée de l'approche d'évaluation

Dans cette section, nous décrivons deux moyens de montrer la valeur ajoutée de l'approche. Le premier est un retour d'expérience d'experts que nous avons interrogés sur notre approche. Le second est une utilisation de l'arborescence pour anticiper une action.

4.3.1. Validation par des experts

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.

4.3.2. Anticiper pour agir

Sur la base du cas d'usage décrit plus haut où une entreprise pharmaceutique acquiert des données ADN pour 60 dollars par individu, nous illustrons un autre usage possible de l'arborescence. Imaginons un laboratoire européen privé qui souhaite disposer de telles données. Il fait l'acquisition d'un fichier pour 300 millions d'euros. L'objectif est de produire un médicament pour traiter une maladie génétique. Si l'on estime pouvoir obtenir un chiffre d'affaires d'un milliard d'euros par la vente de ce médicament. L'arborescence (résumée sur le Tableau 6) permet de calculer le coût total d'exploitation de ces données, achat inclus. Les poids sont adaptés au contexte de données complexes et volumineuses. Le total du coût est de 374 millions d'euros. Le laboratoire peut donc supporter un risque maximal de 1 milliard moins 374 millions, soit 626 millions d'euros. Les poids permettent d'anticiper les types de risque (stratégie et réputation, légal et réglementaire, opérationnel) correspondants. Enfin, cette estimation est un outil pour le laboratoire dans sa démarche auprès d'un assureur pour couvrir ce risque.

Dimension (poids)	Sous-critère (poids)	Achat de données ADN
Valeur (69%)	Valeur d'usage (100%)	1000 M€
	<i>Evaluation dimension Valeur</i>	1000 M€
Risque (17%)	Stratégie et réputation (70%)	440 ? M€
	Légal et réglementaire (10%)	62 ? M€
	Opérationnel (20%)	124 ? M€
	<i>Evaluation dimension Risque</i>	626 ? M€
Coût (14%)	Acquisition (80%)	300 M€
	Production (5%)	19 M€
	Stockage (3%)	11 M€
	Utilisation (10%)	37 M€
	Partage (0%)	0 M€
	Destruction (2%)	7 M€
	<i>Evaluation dimension Coût</i>	374 M€

Tableau 6. Application de l'approche pour l'évaluation a posteriori

Ainsi, l'approche peut être utilisée aussi pour anticiper une action d'assurance à mener avant le lancement d'un projet à haut risque. En conclusion, l'intérêt de l'approche mais aussi sa difficulté de mise en œuvre peut venir du fait que, comme l'ont illustré les trois scénarios décrits dans cet article, les poids sont tantôt utilisés en tant que tels, tantôt générés grâce au scénario, tantôt neutralisés.

5. 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.

Si l'exhaustivité n'est jamais garantie, en revanche, notre approche fondée sur un modèle et une méthode offre une agilité permettant, le cas échéant, de supprimer ou d'ajouter de nouveaux critères, de tester plusieurs scénarios, de modifier les poids des critères en fonction de la connaissance du domaine. Adopter une approche multicritère combinant risque, coût et valeur entraîne la nécessité de recourir à des experts de culture et de métiers différents. Le calibrage des différentes branches de l'arbre s'en trouve complexifié, d'où l'importance de compléter notre approche par une confrontation d'experts sur des cas réels. A cette fin, 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

- [AGU 67] AGUILAR F. J., *Scanning the business environment*. Macmillan, 1967.
- [ANT 21] ANTUCA A., NOBLE R., « Data: how it affects competitive dynamics, how to value it, and whether to provide third-party access to it », *Competition Law Journal*, 20(2), p. 102-110, 2021.
- [ATT 18] 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)*, p. 133-144, 2018.
- [BOD 22] Bodendorf F., Dehmel K., Franke J., « 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*, 2022.
- [CIU 19] Ciuriak D., « Unpacking the Valuation of Data in the Data-Driven Economy », *Conference on Global Data Law*, New York, p. 26-27, 2019.
- [COR 19] Corrado C., « Data as an Asset: Expanding the Intangible Framework », *Conference on the Economics, Governance and Management of AI, Robots and Digital Transformation (EMAEE)*, 2019.
- [COY 21] Coyle D., Diepeveen S., « Creating and governing social value from data », Available at SSRN 3973034, 2021.
- [EUR 20] European Commission, *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, 2020.
- [FED 22] Federal Data Strategy, *Data, accountability, and transparency: creating a data strategy and infrastructure for the future*, <https://strategy.data.gov>, accédé le 26.02.2022.
- [GAR 15] Garifova L.F., « Infonomics and The Value of Information in The Digital Economy », *Procedia Economics and Finance* 23, p. 738 – 743, 2015.
- [MOO 99] Moody D., Walsh P., « Measuring the Value Of Information: An Asset Valuation Approach », *European Conference on Information Systems, ECIS'99*, p.1-17, 1999.
- [OTT 15] Otto B., « Quality and Value of the Data Resource in Large Enterprises », *Information Systems Management*, 32, p. 234–251, 2015.
- [PWC 19] PwC, *Putting a value on data*, <https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf>, 2019.
- [SAA 94] Saaty T. L., « How to make a decision: the analytic hierarchy process », *Interfaces*, 24(6), p. 19-43, 1994.
- [SAV 19] Savona M., « The Value of Data: Towards a Framework to Redistribute It », *SPRU Working Paper Series SWPS 2019-21*, 2019.

- [SHO 17] Short J., Todd S., « What's your data worth? », *MIT Sloan Management Review*, Spring, 2017.
- [TAL 13] Tallon P. P., « Corporate governance of big data: Perspectives on value, risk, and cost », *Computer*, 46(6), p. 32-38, 2013.
- [WAN 20] Wang Y., Zhao H., « Data Asset Value Assessment Literature Review and Prospect », *Journal of Physics: Conference Series*, 1550, 2020.
- [WDO 20] Wdowin J., Diepeveen S., *The Value of Data – Literature review*, Bennett Institute for Public Policy, Cambridge, 2020.
- [WOR 21] World Economic Forum, *Data for Common Purpose: Leveraging Consent to Build Trust*, White paper, Novembre 2021.
- [ZEI 21] Zeiter A., Hagel J., Snyder S, *Articulating Value from Data*, White Paper, World Economic Forum, Novembre 2021.