
Gouvernance des données

Vers un cadre conceptuel

Jacky AKOKA¹, Isabelle COMYN-WATTIAU²

1 Laboratoire CEDRIC-CNAM, 2 Rue Conté, 75003 PARIS

2. ESSEC Business School, 3 Av. B. Hirsch, 95021 CERGY Cedex

RESUME. La gouvernance des données est une activité essentielle pour les organisations qui cherchent à exploiter les données comme un atout stratégique. Son objectif est de maximiser la valeur tout en minimisant les coûts et les risques. Dans cet article, nous présentons un cadre conceptuel pour la gouvernance des données offrant une vue holistique enrichissant les cadres et modèles académiques et professionnels existants. En utilisant des techniques bibliométriques, nous analysons la littérature existante afin d'identifier les éléments clés de la gouvernance des données, notamment sa structure intellectuelle, ses thèmes de recherche et les articles les plus influents qui en forment la colonne vertébrale. Dans un deuxième temps, nous proposons un cadre conceptuel enrichi fondé sur la théorie des systèmes. Ce cadre englobe cinq dimensions primordiales : le but, la structure, les activités, l'environnement et le résultat. Il permet également de prendre en considération l'interaction entre ces dimensions. Afin d'illustrer ce cadre conceptuel, nous décrivons comment celui-ci a permis de structurer les questions d'un baromètre dédié à l'évaluation de la maturité de la gouvernance des données dans les organisations. Nous présentons ensuite quelques cas d'usage, puis nous discutons des implications pour les chercheurs et les praticiens.

ABSTRACT. Data governance constitutes an indispensable activity for organizations intent on leveraging data as a strategic asset. Its objective is to maximize value while minimizing costs and risks. In this article, we present a conceptual framework for data governance that offers a holistic view, thereby enriching existing academic and professional frameworks and models. Using bibliometric techniques, we analyze the extant literature to identify the key elements of data governance, including its intellectual structure, research themes, and the most influential articles that form its backbone. Secondly, we propose an enriched conceptual framework based on systems theory. This framework encompasses five overarching dimensions: goal, structure, activities, environment, and outcome. It also considers the interaction between these dimensions. To illustrate this conceptual framework, we describe how it was used to structure the questions of a barometer dedicated to assessing data governance maturity in organizations. We then present a few use cases and discuss the implications for researchers and practitioners.

MOTS-CLES : gouvernance de données, cadre conceptuel, valeur, risque, coût, étude bibliométrique.

KEYWORDS: data governance, conceptual framework, value, risk, cost, bibliometric study.

1. Introduction

Les données sont un atout stratégique et une ressource pour les organisations, car elles conditionnent la prise de décision. La gouvernance des données permet la gestion des données de manière efficace pour améliorer la performance de l'organisation tout en respectant les obligations réglementaires et en minimisant les coûts et les risques. La gouvernance des données désigne l'ensemble des processus, politiques et outils qui permettent de déterminer qui est responsable de la prise de décision d'une organisation au sujet de ses données (Khatri et Brown, 2010). En cela, ils distinguent la gouvernance des données de leur management, lequel consiste à prendre et mettre en œuvre ces décisions. (Smallwood, 2019) la définit comme un ensemble de structures, politiques, procédures, processus et technologies utilisés pour collecter, organiser, utiliser et sécuriser les données. Plusieurs revues de la littérature sur la gouvernance des données ont été publiées (Nguyen, 2016 ; Abraham *et al.*, 2019 ; Merkus *et al.*, 2019). La gouvernance des données n'est pas simplement une question de technologie ou de management des données ; il s'agit fondamentalement d'établir les règles et les responsabilités relatives au management des données. Il manque, à ce jour, une définition acceptée par tous, tant chercheurs que praticiens. De plus, de nombreux cadres de référence (« frameworks ») ou modèles ont été proposés sans qu'aucun ne fasse consensus. Notre recherche est ainsi guidée par les questions suivantes : QR1 : *Quelle est la structure intellectuelle, les thèmes de recherche et les articles influents dans le domaine de la gouvernance des données ?* QR2 : *Quel cadre théorique peut-on proposer pour structurer les composants de ce domaine ?*

Pour répondre à ces questions, nous effectuons une analyse rétrospective complète de la littérature relative à la gouvernance des données dans la base de données Scopus (QR1). Pour ce faire, trois méthodes d'analyse bibliométrique quantitative sont employées : l'analyse des co-citations (CCA), l'analyse du couplage bibliographique (BCA) et l'analyse des chemins principaux (MPA). Grâce à cette étude bibliométrique, nous identifions les dimensions les plus saillantes et les lacunes, ce qui nous permet de développer un nouveau cadre conceptuel (QR2). Notre cadre n'est pas destiné à remplacer les cadres existants, mais plutôt à fournir une carte conceptuelle qui montre comment les différents éléments s'assemblent pour définir la gouvernance des données d'une manière multidimensionnelle.

Dans la section qui suit, nous présentons un état de l'art structuré autour d'une étude bibliométrique. Le cadre conceptuel fondé sur la théorie des systèmes est décrit dans la section 3. Dans la section 4, nous illustrons l'utilisation de ce cadre pour générer la structure d'un baromètre dédié à l'étude de la maturité des organisations en gouvernance des données et décrivons quelques cas d'usages. La section 5 conclut et suggère des voies de recherche future.

2. Étude bibliométrique de la gouvernance des données

L'objet principal de cette section est de dresser un état de l'art de la littérature scientifique sur la gouvernance des données. Au-delà, nous visons à cartographier

cette littérature à l'aide de techniques bibliométriques. A cette fin, nous utilisons trois techniques d'analyse des citations, après une étude descriptive du domaine. Dans la suite, nous présentons le dispositif méthodologique fondé sur trois étapes : la constitution du jeu de données, l'analyse descriptive et l'analyse par les citations.

2.1. Constitution du jeu de données

La constitution du jeu de données requiert le choix d'une base de données bibliographique, la construction d'une requête, l'extraction des résultats et le nettoyage des données. Dans cette recherche, nous avons soumis à la base de données bibliographiques Scopus, réputée pour sa pertinence dans le domaine de l'informatique et des systèmes d'information, la requête ci-dessous :

```
TITLE-ABS-KEY("data governance") AND ( EXCLUDE ( DOCTYPE,"ch" ) OR
EXCLUDE ( DOCTYPE,"bk" ) OR EXCLUDE ( DOCTYPE,"no" ) OR EXCLUDE (
DOCTYPE,"ed" ) OR EXCLUDE ( DOCTYPE,"sh" ) OR EXCLUDE ( DOCTYPE,"er" )
OR EXCLUDE ( DOCTYPE,"le" ) OR EXCLUDE ( DOCTYPE,"tb" ) ) AND ( LIMIT-
TO ( LANGUAGE,"English" ) )
```

La chaîne de recherche choisie est « data governance » limitée aux titres, résumés et mots-clés des publications en anglais dans des revues ou des conférences, ce qui nous retourne 2701 documents. Certains auteurs utilisent de manière interchangeable l'expression « gouvernance de l'information ». L'ajout de cette chaîne dans la requête n'a pas apporté de modifications significatives pertinentes. Après exportation des données, à l'aide du logiciel CRExplorer (Thor *et al.*, 2021), nous avons procédé à la déduplication des références. En effet, Scopus ne produit pas de documents en double mais, dans ces documents, de nombreuses références sont mal codées, rendant difficile leur rapprochement par les logiciels d'analyse bibliométrique. CRExplorer dispose d'un outil de rapprochement et d'harmonisation de ces références. Un nettoyage « manuel » a aussi été nécessaire pour éliminer quelques références incomplètes qui perturbaient les analyses, en l'occurrence une référence, ce qui nous a conduit à un jeu de données de 2700 articles.

2.2. Analyse descriptive

Avant d'utiliser les techniques d'analyse des citations, nous avons procédé à une analyse descriptive du jeu de données résultant de l'étape précédente. La distribution des articles obtenus est représentée à la figure 1. Le terme « gouvernance des données » apparaît pour la première fois en 2005 et le phénomène reste embryonnaire jusqu'en 2011 environ.

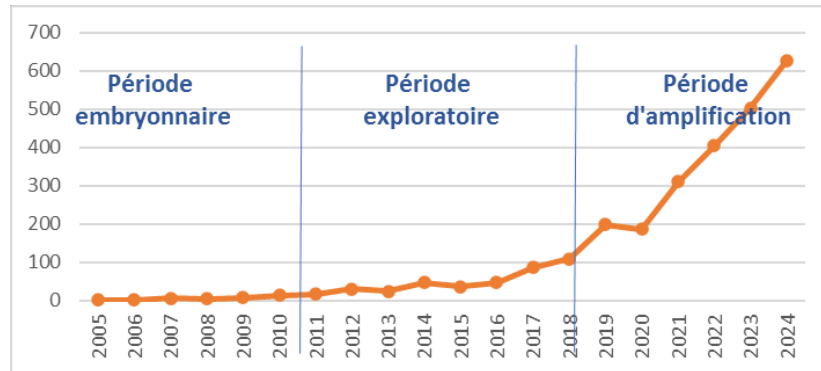


Figure 1. Nombre de publications par an

La période 2011 à 2018, qu'on peut qualifier d'exploratoire, révèle une augmentation régulière passant de 16 publications en 2011 à 109 en 2018. La période de 2018 à 2024, ou période d'amplification, voit le nombre de publications augmenter beaucoup plus rapidement jusqu'à atteindre 626 en la seule année 2024. Le terme de gouvernance de données est ainsi adopté progressivement. Le domaine de la gouvernance des données fait ainsi l'objet d'une recherche plus abondante. Scopus fournit un ensemble de tableaux et de graphiques décrivant certains traits saillants du jeu de données renvoyé par la requête. Nous avons sélectionné les plus pertinents. Ainsi, en ce qui concerne les auteurs les plus prolifiques (Fig. 2), Rob Brennan se détache avec plus de 20 publications centrées autour de la valeur de la donnée, sa qualité et la mesure de sa valeur (Brennan *et al.*, 2018). Les auteurs ne sont sélectionnés ici que sur leur nombre de publications dans le jeu de données, c'est-à-dire dans les articles « étiquetés » gouvernance de données.

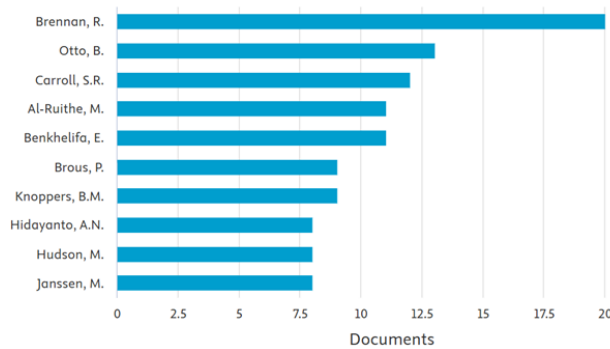


Figure 2. Auteurs les plus prolifiques

La distribution des documents par type montre deux pôles équilibrés entre les articles publiés dans des revues et ceux parus dans les actes de conférence (Fig. 3).

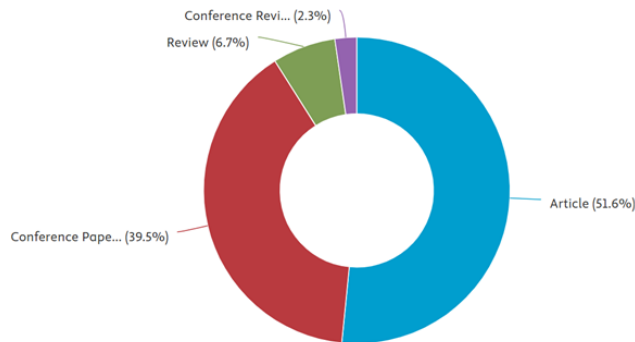


Figure 3. Documents par type

La recherche relative à la gouvernance des données est multidisciplinaire, avec toutefois près de 30% pour l’informatique (« computer science ») (Fig. 4). A noter qu’un domaine d’application, la médecine, émerge nettement (près de 9% des publications y sont rattachées).

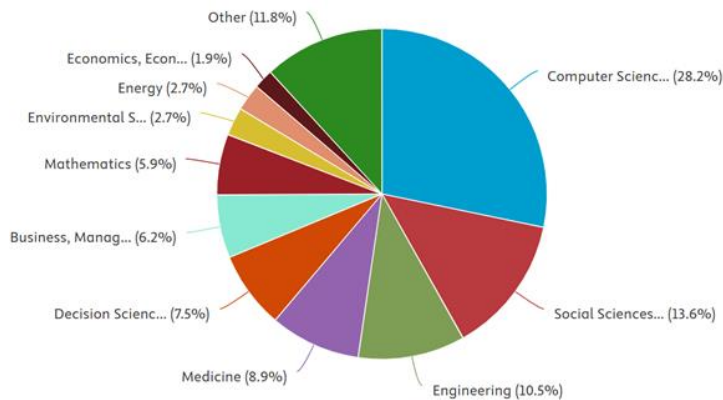


Figure 4. Documents par discipline

En résumé, cette analyse permet d’identifier une recherche en croissance en gouvernance des données, avec quelques auteurs très présents et des sujets comme la qualité de la donnée, sa confidentialité, sa valeur et donc le besoin de gouvernance.

2.2. Analyse par les citations

Les publications extraites sur la gouvernance des données sont trop nombreuses pour en faire une analyse individuelle. Dès lors, le recours aux techniques bibliométriques est judicieux. Dans cette section, nous présentons l’application des techniques de co-citation (« co-citation analysis ») pour identifier la structure

intellectuelle du domaine. Puis le couplage bibliographique (« bibliographic coupling analysis ») nous permet de repérer les thèmes principaux. Enfin, au moyen du calcul de chemins principaux (« main path analysis »), nous mettons en évidence les articles les plus influents et les liens de citation qui les relie.

2.2.1. Analyse des co-citations (CCA)

L'analyse des co-citations consiste à construire des classes (clusters) d'articles (ou références) cités par un ensemble d'articles citants. Plus deux références sont citées ensemble, plus elles sont proches et susceptibles d'appartenir au même cluster. Le logiciel utilisé ici est Vosviewer (pour Visualization of Similarities) met en œuvre une similarité appelée force d'association (association strength) pour répartir les articles dans le plan (Van Eck et Waltman, 2013). À l'aide de Vosviewer, nous avons catégorisé trois domaines de références regroupant les 28 articles cités plus de 30 fois dans notre ensemble de publications (Fig. 5). Le choix du seuil de 30 est empirique et dicté par le souci de ne retenir que des articles influents et d'obtenir une visualisation de taille raisonnable. La première catégorie (14 articles en rouge) regroupe les publications proposant des modèles et des cadres conceptuels (« frameworks ») sur le thème de l'organisation, les rôles, les responsabilités et la prise de décision relatifs à la gouvernance des données. Ainsi, Khatri et Brown (2010) différencient la gouvernance des données de son management et proposent un cadre distinguant les cinq domaines de décision relative à la gouvernance des données : les principes qui gouvernent l'usage des données, la qualité des données, les méta-données, l'accès aux données et le cycle de vie de la donnée. L'article de Weber *et al.* (2009) introduit un modèle matriciel de gouvernance de la donnée qui croise les rôles et les activités. Un autre article séminale décrit une étude empirique de la gouvernance de l'information concernant trente organisations (Tallon *et al.*, 2013). Ils en déduisent un modèle composé d'antécédents, facilitateurs ou inhibiteurs, des mécanismes structurels, procéduraux ou relationnels de la gouvernance de l'information avec des résultats en termes de performance et de mitigation des risques. À noter qu'il s'agit de la gouvernance de l'information et non de la gouvernance des données, bien que ces deux termes soient souvent utilisés de manière interchangeable.

La seconde catégorie (9 articles en vert) regroupe les publications proposant des cadres conceptuels de gouvernance de données intégrant davantage de dimensions. Ces articles paraissent dans des revues relevant de domaines très différents (informatique, juridique, politique, psychologique, managérial). Le plus cité est (Abraham *et al.*, 2019) qui étudie 145 sources afin d'identifier les composants de la gouvernance des données selon six dimensions : le domaine concerné (la qualité des données, la sécurité, les méta-données, raffinant les cinq domaines de Khatri et Brown), le périmètre organisationnel concerné (un département, toute l'organisation, l'inter-organisationnel), le type de données concerné (traditionnelles ou massives), les mécanismes de gouvernance (structurels, procéduraux ou relationnels), les antécédents tant internes qu'externes et les conséquences (performances et risques). Pour les données scientifiques, Wilkinson *et al.* (2016) proposent les quatre principes FAIR (pour facile à trouver, accessible, interopérable et réutilisable) pour

un bon management des données et une bonne intendance (stewardship). Janssen *et al.* (2020) en proposent treize à appliquer aux systèmes algorithmiques des big data, dont beaucoup reprennent les règles relatives aux données personnelles (par exemple la minimisation de l'accès aux données ou encore la nécessité d'informer).

La troisième catégorie (5 articles en bleu) est dominée par des analyses de la littérature naissante sur la gouvernance de données. Ainsi, Alhassan *et al.* (2016) proposent une analyse de fréquence des activités de gouvernance de données selon l'action (définir, mettre en œuvre et surveiller) effectuée dans un espace (standards liés aux données, exigences liées aux données, stratégie en matière de données, politique des données, etc.) pour un des cinq domaines définis par Khatri et Brown (2010). Al-Ruithe *et al.* (2019) présentent une analyse systématique de la littérature qui compare la gouvernance des données en général à celle liée au cloud. Six dimensions sont ainsi comparées : la politique de gouvernance de la donnée, son administration, sa structure organisationnelle, la dimension technologique, la dimension réglementaire et ses outils de mesure et de surveillance.

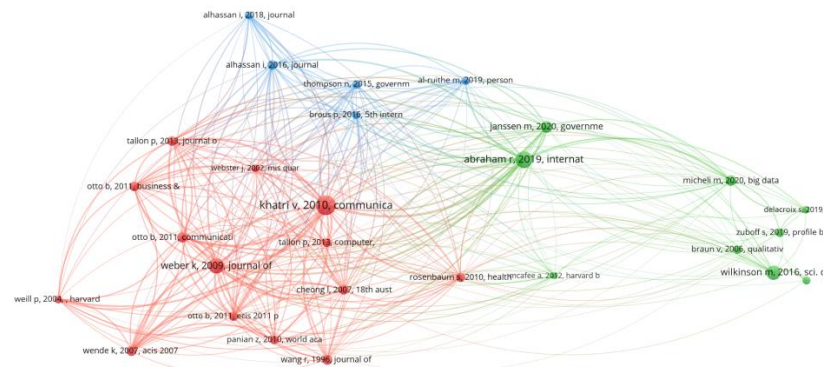


Figure 5. Analyse des co-citations

En conclusion, cette analyse nous a permis d'identifier la structure intellectuelle de la gouvernance des données, une première composante principalement orientée vers la dimension organisationnelle du sujet, une seconde élargissant la dimension à des cadres conceptuels plus complets et l'appliquant à des domaines variées, la troisième composante s'attachant à l'étude de la littérature naissante sur le sujet.

2.2.2. Analyse des couplages bibliographiques (BCA)

L'analyse des couplages bibliographiques consiste à construire des classes d'articles considérés proches parce qu'ils citent un même ensemble de références. Plus cet ensemble commun est grand, plus ils sont susceptibles d'appartenir au même cluster. Elle permet de rapprocher les articles travaillant sur des thématiques proches. D'autre part, elle complète l'analyse des co-citations puisqu'elle permet

d'intégrer aussi des articles récents, lesquels ne sont pas encore assez cités pour apparaître dans l'analyse des co-citations.

L'analyse des couplages bibliographiques produite par Vosviewer catégorise en 6 classes les 102 articles cités plus de 50 fois et connectés (Fig. 6). Le seuil de 50 est défini de façon empirique pour retenir les articles les plus influents. Vosviewer permet aussi de choisir le nombre de classes souhaité en modifiant la valeur de la résolution. Nous avons effectué ce calibrage en examinant, par essai-erreur, le contenu des classes et surtout leur interprétabilité.

La première classe (35 nœuds en rouge) regroupe des articles étudiant les problématiques de gouvernance de données dans le domaine de l'apprentissage automatique, de la santé et de la « blockchain ». La deuxième classe (28 nœuds en vert) contient de nombreux articles déjà mentionnés dans l'analyse des co-citations. Elle regroupe donc les articles fondateurs sur la gouvernance de données, notamment les différents cadres conceptuels. La troisième classe (16 nœuds en bleu) traite de la gouvernance de données appliquée à différents domaines dont les « smart cities », l'aspect juridique, les plates-formes de données, le « big data ». La quatrième classe (14 nœuds en jaune) est aussi caractérisée par le big data mais aussi d'autres domaines d'application, comme l'agriculture. La cinquième classe (7 nœuds en violet) cible le domaine de l'industrie 4.0 et celui des « smart cities ». Enfin, la sixième classe (2 nœuds en bleu clair) porte sur l'interdépendance entre bases de données et intelligence artificielle. A noter que le thème des « smart cities » commun aux classes 3 et 5 est traité sur un plan organisationnel dans la première et plus technologique dans la seconde. En conclusion, on a une seule classe (la deuxième) qui concentre les articles sur les fondamentaux de la gouvernance de données tandis que les autres les appliquent à différents domaines.

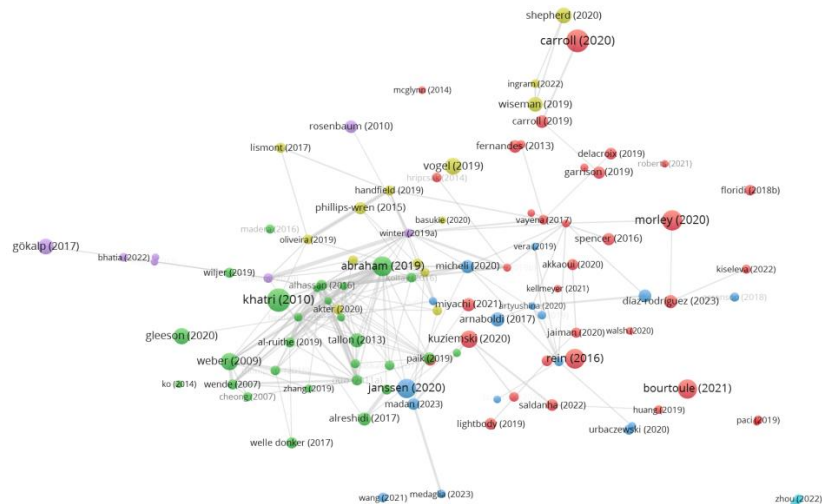


Figure 6. Analyse des couplages bibliographiques

Ainsi, deux thèmes principaux émergent. Le premier est lié aux fondements de la gouvernance de données. Le second, regroupant le reste des classes, est dédié aux applications dans différents domaines.

2.2.3. Analyse des chemins principaux (MPA)

L'analyse des chemins principaux s'effectue sur le graphe de citations entre les articles du jeu de données. Elle consiste à rechercher les trajectoires principales de ce graphe. La métrique utilisée ici est SPLC (Search Path Link Count) qui compte le nombre de fois où un arc est traversé si l'on parcourt tous les chemins possibles depuis tous les ancêtres du nœud d'origine vers tous les puits du graphe. A l'aide de cette métrique, on peut ensuite calculer les chemins les plus influents au moyen de différents algorithmes. Le logiciel Pajek propose plusieurs algorithmes. Celui utilisé ici est le global key-route (ici 10 key-routes) qui consiste à retenir les dix arcs ayant la valeur de SPLC la plus importante et, pour chacun de ces arcs, à calculer les chemins passant par cet arc. On obtient le graphe de la figure 7 qui représente une vision synthétique de la structuration du domaine de la gouvernance des données et les différents canaux de diffusion de la connaissance. Parmi ces chemins, le plus impactant est en rouge sur la figure.

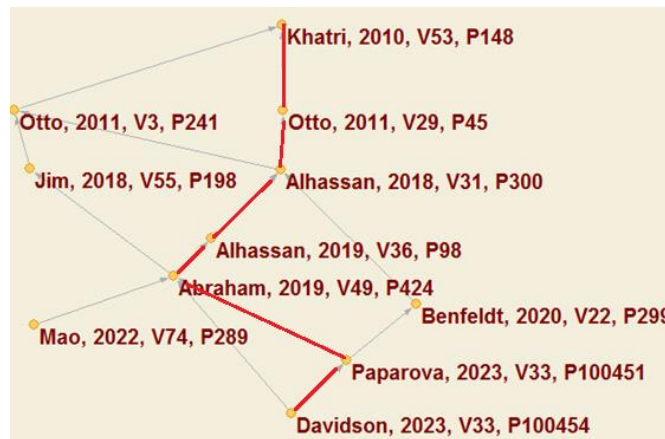


Figure 7. Analyse des chemins principaux de la gouvernance de données

Le nœud racine est l'article de Khatri et Brown (2010) déjà mentionné dans la CCA. Otto (2011) compare l'organisation de la gouvernance de données dans deux entreprises de télécommunications selon trois dimensions (le but, la structure et les processus). Alhassan *et al.* (2018) comparent la littérature scientifique et la presse professionnelle selon un cadre qui croise les cinq domaines de décision définis par Khatri et Brown (2010) et les trois actions Définir, Mettre en œuvre et Surveiller. Les mêmes auteurs identifient aussi les facteurs clés de succès de la gouvernance des données au travers de la littérature (Alhassan *et al.*, 2019). Abraham *et al.*

(2019) proposent un cadre de la gouvernance des données qui combine 3 dimensions (types de données, niveau organisationnel, processus de mise en œuvre) pour analyser les mécanismes de gouvernance, leurs antécédents et les résultats de la gouvernance des données. Paparova *et al.* (2023) introduisent le concept d'espace de gouvernance de données qui combine une logique verticale et une logique horizontale en matière de rôles et responsabilités. Cet article fait partie d'un numéro spécial sur la gouvernance des données dont l'éditorial porte sur les défis de la gouvernance des données à l'ère de l'innovation digitale (Davidson *et al.*, 2023). Ce chemin principal constitue une sorte de colonne vertébrale (« backbone ») de diffusion de la connaissance du domaine naissant de la gouvernance de données

Pour des raisons d'espace, nous n'avons pas pu effectuer une analyse par période. Les différentes analyses menées font ressortir le concept de but de la gouvernance de donnée et celui de périmètre. Certains articles mettent l'accent sur les activités liées à la gouvernance des données, par exemple les efforts de mise en conformité. D'autres pointent sa structure qui comprend les rôles et les responsabilités. D'autres encore identifient des éléments facilitateurs ou des facteurs clés de succès. Enfin, certains auteurs structurent tout ou partie de ces éléments en un cadre conceptuel. A notre connaissance, aucun de ces cadres n'utilise un fondement théorique ni ne considère une vue holistique de la gouvernance de données. C'est un tel cadre que nous proposons dans la section suivante.

3. Vers un cadre de référence pour la gouvernance des données

Notre cadre de référence est fondé sur la théorie des systèmes (Skyttner, 1996). Comme nous l'avons établi dans (Akoka et Comyn-Wattiau, 2019), la gouvernance des données est un artefact de conception que l'organisation développe et met en œuvre pour atteindre ses objectifs. Simon (1996) caractérise les artefacts en termes de fonctions (*activités*), d'*objectifs* et d'adaptation (*évolution*). Il distingue également la *structure* de l'artefact de *l'environnement* dans lequel il opère. De plus, la gouvernance des données est au cœur d'un processus de pilotage dont elle ne peut être définie indépendamment. Son objet est la prise de décision, à la fois pour fixer les orientations de l'action (les *objectifs*), et pour définir et ajuster le cadre de fonctionnement (*structure* et *environnement*) et les *activités* correspondantes. En fixant les objectifs, les modalités et les règles du jeu des activités de management de données, la gouvernance des données est également amenée à les faire *évoluer* si nécessaire, en fonction des *résultats* obtenus : c'est là son rôle de pilote principal de la donnée. La gouvernance des données trouve sa signification dans la mise en relation des éléments qui la composent (Le Moigne, 1994 ; Skyttner, 1996). Ces dimensions forment un ensemble indissociable donnant lieu à une évaluation des *résultats* bénéficiant d'une *rétroaction*. À ce titre, la gouvernance des données possède toutes les caractéristiques d'un système. Nous présentons ci-dessous les composants de ce système et leur interaction (Fig. 8).

La dimension **But** de notre système comporte deux sous-dimensions. La première sous-dimension se rapporte à l'objectif de la gouvernance des données.

Bien sûr son but est dicté par l'alignement avec la stratégie mais il reste celui de maximiser sa valeur en minimisant les coûts et les risques (Bennett, 2017).

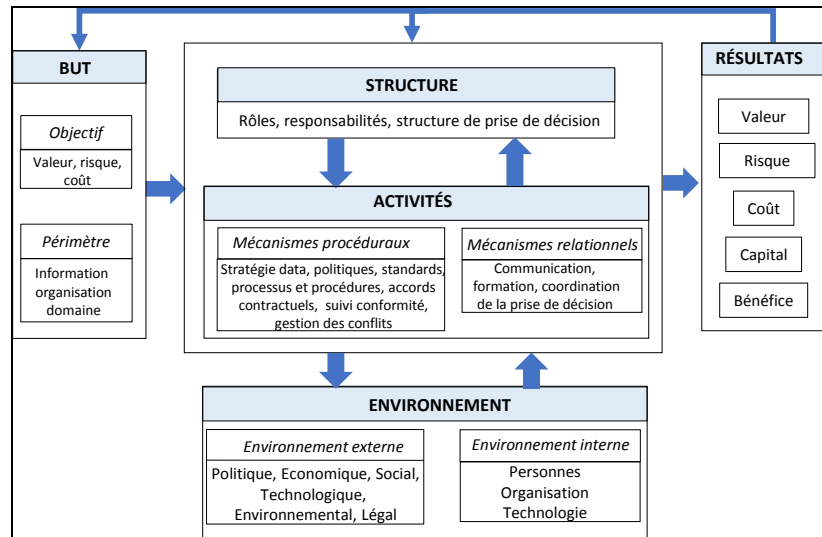


Figure 8. Le cadre de référence

La deuxième sous-dimension précise le périmètre auquel s'applique la gouvernance. Trois composants définissent ce périmètre, à savoir le périmètre « donnée », celui de « l'organisation » et celui du « domaine », tel que suggéré par Abraham *et al.* (2019). Dans notre cadre, le périmètre « donnée » intègre les données traditionnelles (structurées), les données massives (semi-structurées ou non structurées) mais aussi les données synthétiques générées par l'intelligence artificielle. Le périmètre « organisation » englobe les niveaux intra- et inter-organisationnels, mais peut aussi ne s'appliquer qu'à une entité de l'organisation. Enfin, le périmètre « domaine » est décrit par sept éléments, à savoir la qualité des données, la sécurité, l'architecture, le cycle de vie, les métadonnées, le stockage et l'infrastructure, ainsi que la gestion des documents et du contenu. La deuxième dimension a trait à la **Structure**, en liaison avec les mécanismes structurels proposés dans (Tallon *et al.*, 2013). Ces mécanismes englobent les rôles et les responsabilités ainsi que l'attribution du pouvoir de décision. La troisième dimension, appelée **Activités**, identifie les mécanismes procéduraux et les mécanismes relationnels de la gouvernance tels qu'ils sont proposés dans (Tallon *et al.*, 2013). Les premiers englobent la stratégie d'information, les politiques, les règles, les normes, les processus et les procédures, les accords contractuels, le contrôle de la conformité et la gestion des problèmes. Les mécanismes relationnels font référence à la communication, à la formation et à la coordination de la prise de décision. La quatrième dimension ou **Environnement** du système est composée de facteurs externes et internes à l'organisation et ayant un impact sur sa gouvernance des

données. Tallon *et al.* (2013) identifient un facteur externe, celui de la réglementation. Nous proposons d'enrichir cet ensemble de facteurs externes à l'aide du modèle PESTLE (Rastogi *et al.*, 2016), incluant tant l'impact de la politique, que l'économie et le social, mais aussi les éléments technologiques, légaux et environnementaux. Comme le montrent Tallon *et al.* (2013), l'environnement est également composé de différents facteurs internes à l'organisation, ayant un impact potentiel sur le système de gouvernance de la donnée. Plusieurs auteurs mentionnent une liste d'éléments (culture, stratégie informatique, soutien de la direction générale, etc.) Nous proposons de structurer ces facteurs en trois catégories : les personnes, l'organisation et la technologie (Hevner *et al.*, 2004). La cinquième dimension **Evolution** décrit les changements du système de gouvernance des données au fil du temps. Elle n'apparaît pas dans la figure 8 qui représente la gouvernance à un moment donné. Ces changements peuvent affecter toutes les autres dimensions. Elle n'est pas décrite dans les cadres de référence de la littérature, autres que les modèles de maturité. Elle constitue une dimension fondamentale permettant aux entreprises de construire et d'alimenter ces modèles de maturité. Enfin, le système comprend une dimension de performance (valeur, risque, coût, bénéfice, capital investi) qui permet d'exercer la boucle de rétroaction pour faire évoluer le système.

En conclusion, en mobilisant la théorie des systèmes comme fondement de ce cadre conceptuel, nous offrons une vue holistique de la gouvernance des données composée de cinq dimensions, permettant d'identifier les interactions existant entre elles. Notre contribution consolide les différents axes de la recherche en matière de gouvernance des données en un modèle unique enrichi avec les composants d'autres domaines (Hevner *et al.*, 2004 ; Rastogi *et al.*, 2016 ; Bennett, 2017).

Il existe des cadres de référence proposés par les praticiens¹ (DAMA-DMBoK, COBIT, DGI, SAS, BCG). Malgré plusieurs points communs (principes, rôles et responsabilités organisationnels, politiques, règles et normes, processus de base, etc.), chaque cadre a un centre d'intérêt et son propre champ d'application : certains se concentrent sur la gouvernance stratégique, d'autres sur le management en termes de pratiques détaillées de gestion des données, d'autres encore sur des objectifs spécifiques (qualité, conformité, valeur). Notre cadre conceptuel offre une vue globale qui peut être déclinée par niveau et par domaine d'application, par exemple données massives et intelligence artificielle, secteur public et villes intelligentes.

4. Applications du cadre de référence

Ce cadre a été utilisé pour générer la structure d'un baromètre d'évaluation de la maturité des entreprises et organisations françaises en matière de gouvernance des données, appelé MetraData². Cette enquête a été menée auprès de 150 entités. Nous

¹ <https://www.kellton.com/kellton-tech-blog/popular-data-governance-frameworks#:~:text=activities%20like%20metadata%20and%20architecture>

² <https://drive.google.com/file/d/1Gx3-mE0eRdizmlw4npObeXKBLm1daX3t/view>

avons généré un questionnaire en ligne comprenant environ 40 questions auprès de dirigeants d'entreprise, directeurs des systèmes d'information, directeurs de données et directeurs fonctionnels dans tous les secteurs (privé, public, parapublic, association), tous les domaines d'activité (assurance, technologie, banque, etc.) et toutes les tailles d'entreprises. On a recueilli la perception du concept de gouvernance des données, son opérationnalisation dans l'organisation, ses acteurs-clés, le contexte et les plans d'évolution. Le baromètre met en lumière l'importance des facteurs organisationnels et humains permettant d'améliorer la gouvernance des données. La première colonne du tableau 1 reprend chaque dimension du cadre de référence. La seconde colonne contient un exemple de question fondée sur cette dimension. La troisième colonne fournit les modalités de réponse proposées

Tableau 1. Correspondance entre le cadre de référence et le baromètre *MetraData*

Dimension/ Sous-dimension	Question	Réponses
But/objectif	Quels sont les objectifs principaux de la gouvernance des données au sein de votre organisation ?	Maximiser la valeur métier tirée des données Assurer la conformité aux lois et réglementations Minimiser les risques liés aux données Minimiser les coûts de gestion des données
But/périmètre/ donnée	Quels sont les principaux périmètres d'intervention de la gouvernance des données dans votre organisation ?	Bases de données classiques, documents, "big data", emails, sites web, supports numériques (réseaux sociaux), supports papier, données synthétiques de l'IA
Structure/ Rôles et responsabilités	Quels sont les principaux rôles et responsabilités de la gouvernance des données identifiés dans votre organisation ?	Gestionnaire de données, architecte de données, propriétaire de données, directeur de données, analyste des données, délégué à la protection des données
Structure/Prise de décision	Quelles instances de votre organisation ont en charge tout ou partie de la gouvernance des données ?	Comité stratégique, comité risques et sécurité, comité de direction, comité de pilotage dédié, aucun
Activités/Mécanismes procéduraux	Les mécanismes ci-dessous sont-ils mis en place dans votre organisation pour opérationnaliser les activités de gouvernance des données ?	Gestion de la conformité Preuve et signature électronique Sécurité des données Architecture d'entreprise et modélis. données et flux Architecture des données Gestion de la qualité Définition des rôles liés aux données Investigation informatique (e-discovery) Archivage électronique Gestion des méta-données et catalogage Gestion des données maîtres et de référence Analyse des risques informationnels Gestion des documents et contenus
Activités/Mécanismes relationnels	Comment qualifiez-vous le dialogue entre les responsables de la gouvernance des données et les métiers dans votre organisation ?	Excellent, bon, améliorable, tendu, inexistant
Environnement/Externe	Quel a été l'élément déclencheur qui a suscité une préoccupation pour la gouvernance des données ?	Fusion/acquisition Cyber-attaque Fuite de donnée Nouvelle réglementation (e.g. RGPD), etc.

Environnement/ Interne/ Technologie	Quels sont les outils spécifiques que vous avez mis en place dans votre organisation afin d'opérationnaliser les activités de gouvernance des données ?	Master Data Management Mise en qualité des données Gestion documentaire Modélisation et architecture de données Classification et catalogage de données Registre de traitement RGPD, etc.
Environnement/ Interne/ Personnes et Organisation	Quels sont les éléments qui freinent ou pourraient freiner l'implantation de la gouvernance des données dans votre organisation ?	Manque de moyens humains Résistance au changement Culture d'entreprise éloignée de ces préoccupations Manque de dialogue métiers et équipes gouv. donnée Manque d'engagement de la direction Manque de dialogue entre la DSI et les équipes gouvernance donnée La data n'est pas un enjeu clé
Résultats	Quels éléments de la gouvernance des données sont suivis dans le reporting régulier de l'organisation ?	Avancement des projets Budget Niveau de satisfaction des services data Gestion des compétences

D'autres applications sont à l'étude. Premièrement, le cadre fournit un modèle théorique pour les recherches futures sur la gouvernance des données. De nouvelles questions de recherche peuvent émerger en croisant les principales dimensions. À titre d'illustration, l'étude de l'évolution de la gouvernance des données dans une organisation ainsi qu'au niveau sectoriel est un domaine de recherche prometteur. Un cadre offre de nombreux avantages pour l'élaboration d'approches de recherche. La théorie des systèmes en constitue une base solide. Deuxièmement, le cadre peut être utilisé comme modèle pour structurer la charte de gouvernance des données d'une organisation. Certaines chartes sont disponibles sur des sites web, principalement dans le domaine de la santé. Toutefois, elles se concentrent sur les obligations légales et les questions de conformité. Fournir un tel cadre aux experts chargés de rédiger une charte facilite la recherche de l'exhaustivité. Nous avons vérifié sur certaines chartes existantes que notre cadre couvre tous les sujets qu'elles contiennent. Troisièmement, le cadre de référence fournit une base pour l'évaluation de la gouvernance des données. Dans cette optique, chaque dimension peut être associée à des tests d'audit et à des mesures. Nous avons commencé à définir les questions d'audit avec un panel de praticiens. À titre d'exemple, nous avons affiné les objectifs de valeur, de risque et de coût en élaborant des typologies pour chacun d'entre eux. La valeur englobe plusieurs composantes : commerciale, intrinsèque, de performance, économique, etc. Chacune d'entre elles peut être mesurée à l'aide d'un indicateur spécifique. Enfin, le cadre de référence peut servir de point de départ au développement d'un modèle de maturité de la gouvernance des données. De nombreux modèles de maturité sont proposés dans la littérature. Ils sont utilisés pour évaluer, par exemple, la manière dont les parties prenantes comprennent et adoptent les concepts et les méthodes. Pour développer un modèle de maturité, chaque dimension du cadre de référence doit être déclinée en niveaux de maturité et associée à un ensemble de mesures.

5. Conclusion et recherche future

Cette recherche a pour but de structurer le domaine de la gouvernance des données en développant un cadre conceptuel fondé sur la théorie des systèmes qui définit ses dimensions clés. Ce faisant, nous avons répondu au besoin d'une vision holistique qui puisse guider à la fois les chercheurs dans l'élaboration d'hypothèses et les praticiens dans l'organisation de leur gouvernance des données. Cette recherche a abouti à un cadre conceptuel pour la gouvernance des données mais également mis en évidence plusieurs axes dans lesquels des recherches supplémentaires sont nécessaires. Ainsi, une validation empirique du cadre conceptuel auprès de praticiens est en cours, ainsi que l'opérationnalisation d'un modèle de maturité de la gouvernance des données. L'applicabilité du cadre conceptuel à d'autres types de gouvernance (informatique, connaissance, entreprise, etc.) est une autre voie de recherche future.

Remerciements. Les auteurs remercient les relecteurs pour leurs précieux conseils et les partenaires de la Chaire ESSEC Stratégie et gouvernance des données et de l'IA au sein de laquelle cette recherche a été menée.

Bibliographie

- Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International journal of information management*, 49, 424-438.
- Akoka, J., & Comyn-Wattiau, I. (2019, June). Évaluation de la gouvernance de l'information: une approche holistique. In AIM 2019: 24ème Conférence de l'Association Information et Management.
- Alhassan, I., Sammon, D. and Daly, M. (2018), "Data governance activities: a comparison between scientific and practice-oriented literature", *Journal of Enterprise Information Management*, Vol. 31 No. 2, pp. 300-316. <https://doi.org/10.1108/JEIM-01-2017-0007>
- Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. *Journal of Decision Systems*, 25(sup1), 64-75.
- Alhassan, I., Sammon, D., & Daly, M. (2019). Critical success factors for data governance: a theory building approach. *Information systems management*, 36(2), 98-110.
- Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and ubiquitous computing*, 23, 839-859.
- Bennett, S. (2017). What is information governance and how does it differ from data governance? *Governance Directions*, 69(8), 462-467.
- Brennan, R., Quigley, S., De Leenheer, P., & Maldonado, A. (2018). Automatic extraction of data governance knowledge from slack chat channels. In OTM Confederated International Conferences" On the Move to Meaningful Internet Systems" (pp. 555-564). Cham: Springer International Publishing.
- Davidson, E., Wessel, L., Winter, J. S., & Winter, S. (2023). Future directions for scholarship on data governance, digital innovation, and grand challenges. *Information and Organization*, 33(1), 100454.

- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 75-105.
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government information quarterly*, 37(3), 101493.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148-152.
- Le Moigne, J. L. (1994). *La théorie du système général : théorie de la modélisation*. PUF.
- Merkus, J., Helms, R., & Kusters, R. J. (2019, May). Data Governance and Information Governance: Set of Definitions in Relation to Data and Information as Part of DIKW. In ICEIS (2) (pp. 143-154).
- Nguyen, T. C. (2016). *Information governance and management in the context of Gov 2.0* (Doctoral dissertation, Swinburne).
- Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, 29(1), 3.
- Paparova, D., Aanestad, M., Vassilakopoulou, P., & Bahun, M. K. (2023). Data governance spaces: the case of a national digital service for personal health data. *Information and Organization*, 33(1), 100451.
- Rastogi, N., & Trivedi, M. K. (2016). PESTLE technique—a tool to identify external risks in construction projects. *International Research Journal of Engineering and Technology (IRJET)*, 3(1), 384-388.
- Simon, H. A. (1996). *The architecture of complexity: hierarchic systems*.
- Skyttner, L. (1996). *General systems theory: An introduction*. London: Macmillan Press.
- Smallwood, R. F. (2019). *Information Governance: Concepts, Strategies and Best Practices*, Wiley.
- Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. *Journal of management information systems*, 30(3), 141-178.
- Thor, A., Bornmann, L., Haunschild, R., & Leydesdorff, L. (2021). Which are the influential publications in the Web of Science subject categories over a long period of time? CRExplorer software used for big-data analyses in bibliometrics. *Journal of Information Science*, 47(3), 419-428.
- Van Eck, N. J., & Waltman, L. (2013). *VOSviewer manual*. Leiden: Universiteit Leiden, 1(1), 1-53.
- Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality (JDIQ)*, 1(1), 1-27.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1-9.