



Titre: Méthodologie pour évaluer la qualité des données dans le cadre
Title: d'un projet de transformation numérique

Auteur: Khelil Joudane
Author:

Date: 2023

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Joudane, K. (2023). Méthodologie pour évaluer la qualité des données dans le
Citation: cadre d'un projet de transformation numérique [Mémoire de maîtrise,
Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/53352/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/53352/>
PolyPublie URL:

**Directeurs de
recherche:** Christophe Danjou, & Bruno Agard
Advisors:

Programme: Maitrise recherche en génie industriel
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Méthodologie pour évaluer la qualité des données dans le cadre d'un projet de
transformation numérique**

KHELIL JOUDANE

Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie industriel

Avril 2023

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Méthodologie pour évaluer la qualité des données dans le cadre d'un projet de
transformation numérique**

présenté par **Khelil JOUDANE**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Martin TRÉPANIÉ, président

Christophe DANJOU, membre et directeur de recherche

Bruno AGARD, membre et codirecteur de recherche

Camélia DADOUCHI, membre

DÉDICACE

*À tous ceux qui m'ont vu grandir,
Grâce à votre regard bienveillant je m'épanouis . . .*

REMERCIEMENTS

Un grand merci à mon directeur Mr. Christophe Danjou et mon codirecteur Mr. Bruno Agard pour leur patience et leur accompagnement. C'est grâce à leurs conseils et encouragements que mon expérience fut fructueuse.

Merci à La Milanaise, notre partenaire industriel. Le temps et les moyens mis à ma disposition ont été d'une grande aide.

À ma famille, merci pour votre confiance et la force que vous me donnez. Vous êtes mon moteur et à jamais mon inspiration.

RÉSUMÉ

Les technologies de l'industrie 4.0 s'imposent dans les entreprises depuis ces dernières décennies. Cela fait partie d'une volonté des entreprises de maximiser l'utilisation des données et de maîtriser les technologies du numérique. Les technologies qu'apporte l'industrie 4.0 ont un impact important sur la compétitivité des entreprises. Elles permettent un meilleur contrôle des processus et ouvrent les portes pour l'amélioration. Les grandes entreprises investissent massivement dans ce qui est considéré par plusieurs comme la quatrième révolution industrielle. Le terme de révolution industrielle reste à être confirmé par l'histoire, cependant il est indéniable que ces nouvelles technologies impactent la façon de travailler. Les entreprises qui adoptent ce nouveau paradigme de production passent par un processus de transition vers le numérique.

Cependant, les Petites et Moyennes Entreprises (PME) sont en retard dans l'adoption des technologies de gestion de données que l'industrie 4.0 apporte. Les coûts des projets restent élevés et les solutions proposées ne sont pas standardisées. Chaque entreprise aura des besoins spécifiques et les solutions doivent être modulées pour répondre à ces besoins. Il y a quand même une volonté des PME à développer des solutions adaptées à leur taille. Au Québec, le gouvernement appuie les projets visant à accroître la maîtrise du numérique dans les PME. Des projets de recherche sont menés pour trouver des solutions intéressantes pour les PME. Certains de ces projets entrepris sont des projets d'exploitations des données de production. Ces projets visent à récupérer et utiliser les données du plancher de production pour créer de la valeur ajoutée.

Avec l'augmentation du nombre de données industrielles récupérées, la question de la pertinence de ces données se pose. Un projet qui se base sur l'utilisation de données a besoin de s'assurer que les données utilisées sont de qualité. Les projets de transformation numérique sont documentés et plusieurs guides existent pour réussir cette transformation. Vu que les données sont centrales à l'industrie 4.0, un grand intérêt a été porté pour définir au mieux les caractéristiques qui font d'une donnée industrielle une donnée de bonne qualité. Il reste à intégrer cette compréhension de la qualité d'une donnée dans les guides de transformation numérique. Il faut aussi définir des outils d'évaluation de la qualité des données. L'objectif de ce mémoire est d'établir une méthodologie permettant d'évaluer la qualité des données récupérées sur le plancher de production dans le cadre d'un projet de transition numérique.

Le mémoire s'intéresse à la place de la qualité des données lors d'un projet de transformation numérique. Plusieurs questions apparaissent dans la problématique. La première est comment

réussir à identifier les besoins en qualité des données lors du lancement d'un projet. En effet, chaque projet a des besoins spécifiques en données et ces données doivent avoir un certain standard. Ensuite, il est important de définir des outils et des méthodes pour évaluer la qualité des données. Après avoir défini le standard requis en termes de qualité, il faut réussir à évaluer si les données atteignent ce standard. Ainsi la qualité des données peut être suivie en vue d'être améliorée.

Pour résoudre les questions évoquées plus haut, une méthodologie est développée au cours de ce mémoire. La méthodologie repose sur des outils pour définir les besoins en qualité des données d'un projet d'analyse de données. La qualité des données sera évaluée selon différentes caractéristiques. Cette évaluation se fait par le biais d'indicateurs de performance (KPIs). Une méthode pour créer des KPIs adaptés est présentée. Le résultat final de cette méthodologie est un document descriptif des données qui rajoutent du contexte et des informations pour les utilisateurs de la donnée.

Pour illustrer la méthodologie développée, elle est appliquée à des cas réels. Ces cas sont des projets réalisés chez un partenaire industriel. Ce partenaire produit de la farine biologique. Ce sont les difficultés rencontrées lors du travail avec le partenaire qui ont guidé la réflexion. La méthodologie développée est applicable dans ce contexte. Elle nécessite cependant d'être appliquée dans d'autres industries pour juger de sa solidité.

ABSTRACT

In the previous years enterprises started investing in numeric technologies. Those investments reflect the will to adopt industry 4.0 technologies. Those technologies are considered a part of the fourth industrial revolution. The term revolution is yet to be confirmed by history, but there is no denying the impact on the way industry functions nowadays. Enterprises having the funds to invest took the leap to stay competitive and to gain control over their production. The industry 4.0 technologies are spreading around the world at a fast pace.

Moreover, Small and Medium Entreprises (SMEs) have a hard time adopting industry 4.0 technologies. The cost of the installation is high. Furthermore, the solutions that are available still need customization to fit within an enterprise. Each enterprise has unique needs that need to be met. Customization of software and hardware technologies builds up the cost. Adding to that the lack of qualified personnel who can install and manage those technologies, it is harder for SMEs to adopt those technologies. Research in the area of making industry 4.0 technologies accessible to SMEs is encouraged. Projects leading to data collection are being launched.

The exponential growth in the number of data retrieved in the industry brings the question of the quality of data. Projects about numerical transformation are well documented, and road maps to achieve those projects are found in the literature. In those projects data is a central component. Characteristics have been defined to describe the quality of data. But to evaluate the quality of a data there is a need for evaluation tools. This thesis establishes a methodology to evaluate the quality of data retrieved from the production floor in the framework of a numerical transformation project.

This work develops a method to ensure that the quality of the retrieved data matches the needs of a data project. Three main steps were developed. First, defining the needs in data quality of a project and taking a decision on the system to use. There are a lot of choices in terms of sensors and ways of installing them. This step guides decision takers throughout the process. Second is creating a set of Indicateurs de Performances Clé (Key Performance Indicators) (KPI) to trace the data quality. Third, use the KPI to evaluate the data. In this step, a canvas on how to store information related to the collected data will be presented. The meta-data will add context to someone who will use the collected data. The KPI needs to be stored with the meta-data. This method has been applied in the field with an industrial partner.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vii
TABLE DES MATIÈRES	viii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xii
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Introduction	4
2.2 La place des données dans les technologies de la transformation numérique	5
2.2.1 Les leviers technologiques de l'industrie 4.0	5
2.2.2 La places des données dans l'industrie 4.0	8
2.2.3 Les enjeux stratégiques pour les données	10
2.3 La notion de qualité des données	13
2.3.1 Concept de la qualité de l'information	13
2.3.2 Qualité de l'information	14
2.3.3 Indicateurs de performance clés : KPIs	15
2.4 La transition numérique dans les PME	17
2.4.1 Les PME et le numérique	17
2.4.2 Définition d'un plan d'action pour la transition numérique	20
2.4.2.1 Méthode 6P	20
2.4.2.2 Méthodes *10 steps*	21
2.4.3 Structure pour la qualité de l'information	22
2.5 Conclusion	23

CHAPITRE 3	PROBLÉMATIQUE	26
3.1	Retour sur la revue de littérature	26
3.2	Problématique	27
3.3	Solution proposée	27
CHAPITRE 4	MÉTHODOLOGIE D'ÉVALUATION DE LA QUALITÉ DES DON- NÉES	30
4.1	Étapes préliminaires - Conditions de départ	30
4.2	Étape 1- Définition des attentes qualité	32
4.3	Étape 2- Développement de KPIs	35
4.3.1	Méthode de développement de KPIs	37
4.3.1.1	Identifier le problème	38
4.3.1.2	Trouver une méthode de mesure	39
4.3.1.3	Choisir le format	39
4.3.2	Autre exemple de KPI développé	40
4.3.2.1	KPI-Fréquence	40
4.3.3	KPI-Manutention	42
4.3.4	KPI-Évolution	43
4.4	Étape 3- Évaluation de la qualité de la donnée	43
CHAPITRE 5	ÉTUDES DE CAS	47
5.1	Contexte industriel	47
5.1.1	Présentation du partenaire	47
5.1.2	Projet d'amélioration de la farine	48
5.2	Projet 1 : Mesure de l'écartement des rouleaux de broyage	49
5.2.1	Étape 1-Définition des attentes qualité	53
5.2.2	Étape 2- développement des KPIs	55
5.2.3	Étape 3- Évaluation de la qualité de la donnée	56
5.3	Projet 2 : Mesure du niveau dans les silos	57
5.4	Projet 3 : Suivi des palettes de farine	59
5.5	Résultats de l'étude de cas	60
CHAPITRE 6	CONCLUSION ET RECOMMANDATIONS	61
6.1	Synthèse des travaux	61
6.2	Limitations de la solution proposée et solutions	62

LISTE DES TABLEAUX

Tableau 2.1	Sources de données, d'après [Yuan et al., 2021]	11
Tableau 2.2	Besoins QdD-Traduit de [McGilvray, 2021]	22
Tableau 2.3	Méthode POSMAD-traduit de [McGilvray, 2021]	22
Tableau 4.1	Tableau de prise de décision	33
Tableau 4.2	Décision et critères de qualité	36
Tableau 4.3	KPI-Fréquence	41
Tableau 4.4	KPI-Manutention	42
Tableau 4.5	KPI-Évolution de l'installation	44
Tableau 4.6	Gabarit des méta-données	45
Tableau 5.1	Tableau de prise de décision pour le capteur serrage rouleaux	54
Tableau 5.2	Méta-données remises à jour avec KPI.	57

LISTE DES FIGURES

Figure 2.1	Capacités des systèmes - [Porter and Heppelmann, 2019] . . .	19
Figure 2.2	Méthode 6P-Traduite de [Spaltini et al., 2022]	21
Figure 3.1	Méthodologie d'évaluation de la qualité des données	28
Figure 4.1	KPI- Exactitude et précision	41
Figure 5.1	Processus simplifié	49
Figure 5.2	Anatomie du grain de blé	50
Figure 5.3	Poignée d'ajustement des rouleaux de broyage	51
Figure 5.4	Hélice du capteur infrarouge. Par Abdellah Aboukhar	52

LISTE DES SIGLES ET ABRÉVIATIONS

AHP	Procédure Hiérarchique d'Analyse (Analytic Hierarchy Process)
BI	Informatique Décisionnelle (Business Intelligence)
CPS	Système Cyber-Physique
ERP	Progiciel de Gestion Intégré (Entreprise Ressources Planning)
IoS	Système d'Exploitation pour la Connexion des Réseaux (Internetwork Operating System)
IoT	Internet des Objets (Internet of Things)
KPI	Indicateurs de Performances Clé (Key Performance Indicators)
MES	Logiciel de Pilotage de la Production (Manufacturing Execution system)
MM	Modèle de Maturité
OGM	Organismes Génétiquement Modifiés
PMBok	Corps des Savoir en Gestion de Projet (Project Management Body of Knowledge)
PIB	Produit Intérieur Brut
PME	Petites et Moyennes Entreprises
QdD	Qualité des Données
QI	Qualité de l'Information (Information Quality)

CHAPITRE 1 INTRODUCTION

Pour maintenir le Québec à une place compétitive dans l'industrie mondiale, le gouvernement aide les entreprises à réussir leur transformation numérique [Québec, 2016]. En 2016, le "Plan d'action en économie numérique" est un plan d'action qui a été mis en place sur 5 ans [Québec, 2016]. L'État indique, qu'à travers ce plan, des investissements à hauteur de 200 millions de dollars seront déboursés [Québec, 2016]. Les investissements mis en place, dans le cadre des projets de ce plan d'action, visaient à mettre à jour le Québec en termes de transition numérique. Ce sont les PME qui ont été les plus encouragées par l'État, car selon Statistique Canada les PME, les PME participent à plus de 50% du PIB national [Canada, 2020]. Cependant, les industries canadiennes ont un retard dans l'adoption des technologies proposées par l'industrie 4.0 [Danjou et al., 2017].

Dans cet esprit d'aide et de recherche, un sondage sur l'industrie 4.0 a été mené au Québec. Dans ce sondage, 401 gestionnaires des opérations ou de la production de PME ont été interrogés sur leurs connaissances de l'industrie 4.0. Les faits marquants montrent que 39% des PME connaissent le concept d'industrie 4.0 [Fortin, 2019]. Ceci montre que, malgré les efforts du gouvernement, le passage au numérique tarde à se faire connaître. Pour enchaîner sur quelques autres faits, la majorité des entreprises (52%) pensent que l'impact de la transformation numérique sera faible au bout de 3 ans. Cela indique qu'il n'y a pas un fort taux d'engagement de la part des entreprises. Plus de la moitié des entreprises interrogées ne croient pas en l'impact de la transformation numérique, il reste un travail de sensibilisation à mener.

Dans le domaine de la science des données, les évaluations du même sondage sont plus élevées. 78% des PME interrogées pensent que leurs données sont bien collectées et stockées, et 71% des PME sont sensibilisées aux problèmes de sécurité [Fortin, 2019]. Là où un problème persiste, c'est dans le recrutement des compétences. Même si pour la sécurité des données et la gestion des données, plus de 60% des répondants affirment avoir des compétences à l'interne, ce chiffre baisse drastiquement dans des domaines plus expérimentaux. Les PME ayant des compétences en science des données à l'interne ne représentent que 20% des intervenants sondés. Quant à l'intelligence artificielle, seule 15% des PME sondées ont cette compétence à l'interne [Fortin, 2019].

Ces chiffres montrent bien qu'il y a encore du travail à faire dans le domaine du numérique dans les industries. Des concepts généraux et complexes comme l'industrie 4.0 sont peu compris. Dans le domaine de la sécurité et de la gestion des données, les entreprises s'équipent

plus vite. Ceci peut être dû aux caractères critiques de ces domaines. La difficulté survient dans les domaines plus expérimentaux, qui demandent des compétences plus spécialisées. L'intelligence artificielle et l'analyse de données sont très peu utilisées dans les PME québécoise. Pour pallier à ces problématiques, le gouvernement a mis en place un plan d'action. L'un des axes de ce plan d'action est d'impliquer la recherche à développer des outils pour l'industrie. Faire le pont entre le domaine industriel et universitaire aidera à la transition numérique.

La transition numérique est un enjeu important dans la prospérité des entreprises. Les outils qu'elle apporte permettent de mieux visualiser la situation dans une entreprise, d'avoir un suivi sur les performances de l'entreprise. Intégrer des systèmes informatisés aide aux échanges avec les clients et les fournisseurs. Le rapport "Le passage au numérique" de Danjou et al. [2017] liste, parmi ces technologies, les systèmes cyber-physique, les mega-donnée, l'Internet des Objets (Internet of Things) (IoT) et l'intelligence artificielle. Ce sont des technologies qui se basent sur les données industrielles et qui font partie des clés pour entrer dans l'ère du numérique.

Les données industrielles ont un cycle de vie. Plusieurs étapes de ce cycle de vie sont sensibles d'affecter ces données et leur utilisation. Les étapes principales d'un projet de valorisation des données sont : la récupération des données industrielle, le stockage des données dans des bases de données, le nettoyage de ces bases de données, enfin la fouille et l'analyse des données. Chaque étape est critique à la réussite d'un projet numérique et plusieurs méthodes existent pour réaliser chacune d'entre elles. La revue de littérature explorera plus en détail ces méthodes. Une information importante est que le nettoyage d'une base de données avant son utilisation est la partie qui prend le plus de temps.

Un projet de récupération de données sert à mettre en place des processus qui remontent les informations du plancher de production. Ces données sont ensuite stockées dans une base de données permettant leur exploitation. Les données présentes dans les bases de données nécessitent d'être traitées avant leur utilisation. Des techniques de fouille de données sont utilisées pour explorer les bases de données, plus la base de données contient des données complètes et dénuées d'erreurs, plus la fouille sera aisée. Or dans le cadre des PME qui s'équipent en capteurs, les données récoltées ne sont pas des données standards et les équipements n'ont pas toujours été pensés pour intégrer les capteurs qu'une entreprise veut ajouter. Ceci pose la question de l'évaluation de la pertinence de la donnée qui est extraite ainsi que de sa qualité.

Pour entrer dans la vague du numérique, les PME doivent investir dans des projets souvent coûteux. Il est donc logique d'entreprendre des étapes préalables auxancements des projets pour bien définir les besoins et les objectifs. Parmi les précautions qui doivent être prises,

il faut s'assurer que les données utilisées sont bien conformes aux besoins de l'entreprise. Comme exploré plus tôt, plusieurs technologies et méthodes de l'industrie 4.0 se basent sur les données industrielles issues du plancher de production. Ce mémoire suit la volonté des entreprises à réussir leur transition numérique. Une méthodologie pour analyser la qualité des données industrielles récupérées sur le plancher de production est développée. Cette méthodologie vise à définir les besoins en données et à évaluer si les données récupérées répondent à ces besoins.

La revue de littérature reviendra sur les définitions de l'industrie 4.0. Ensuite, la notion de qualité des données sera définie. Les caractéristiques qui définissent la qualité de la donnée seront présentées. Enfin, l'avancement de l'intégration de l'industrie 4.0 dans les PME sera exploré. Des feuilles de route pour l'adoption de l'industrie 4.0 seront discutées. De cette revue de littérature, une problématique sur l'évaluation de la qualité des données sera dégagée. La problématique relèvera le manque en méthodologie pour évaluer la qualité des données pour un projet de transformation numérique. Pour y répondre, une méthodologie avec des outils est développée au chapitre 4. La méthodologie est constituée de 3 grandes étapes, soit la définition des besoins en qualité des données, le développement d'indicateurs (KPIs) pour évaluer cette qualité et enfin de la création d'un gabarit pour enregistrer les métadonnées décrivant la donnée récupérée. Enfin, des applications de cette méthodologie sont présentées au chapitre 5 lors de différentes études de cas.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Introduction

Depuis la troisième révolution industrielle, la robotisation a été largement adoptée. Depuis, les industries ont commencé à vouloir suivre et contrôler leurs productions numériquement. Pour cela, plusieurs technologies ont été développées. Le concept d'industrie 4.0 est le terme sous lequel ces leviers technologiques se regroupent [Danjou et al., 2017]. Ces technologies sont encore en train de se développer. Leurs utilisations sont étudiées dans les centres de recherche, cependant, les planchers de productions ne sont pas tous aussi avancés dans la numérisation de leurs données de production.

Un lien majeur existe entre les données et l'installation de l'industrie 4.0. Les technologies du numérique se basent principalement sur l'échange d'informations entre les machines de production et les systèmes d'information. La récupération, le traitement et l'envoi de données sont des aspects primordiaux à l'intégration des outils de la production intelligente.

Comme exploré dans l'introduction, l'industrie 4.0 présente plusieurs défis d'installation. Ces défis se trouvent dans les technologies à développer, les nouvelles connaissances à maîtriser et la difficulté de s'adapter aux changements qu'un tel paradigme apporte. Une PME qui veut introduire des technologies de l'industrie 4.0 va être confrontée à d'autres défis plus spécifiques. Dans le rapport Danjou et al. [2017], les auteurs attirent l'attention sur le manque de connaissances en technologies du numérique dans les PME. En effet, contrairement à une grande entreprise qui internalise des services informatiques, les PME n'ont pas souvent un département qui s'occupe des projets d'innovations technologiques. Les technologies développées par l'industrie 4.0 sont rarement des solutions prêtes à utiliser, un certain travail d'adaptation est nécessaire. L'existence de personnes capables de faire les suivis au sein de l'entreprise aide à la réussite de ces projets. Il y a donc une difficulté supplémentaire pour une PME pour se maintenir à jour avec les moyens de production actuels.

Ce mémoire se concentre sur la place des données dans les technologies de la transformation numérique, particulièrement comment établir leur qualité. Les chapitres suivants présenteront d'abord une définition des concepts abordés, tels que l'industrie 4.0 et son interaction avec les données industrielles. Ensuite, la notion de qualité des données sera abordée ainsi que la capacité de l'évaluer. Enfin, il s'agira de présenter l'état actuel de l'industrie 4.0 dans les PME, ainsi que leur capacité à exploiter les données.

2.2 La place des données dans les technologies de la transformation numérique

La transformation numérique est un enjeu de plus en plus important pour les entreprises. Les industries de production et de service voient leurs chaînes de valeur changer par l'effet du "tout numérique". Plusieurs expressions font référence à la transformation numérique, une part de la littérature s'accorde même à parler de révolution industrielle. L'ère de l'industrie 4.0 fait référence aux changements que la transformation numérique apporte.

La définition de l'industrie 4.0, telle que citée par Hermann et al. [2015] est : "L'Industrie 4.0 est un terme qui rassemble les technologies et les concepts d'organisation de la chaîne de valeur. Au sein des "Smart Factories" à structure modulaire de l'Industrie 4.0. Les CPS surveillent les processus physiques, créent une copie virtuelle du monde physique et prennent des décisions décentralisées. Grâce à l'IoT, les Système Cyber-Physique (CPS) communiquent et coopèrent entre eux et avec les humains en temps réel. Via le Système d'Exploitation pour la Connexion des Réseaux (Internet of Things Operating System) (IoS), des services internes et interorganisationnels sont proposés et utilisés par les participants de la chaîne de valeur." Cette définition, qui résume les recherches effectuées par Hermann et al. [2015], nous indique que pour comprendre l'industrie 4.0, il faut comprendre les technologies qui sont en jeu.

Dans ce but, la section 2.2.1 présente les leviers technologiques de l'industrie 4.0. la section 2.2.1 présente aussi les différentes étapes qui mènent vers une usine intelligente. Les technologies ainsi que les étapes de leurs incorporations dans une usine seront présentées. La section 2.2.2 présente l'omniprésence des données dans les leviers technologiques. La section 2.2.3 s'attarde sur les enjeux stratégiques de l'utilisation et la gestion des données.

2.2.1 Les leviers technologiques de l'industrie 4.0

Le rapport Danjou et al. [2017] liste un ensemble de leviers technologiques sur lesquels s'appuie l'industrie 4.0. La liste qui suit cite 10 groupes technologiques :

1. Les données massives se caractérisent par leur volume, leur variété et leur vitesse. Une quantité énorme de données issue de sources hétéroclites et reçues à un rythme soutenu. Les traiter est un défi qui requiert des méthodes nouvelles pour en extraire de la valeur
2. L'intelligence artificielle est un domaine d'étude visant à doter les ordinateurs d'intelligence. Plusieurs définitions de l'intelligence artificielle existent. Elle permet aux machines d'accomplir des tâches qui demandent de la réflexion aux humains ou d'agir de façon rationnelle dans différentes situations[Russell and Norvig, 2010]
3. Les systèmes cyber physiques sont des réseaux d'éléments physiques interconnectés

entre eux. Cette connectivité permet de réguler les intrants et les extrants de ces éléments. [Chang et al., 2015]

4. La communication intermachines dis Le M2M se base sur des réseaux locaux pour la communication entre machines.
5. Les machines autonomes : Sont les machines qui permettent la connectivité et l'autonomie.
6. L'internet des objets est une application qui permet de connecter les objets aux réseaux internet et ainsi permettre leur autonomie. Cette application donne une identité propre à chaque objet physique ou virtuel sur le réseau internet. Grâce à cette indexation, il est possible d'accéder aux informations du monde physique et virtuel, mais aussi d'interagir avec le monde physique. [Saleh, 2018]
7. L'infonuagique est un outil permettant d'externaliser certains services informatiques tels que le stockage ou la puissance de calcul. L'infonuagique permet d'accéder aux ressources informatiques de partout et à tout moment.
8. La cybersécurité prend de l'importance car les risques liés à des cyberattaques augmentent. La protection des données, mais aussi des machines connectées devient importante avec la démocratisation de la connectivité.
9. La simulation permet d'effectuer des essais sur une copie numérique de l'usine. En combinant les technologies de connectivité et de récupérations des données, on peut avoir un jumeau numérique de l'usine.
10. La réalité augmentée permet de visualiser des informations supplémentaires tout en interagissant avec le monde extérieur.

Dans le domaine manufacturier, l'application de ces concepts de façon concrète donne naissance au "Smart Manufacturing" [Etz et al., 2020]. Etz et al. [2020] différencie l'industrie 4.0 et le "Smart Manufacturing". L'industrie 4,0 est une vision stratégique de l'application des technologies du numérique, elle contient toutes les technologies et techniques faites pour introduire la connectivité et l'informatique dans l'industrie. Les travaux faits dans le domaine de l'industrie 4.0 visent à inventer des outils et techniques pour utiliser les leviers technologiques. Le "smart manufacturing" est la réalisation concrète de cette vision dans les usines de production. Le passage au "smart manufacturing" requiert l'utilisation des leviers technologiques dans le cadre industriel. L'implémentation demande une préparation et une planification. Les technologies étant complexes et coûteuses, plusieurs recherches visent à définir une feuille de route vers une usine intelligente. Sufian et al. [2019] développe une feuille de route en 6 étapes. Les 6 étapes, appelées aussi 6 engrenages, sont [Sufian et al., 2019] :

1. **Stratégie** : Les concepts de l'industrie 4.0 attirent les entreprises en offrant un meilleur contrôle de la production, un lien plus accru avec la demande et une adaptabilité rapide au marché. Cependant, pour transformer la théorie en projet concret, il faut définir les objectifs désirés. Le point de départ de chaque entreprise est différent ainsi que leurs priorités. Il faut bien choisir un plan d'action pour avoir un retour sur investissements avantageux. Au chapitre 2.4 des méthodes pour définir des stratégies seront présentées.
2. **Connectivité** : L'usine intelligente se base sur des technologies de connectivité. L'internet prend une place importante et omniprésente. L'ajout de capteurs et de machines capables de s'arrimer au système informatique complexifie le système informatique. Avec cette augmentation des ports d'entrée, le système devient vulnérable à des cyberattaques. L'engagement du département informatique dès le départ d'un projet de transformation numérique est requis.
3. **Système d'information intégré** : Les technologies de récupération de données sont de plus en plus présentes sur le plancher de production. Un pont entre les technologies d'opération et les technologies d'informations n'a jamais été aussi important. Des progiciels comme les Progiciel de Gestion Intégré (Entreprise Ressources Planning) (ERP) et les Logiciel de Pilotage de la Production (Manufacturing Execution system) (MES) permettent une fluidité dans le transfert d'information dans l'entreprise. Le MES gère la partie production et relie l'information au système ERP dans les bureaux. Intégrer toute l'information dans une seule base de données commune reste un défi de taille pour toute entreprise.
4. **Analyse de données** : L'une des caractéristiques de l'usine intelligente est la quantité de données en circulation. Ces données massives ne sont utiles qu'après traitement. Les techniques d'analyse de données permettent de transformer les données en informations utiles. Grâce aux informations, des décisions plus précises et efficaces sont prises. La difficulté est donc de gérer ce flux de données et de s'assurer de sa qualité pour baser l'analyse. De plus, il faut trouver le bon format pour présenter les informations.
5. **Intelligence artificielle** : Dans les entreprises de production l'IA est souvent l'outil utilisé pour faire évoluer la production automatique à la production autonome. L'IA sert à augmenter les capacités cognitives des machines et des humains. La perception, le raisonnement, l'apprentissage et la prise de décision autonome sont les buts de l'IA. Actuellement, l'IA est utilisée pour faire de la maintenance prédictive et aider les machines de production à se réguler en implémentant des règles. Cependant, un grand enjeu se trouve dans la disponibilité et la qualité des données utilisées pour les modèles de l'IA.
6. **Évolutivité** : L'usine intelligente ne veut pas se contenter de suivre l'état des machines

utilisées. Utiliser l'IOT permet de faire des liens avec plusieurs facteurs qui sont extérieurs à l'entreprise. Tirer profit du réseau d'informations aide à faire évoluer les sources d'informations et améliore la compréhension des enjeux de productions.

Ces engrenages donnent une idée de ce qui est attendu pour les usines de nouvelle génération. Ce qui est observable c'est l'omniprésence des données dans ces engrenages. Dans la section suivante, l'importance des données est mise en avant. Les données massives sont présentées en plus d'un passage sur la fouille des données. Ensuite des applications des données sont présentées comme dans l'IA et l'IOT.

2.2.2 La places des données dans l'industrie 4.0

Les données générées par les machines de production sont de plus en plus nombreuses. Ce fait est en train de mettre en difficulté les bases de données standard. De plus en plus, les méthodes de gestion des données massives sont en train d'être appliquées [Winkler et al., 2021]. L'utilisation des données massives est encouragée par les entreprises, car elle permet une prise de décision mieux informée [Jagatheesaperumal et al., 2022]. Ces données massives permettent entre autres d'appliquer les techniques de valorisation des données et de l'intelligence artificielle (IA) [Jagatheesaperumal et al., 2022].

La fouille des données est un outil important pour extraire des informations dans des bases de données massives. La fouille de données consiste en plusieurs algorithmes qui permettent de valoriser les données brutes. Vu que la présence de données massives empêche le traitement manuel, faire des liens entre les données et en extraire de l'information se fait par des programmes informatiques. Ces algorithmes utilisent plusieurs méthodes pour permettre à un utilisateur de valoriser les données, la transformer par la suite en informations intelligible et enfin acquérir du savoir. Même si ces algorithmes sont le cœur de la fouille des données ils ne représentent que 15% à 25% de l'effort [Braha, 2013]. En effet, des étapes préalables et ultérieures sont nécessaires. Avant de pouvoir appliquer les algorithmes de fouille, il faut comprendre le milieu d'étude. Comprendre où et comment les données sont stockées ainsi que leur origine permet une meilleure analyse par la suite. L'outil utilisé dans la fouille dépendra du contexte. Après, il faut sélectionner les données qui seront analysées. Ensuite, nettoyer la base de données et s'assurer que les données sont fiables, qu'il ne manque pas d'information et qu'il n'y a pas de doublons. Finalement, il faut préparer la fouille en prenant des décisions sur les données qui seront mises en valeur. Cette préparation vise à s'assurer que les données sont de bonne qualité et prêtent à être analysées. Avoir un système de récupération de données qui favorise la qualité des données permet de diminuer le temps perdu. Ce n'est que là où les algorithmes de fouille peuvent être utilisés [Braha, 2013]. Après la fouille, il faut

interpréter les résultats et les présenter.

L'IA peut être utilisée dans l'industrie pour prédire la qualité, pour faire de la maintenance prédictive, pour organiser la production intelligente, pour être implémentée dans la robotique industrielle, pour créer des conceptions générées par l'IA et enfin pour mieux prédire le marché [Jagatheesaperumal et al., 2022]. Ces outils se basent principalement sur l'utilisation des données industrielles. L'IA distille la donnée industrielle récupérée et la valorise. Les données sont donc traitées pour améliorer le contrôle des processus dans l'industrie. Ce contrôle se base sur la réduction des arrêts de production non voulus et l'augmentation de la productivité [Angelopoulos et al., 2022].

L'IoT permet d'automatiser l'échange d'information entre machine et humain. L'IoT permet aux CPSs de communiquer avec le monde environnant [Danjou et al., 2017]. Ces deux technologies permettent la récupération de données, le partage de ces données, leur analyse, et enfin, de prendre des décisions se basant sur ces données. C'est ce qui permet à l'industrie 4.0 de prendre des décisions décentralisées [Danjou et al., 2017].

Toutes les technologies citées comme leviers de l'industrie 4.0 utilisent les données industrielles. Ces données sont valorisées et permettent un meilleur suivi des processus et des prises de décisions éclairées. D'autres outils 4.0 entrent en jeu dans la gestion des données. Par exemple, l'infonuagique permet de stocker des données et de les traiter avec une puissance de calcul supérieure au besoin. Quant à la cybersécurité, elle permet de protéger les données sensibles.

Pour résumer, les données ou l'information sont au centre de la transformation numérique. Les technologies qui supportent l'industrie 4.0 se basent sur cet échange de données. Les leviers technologiques présentent les moyens de récupérer, stocker, analyser et protéger les données de l'entreprise. Cependant, la question qui n'est pas traitée est l'utilité de tout cela. Scott [2000] dit que le savoir humain est exploitable directement, car il vit dans l'esprit des personnes le possédant. En opposition au savoir dans les bases de données, qui est un savoir figé, qui n'est pas exploitable directement. Cette distinction ouvre le regard sur l'importance du traitement et de la valorisation des données industrielles. Pour que l'idée de l'industrie 4.0 soit utile, les données doivent être transformées en connaissance claire, sûre et exploitable. Pour comprendre cette exploitabilité, Sen [2001] fait la distinction entre données et informations. La donnée est l'extrait d'un système qui mesure des faits sur des événements, des personnes ou des opérations. L'information est de la donnée valorisée qui permet de prendre des décisions informées. Cette valorisation se fait en mettant en contexte de la donnée brute ou en l'analysant.

Une fois la distinction entre données et informations faite, il est possible d'appliquer de

méthodes pour transformer la donnée en informations. L'analyse de données, présentée précédemment, en fait partie. Sen [2001] présente les utilités diverses de l'analyse de données. Les systèmes qui se basent sur les informations soutiennent la réingénierie des processus avec efficacité. La compréhension des processus de production et d'affaire, permise par les informations compilées, permet de découvrir plus facilement les changements à apporter. D'autres outils, comme la régression et l'intelligence artificielle permettent même la prise de décision décentralisée. Toute cette gestion des données massive donne la possibilité de se rapprocher des besoins du client.

Il est important de souligner l'imbrication de tous ces outils. En effet, pour former un système d'aide à la prise de décision efficace, il ne suffit pas d'avoir quelques machines intelligentes qui fonctionnent en silo. Il y a un certain nombre de critères que les données d'une industrie doivent atteindre, pour avoir un système d'aide à la prise de décision intelligent. Dans l'article de Guerlain et al. [2000], six critères sont présentés : (1) les différentes parties prenantes utilisant les données doivent interagir ; (2) les événements et les changements doivent être détectés automatiquement et les preneurs de décisions doivent être alertés ; (3) les données doivent être représentées efficacement pour une prise en main facile ; (4) les erreurs doivent être détectées pour minimiser les mauvaises interprétations ; (5) l'information doit être extraite correctement à partir de la donnée brute ; et enfin (6), des tendances ou résultats doivent être prédits en se basant sur le travail en amont. Tous ces éléments mènent vers une meilleure prise de décision, car ils offrent du savoir facilement utilisable par les gestionnaires. Cependant, l'instauration de ces systèmes de prise de décisions reste délicate. Il existe des enjeux liés aux données et à leur exploitation. La sous-section suivante explore les sources de difficultés pour l'exploitation des données.

2.2.3 Les enjeux stratégiques pour les données

Comme vu précédemment, les données industrielles peuvent servir à analyser divers processus d'une entreprise. Le but d'utiliser ces données est de mieux comprendre son entreprise et par conséquent prendre des décisions stratégiques plus éclairées [Ghasemaghaei, 2019]. Selon l'ampleur, les décisions seront prises de façon plus ou moins décentralisées. L'idée d'automatiser certaines décisions en se basant sur les leviers technologiques de l'industrie 4.0 reste valide pour les décisions de type opérationnel. Cependant, les décisions à moyen et long terme seront souvent prises par des gestionnaires ou des investisseurs. Le flux d'information doit réussir à circuler pour atteindre ce niveau décisionnel. Sen [2001] affirme que l'une des problématiques sera d'aligner les besoins en informations des décideurs et les informations produites par les gestionnaires du système d'information. Pour réussir cet alignement, il faut

que la stratégie du système d'information reflète la stratégie de l'entreprise. L'infrastructure du système d'information doit suivre le schéma organisationnel de l'entreprise.

Après un alignement des besoins en matière d'information, il reste à donner la possibilité d'explorer les données sous plusieurs angles. Pour un preneur de décision, cette exploration doit être simplifiée et éviter le passage par les bases de données. Ayoubi and Aljawarneh [2018] insiste sur l'importance d'utiliser des outils de Informatique Décisionnelle (Business Intelligence) (BI). La "Business Intelligence", ou "informatique décisionnelle" est un ensemble d'outils informatiques qui soutient l'analyse de données dans les entreprises. La BI facilite la prise de décision, car elle présente et synthétise l'information obtenue. Ces outils informatiques permettent de représenter les données de façon intelligible et en extraire l'information. Au-delà de la simple représentation, la BI permet aussi une autonomie dans l'exploration de l'information. Comparer les informations, générer des graphiques instantanément et changer l'étendue d'une analyse peut donner plus de clarté lors d'une prise de décision.

Les sources des données industrielles sont un enjeu important pour la maîtrise de l'information. Yuan et al. [2021] liste les différentes sources de données existantes. Quatre sources peuvent être identifiées, les capteurs, les entrées de données humaines, les bases de données existantes et les systèmes de contrôle. Le tableau 2.1 liste les sources, explique leur utilisation et donne des exemples d'utilisation.

TABLEAU 2.1 Sources de données, d'après [Yuan et al., 2021]

Sources de données		
Sources	Description	Exemples
Capteurs	Données objectives, récupérées par des capteurs analogues ou numériques.	Températures, voltage, humidité.
Bases de données existantes	Données récupérées dans des systèmes d'information comme les ERP.	Nomenclature, matières premières.
Entrées humaines	Données entrées par des humains en utilisant les interfaces humain-machine.	Qualité.
Système de contrôle	Données récupérées grâce au système de contrôle de certaines machines de productions.	Date de production, quantités.

Le tableau 2.1 liste les sources de données dans la colonne de gauche. La deuxième colonne contient des explications relatives aux sources. La dernière colonne liste certains exemples de données récupérées par ces sources de données. Le tableau montre que les sources doivent coexister pour avoir une image complète de l'entreprise. Mais, chaque source de données aura des avantages et inconvénients. Par exemple, la source humaine est subjective et les systèmes de contrôle n'affichent pas toujours toutes les informations désirées.

Les capteurs jouent un rôle de plus en plus important dans la récupération des données. Ils représentent la partie sensorielle de tout système IoT [Teh et al., 2020]. De plus en plus de recherches sont effectuées pour améliorer la qualité des données directement récupérées par les capteurs. La qualité des données était étudiée dans les bases de données. Des problèmes comme les données manquantes ou les données erronées sont traités par diverses méthodes de nettoyage des bases des données. Or, il y a certains critères de qualité des données qu'il faut traiter en amont en améliorant l'utilisation des capteurs [Teh et al., 2020]. Ces critères de qualité seront détaillés dans la section 2.3. Avec l'accessibilité des moyens de récupération de données, les données industrielles sont de plus en plus volumineuses. Les problématiques de stockage sont des enjeux qui prennent de l'ampleur [Winkler et al., 2021]. Des solutions comme l'infonuagique et les données massives se démocratisent. Ces bases de données doivent être pensées pour supporter les données massives, mais aussi pour être accessibles et facile à utiliser. Les problèmes de qualité des données surviennent aussi dans le stockage. La perte d'information, la pollution des bases de données et les erreurs sont des problèmes récurrents. L'économie actuelle s'appuie fortement sur la présence de PME. Ces PME doivent rester compétitives en améliorant l'utilisation de leurs données. Coleman et al. [2016] énumère les difficultés qui rendent les PME moins efficaces pour utiliser les données. Les données massives requièrent un certain ensemble de compétences pour être utilisées. Une PME à des particularités qui la mettent en difficulté, par exemple :

- Manque de compréhension : Les données ne sont pas comprises et l'utilité de l'analyse de données est remise en question ;
- Travailleurs très spécialisés : Les PME sont des industries spécialisées dans un domaine d'activité précis et les travailleurs ont généralement des compétences éloignées de l'analyse de données ;
- Manque d'experts des données à l'interne : Les PME manquent d'unités spécialisées en analyse et exploitation des données ;
- Manque de progiciels accessibles : Les progiciels présents sur le marché restent complexes à utiliser et implémenter ;
- Etc.

Ces difficultés montrent qu'une PME a besoin d'outils et de méthodologie simples pour utiliser la donnée. Le manque d'expert relègue la gestion des projets d'exploitation de données à des preneurs de décision inexpérimentés. Il y a une volonté de l'industrie à rendre l'analyse de données accessible pour les entreprises. Pour cela, il est important de fournir des guides pour soutenir la prise de décision. L'évolution de la transformation numérique dans les PME sera explorée en profondeur dans la section 2.4.

2.3 La notion de qualité des données

2.3.1 Concept de la qualité de l'information

Sen [2001] dit qu'avant de mettre en place une étude de qualité il faut d'abord définir ce qu'est la qualité. La qualité d'une donnée est associée à sa capacité à répondre au besoin d'une analyse de données [Yuan et al., 2021]. Ce concept de qualité remet la définition de la qualité sur les différents besoins que peut avoir une application d'analyse de données. Reposer la définition de la qualité des données uniquement sur la précision d'une mesure est réducteur. Comme présenté dans la section "Enjeux stratégiques pour les données", il y a plusieurs sources de données possibles. Pipino et al. [2002] affirme qu'il y a des sources subjectives et d'autres objectives. Cette différence permet d'ajuster son rapport à la donnée présentée.

L'article de Yuan et al. [2021] contient une liste des caractéristiques à prendre en compte pour définir la qualité d'une donnée. Une partie importante du maniement de la qualité d'une donnée est de pouvoir définir des attentes de qualité. Avoir une liste de caractéristiques aide à définir des attentes dans différents domaines. Cette liste contient [Yuan et al., 2021] :

- Accessibilité : décrit si l'analyste a la possibilité d'accéder à une donnée qui est enregistrée. L'inaccessibilité peut être due à cause de problèmes techniques ou d'autorisations ;
- Précision : La donnée doit être une valeur numérique qui décrit bien le phénomène voulu. Elle doit aussi être présentée selon la nomenclature précise de la base de données. Si ces deux critères ne sont pas remplis, un analyste ne peut pas utiliser les données ;
- Quantité de données appropriées : Pour effectuer du data-mining, une certaine quantité de données est nécessaire. La quantité de données nécessaire dépend de la complexité de la donnée et de la méthode utilisée ;
- Disponibilité : Les données doivent être disponibles et stockées dans une base de données. S'il y a un problème dans le système de récupération, les données ne pourront pas être exploitées ;
- Crédibilité : À quel point le signal récupéré par les capteurs est vrai et de confiance. Cela revient à déterminer la calibration des capteurs et le risque d'erreurs. Des données corrompues ne doivent pas être utilisées ;
- Données complètes : Décrit la possibilité d'une perte de données lors d'une récupération sur une période ;
- Consistante : Indique que l'ensemble des données sont cohérente et ne se contredit

- pas ;
- Données pertinentes : Représente l'importance des données utilisées pour le but de l'analyse. L'utilisation de données non pertinentes ralentit le processus de data mining ;
- Traçabilité : Capacité à retracer les changements subis par les données. L'incapacité de suivre l'origine des données peut compromettre l'analyse ;
- Variabilité : La présence de variabilité dans les données récupérées indique la présence d'information. Si la donnée récupérée est constante, peu d'information peut en être récupérée ;
- Représentation de la réalité : La représentation que les capteurs donnent est proche de la réalité ;
- Objectivité : Les données sont récupérées avec des outils objectifs. La donnée ne représente pas un avis, mais un fait ;
- Réputation : La source de la donnée à une bonne réputation. Quelles soit humaine ou informatique la source doit être de confiance ;
- Sécurité : Représente la sécurité d'un ensemble de données. Des données récupérées qui sont à risque de compromettre la sécurité des bases de données sont dangereuses ;
- Données compréhensibles : La capacité de comprendre le sens des données qui se trouve dans la base de données. Reflète le respect des nomenclatures et l'utilisation d'annotation pour expliquer les entrées ;
- Marge d'erreur : Représente la tolérance des écarts de mesure acceptés.

2.3.2 Qualité de l'information

Les caractéristiques de la qualité des données vue précédemment peuvent être catégorisées de plusieurs façons. Kirchen et al. [2017] les divise en caractéristiques en aval et en amont. Les caractéristiques en amont jugent la capacité de récupération de données. Les caractéristiques en aval jugent la capacité à analyser les données. Des articles comme Goodhue [1995], Lee et al. [2002], Strong et al. [1997], Wang and Strong [1996] donnent des exemples de caractéristiques et les évaluent selon des dimensions. Al-Hakim [2007] résume ces travaux et définit quatre dimensions avec lesquelles analyser Qualité de l'Information (Information Quality) (IQ) :

Intrinsèque : Qualité propre à la donnée recueillie. Exemple : la précision, l'objectivité et consistance.

Contextuelle : La qualité de la donnée doit être mise en perspective avec l'objectif du projet dans lequel elle est utilisée. Exemple : importance, valeur ajoutée, niveau de détail.

Accessibilité : L'information est accessible et facilement compréhensible. Exemple : accessibilité, facilité d'utilisation, compatibilité avec le système, sécurité.

Représentation : L'information est représentée de façon concise et facilement compréhensible. Exemple : clarté, format de qualité, consistance de la représentation.

Ces articles viennent apporter de la profondeur et de la nuance à l'analyse de la qualité d'une donnée. En effet, si à une époque la qualité était synonyme de précision, une vision plus complexe est adoptée. Observer la qualité de données sur tout le cycle de vie de la donnée ainsi que dans toutes ses dimensions, change le rapport à avoir avec la qualité. L'intérêt de comprendre le but de l'utilisation d'une donnée devient important. La qualité d'une donnée correspond alors à la capacité de l'utiliser. Ce sont tous les éléments cités plus haut qui indiquent si son utilisation sera avantageuse.

L'article de Alrae et al. [2020] remet la IQ en relation avec L'IoT. Il propose de mettre en place un cadre sur lequel bâtir la qualité d'information. Cette qualité dépend du système d'information utilisé, de l'efficacité avec laquelle il est utilisé ainsi que de sa complexité. La présence de L'IoT rajoute des niveaux de complexité.

2.3.3 Indicateurs de performance clés : KPIs

Pour évaluer la qualité de données plusieurs outils sont possibles. Durant la revue de littérature, les KPIs sont apparus comme l'outil d'évaluation le plus utilisé par les entreprises. Aborder un outil qui est déjà présent dans les pratiques industrielles est une bonne chose. Une question se pose cependant : existe-t-il dans la littérature des KPIs qui décrivent la qualité d'une donnée ? Il se trouve que dans le domaine médical, où les données jouent un rôle capital dans la survie des patients, des KPIs ont été explorés. L'article Stausberg et al. [2019] liste des propositions de KPIs pour évaluer la validité des données utilisées dans le domaine médical. Une revue de littérature systématique des indicateurs de performances indique la présence de 34 KPIs différents présents dans 31 articles. Dans ces articles, les KPIs qui reviennent le plus souvent sont la précision, la clarté et l'exactitude. De plus, un guide d'utilisation des KPIs a été mis au point par "Technology, Methods, and Infrastructure for Networked Medical Research"(TMF) [Stausberg et al., 2019]. Ce guide contient 51 indicateurs divisés en 3 catégories. La première catégorie est l'intégrité, elle rassemble les KPIs indiquant que les données récoltées correspondent à la norme et n'ont pas de valeurs aberrantes. La deuxième catégorie est l'organisation, elle s'assure que les données sont bien identifiables, par exemple qu'il n'y ait pas de redondance ou de synonymes. La troisième catégorie est la véracité, elle s'assure que la donnée est précise et provient d'une source d'information fiable.

Cette revue de littérature montre que la qualité des données peut effectivement être évaluée. La notion d'une donnée de qualité existe bel et bien et peut aussi être mesurée. Dans un domaine comme le médical, juger de cette qualité de la donnée est une priorité. Avec l'introduction du numérique dans les hôpitaux, la gestion de la qualité de la donnée est devenue une priorité. Or ce qui est observé est l'abondance des KPIs développés. Les KPIs doivent évaluer au mieux la donnée ; or la nature de la donnée dépend du milieu d'extraction. Il n'existe pas une banque de KPIs universelle applicable aux données provenant de toutes les industries. Il est donc intéressant d'observer comment des KPIs sont choisis et développés.

L'article de Mourtzis et al. [2018] parle de l'importance de bien choisir les KPIs qu'une entreprise utilise. Les KPIs permettent de traquer les avancements d'un projet, quantifier la qualité d'un produit ou d'un processus. Les KPIs constituent une aide pour la décision car ils donnent une représentation parlante des données. Cependant, il faut savoir choisir les données à suivre et surtout les KPIs à mettre en place. Plusieurs entreprises ne voient pas d'amélioration après la mise en place de KPIs, car leurs choix n'ont pas été judicieux. Le livre de Parmentier [2015] présente une méthodologie pour bien choisir ces KPIs. Il faut définir les besoins en matière de suivi, impliquer le monde pour avoir une participation accrue et mettre en place une équipe pour définir les KPIs. Il est facile de voir dans la littérature le rôle des KPIs et leur importance pour la représentation de données.

Les KPIs sont actuellement utilisés par les organisations comme un guide pour atteindre des objectifs de développement [Larsson et al., 2021]. Ces KPIs vont souvent se concentrer sur des aspects financiers et pour permettre le suivi des actifs tangibles. Cependant, mesurer des actifs intangibles comme le savoir-faire ou la robustesse du système d'information représente un défi plus complexe [S. and P., 2004]. La qualité des données produites sur le plancher de production fait partie des actifs intangibles qu'il serait intéressant de suivre. Dans la littérature, les chercheurs n'ont eu de cesse de développer des KPIs pour toute sorte de mesures. Des listes de ces KPIs existent pour tout type d'industrie et pour les différents aspects organisationnels d'une entreprise. Une recherche sur Compendex associant les mots clés "KPIs" et "Framework" affiche 448 résultats. Ces résultats sont des articles expliquant comment mettre en place une structure de KPIs efficace pour évaluer les performances d'une entreprise dans un cadre particulier. Ou encore, comment générer automatiquement les KPIs et les suivre. L'article de Efkarpidis et al. [2022] par exemple explique en détail le choix des KPIs et leurs élaborations pour suivre les systèmes énergétiques intelligents. Cet article montre la complexité d'établir des KPIs ainsi que l'ensemble des aspects à prendre en compte. La mise en place de KPIs doit prendre en compte l'avis des parties prenantes qui vont les utiliser. La mise en place doit aussi prendre en compte différentes échelles pour différencier les KPIs, qu'elle soit géographique ou temporelle.

D'autres articles s'intéressent en profondeur au choix des KPIs. La thèse de Zheng [2018] indique que plusieurs articles se penchent sur la façon de choisir les KPIs les plus intéressants pour une situation en particulier. Cette façon de faire vise à récupérer tous les KPIs disponibles sur un sujet et évaluer leurs pertinences. Les articles de Larsson et al. [2021] et Chorfi et al. [2021] offrent un exemple de ces méthodes de sélection. Les deux articles se basent sur l'utilisation de la méthode de prise de décision Procédure Hiérarchique d'Analyse (Analytic Hierarchy Process) (AHP). L'AHP est une méthode de prise de décision multicritère. Elle permet à un preneur de décision de comparer plusieurs solutions possibles en utilisant différents critères d'évaluation. Ainsi en comparant les KPIs grâce à l'AHP il est possible de choisir judicieusement les indicateurs. Cependant, la création de KPIs sur mesure est un sujet bien moins développé [Zheng, 2018]. La thèse de Zheng [2018] met en place une méthodologie pour développer de nouveau KPIs sur mesure. Cette méthodologie se base sur le guide Corps des Savoir en Gestion de Projet (Projet Management Body of Knowledge) (PMBok) qui offre des standards en termes de gestion de projet. Le guide désigne 10 aires de connaissances qui doivent être maîtrisées. Ces aires de connaissance sont la gestion de l'intégration du projet, la gestion de la portée du projet, la gestion du temps, la gestion des coûts, la gestion de la qualité, la gestion des ressources humaines, la gestion de la communication, la gestion du risque, la gestion de l'approvisionnement et la gestion des parties prenantes.

Pour créer les KPIs, Zheng [2018] utilise 6 étapes. La première est de spécifier les informations prioritaires. La deuxième est d'associer ces informations à des aires de connaissances. La troisième et la quatrième entrent en détail dans la structure du PMBoK. Il faut y spécifier des processus ainsi que des ensembles de mesures spécifiques à chaque aire de connaissance. La cinquième étape se base sur ces mesures spécifiques pour créer des indicateurs. La sixième étape consiste à intégrer les KPIs à la gestion de projet. Cette méthodologie est très complète et permet de mettre en place des indicateurs pour un projet, quel qu'il soit. Il est maintenant intéressant de voir comment ces projets se développent dans le cadre particulier des PME.

2.4 La transition numérique dans les PME

2.4.1 Les PME et le numérique

Il y a beaucoup d'articles sur l'industrie 4.0. Dans ces articles, les technologies mises en avant, les nouveaux processus, les méthodologies d'implantation sont présentés. Ces technologies sont en train de se démocratiser dans les industries possédant des budgets conséquents. Cependant, pour la réalité des PME l'adoption de ces technologies n'est qu'à ses début.

Dans le but de sensibiliser les PME aux principes d'industrie 4.0 le Ministère de l'Économie,

de la Science et de l'Innovation du Québec (MESI) a mis en place le projet industrie 4.0. Dans le cadre de ce projet, un rapport a été produit pour introduire l'industrie 4.0 au grand public. Ce rapport de Danjou et al. [2017] met en avant deux défis pour maîtriser le concept d'industrie 4.0 dans les PME :

1. Se projeter dans les utilisations possibles des nouvelles technologies, et imaginer l'impact qu'elles peuvent avoir sur les produits et les processus ;
2. Incorporer ces technologies dans les entreprises qui n'ont pas la maîtrise de ces outils, souvent loin de leurs domaines d'expertise.

Dans ce rapport, les auteurs expliquent qu'il y a différents niveaux d'intégration numérique. La figure 2.1 montre ces différents niveaux et s'inspire de l'article [Porter and Heppelmann, 2019]. Les différents niveaux de numérisation présentés permettent aux PME de découvrir les possibilités offertes.

La figure 2.1 présente 4 paliers évolutifs des utilisations les plus basiques aux plus complexes. Chaque entreprise qui veut maximiser son utilisation des systèmes numériques passe par le suivi, le contrôle, l'optimisation et l'automatisation de ses processus. La figure a une forme pyramidale qui reflète l'importance de respecter les paliers technologiques. Chaque palier sert de base pour le suivant, ainsi :

- Réussir à suivre la production et les processus d'affaires permet de commencer la phase de contrôle ;
- Si les opérateurs ont un contrôle efficace sur la production, l'optimisation devient envisageable ;
- Une fois que les algorithmes d'optimisation sont créés et que les capacités de production sont améliorées, il est possible d'automatiser ces processus.

Grâce à cette vision par palier de l'intégration des outils numériques, il est possible de classifier l'évolution des PME au Québec. Les chiffres présentés dans le sondage sur l'industrie 4.0 de Fortin [2019] montrent que la majorité des PME sont au niveau du suivi. Par exemple, l'enregistrement des données est effectivement maîtrisé, mais l'utilisation de ces données pour des applications ne l'est pas.

L'article de Zhou et al. [2020] parle de la réalité sur le terrain des PME. En citant Robert Lowe, l'article parle de la différence entre l'état souhaité et l'état actuel du plancher de production. Le plancher de production parfait, appelé le "green field", est équipé avec des machines de dernières générations capables de communiquer leur état. Ces machines permettent une récupération des données facile et rapide. Ces machines peuvent toutes communiquer entre elles. Mais cela ne se trouve pas dans la réalité du terrain où le plancher de production n'est pas aussi développé.

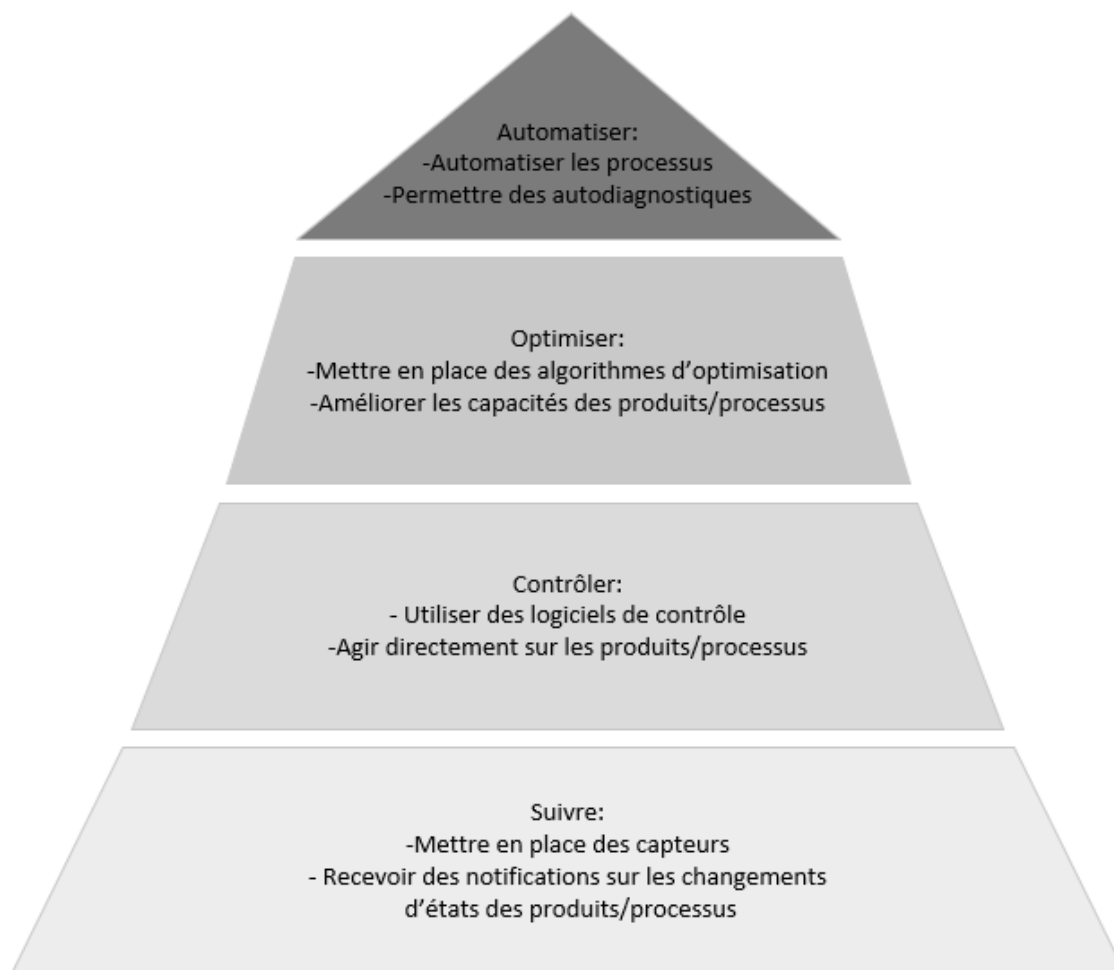


FIGURE 2.1 Capacités des systèmes - [Porter and Heppelmann, 2019]

Le plancher de production actuel est appelé "Brown field", il contient souvent des machines archaïques et les équipements ne sont pas toujours compatibles [Zhou et al., 2020]. Ces équipements qui permettent une bonne production ne sont pas toujours munis de capteurs et ne permettent pas une communication efficace. Généralement, la récupération des données se fait encore manuellement avec l'aide des opérateurs.

Les planchers de production ne sont pas tous au même niveau de préparation pour accueillir les technologies du numérique. Sachant cela, la littérature propose des façons pour évaluer cette préparation. En anglais "readiness assessment" pour l'industrie 4.0 est un ensemble de méthodes qui attribuent une note pour juger la préparation d'une entreprise à se lancer dans un projet de numérisation. La note évalue l'avancement dans le processus du virage numérique. Elle évalue aussi le milieu de l'entreprise et sa capacité à s'adapter au changement amené par l'industrie 4.0. Les articles de Simetingera and Basla [2022], Kolla et al. [2019],

Spaltini et al. [2022], Suleiman et al. [2021], R. et al. [2018] et De Carolis [2017] présentent plusieurs méthodes pour évaluer cette préparation. Ces articles présentent des Modèle de Maturité (MM). Ces modèles veulent juger de la maturité numérique d’une entreprise.

En général, plusieurs dimensions sont prises en compte et chacune est étudiée. Une liste non exhaustive de ces dimensions est listée ici :

- Technologie,
- Organisation,
- Exécution,
- Surveillance,
- Données,
- ...

Ces articles insistent sur la difficulté qu’une PME peut avoir à analyser son état d’avancement. En effet, les méthodes présentées (DREAMY 4.0, test I4.0 etc..) sont difficiles à utiliser dans les PME. Cela mène à plusieurs études mettant en place des adaptations de ces méthodes pour le cas particulier des PME. Après la définition de la maturité de la PME pour entreprendre sa transition numérique il faut entreprendre cette transition. Dans la section suivante, des méthodes pour mettre en place un plan d’action pour la transition numérique sont présentées.

2.4.2 Définition d’un plan d’action pour la transition numérique

Les MM servent de base pour construire un plan d’amélioration de l’entreprise. Connaître le point de départ permet de savoir les améliorations à apporter. Il reste cependant à déterminer un plan et prioriser les actions à entreprendre. Ici deux méthodes sont présentées.

2.4.2.1 Méthode 6P

Plusieurs méthodes de définition d’un plan d’action existent. Chacune de ces méthodes se base sur des outils particuliers. La littérature actuelle ne manque pas de telles références. Pour le travail actuel, il n’a pas été nécessaire de développer une méthode qui permet de mettre en place un plan d’action.

Une de ces méthodes est présentée dans l’article [Spaltini et al., 2022]. Les chercheurs présentent plusieurs méthodes de définition de la situation initiale. Ils s’inspirent de la méthode DREAMY 4.0 et ensuite mettent au point la méthode 6P. Cette méthode fait référence aux six dimensions à utiliser pour établir les objectifs à atteindre. La figure 2.2 résume la méthode 6P.

La figure 2.2 décrit la méthode en présentant d’abord les six dimensions. Ces dimensions

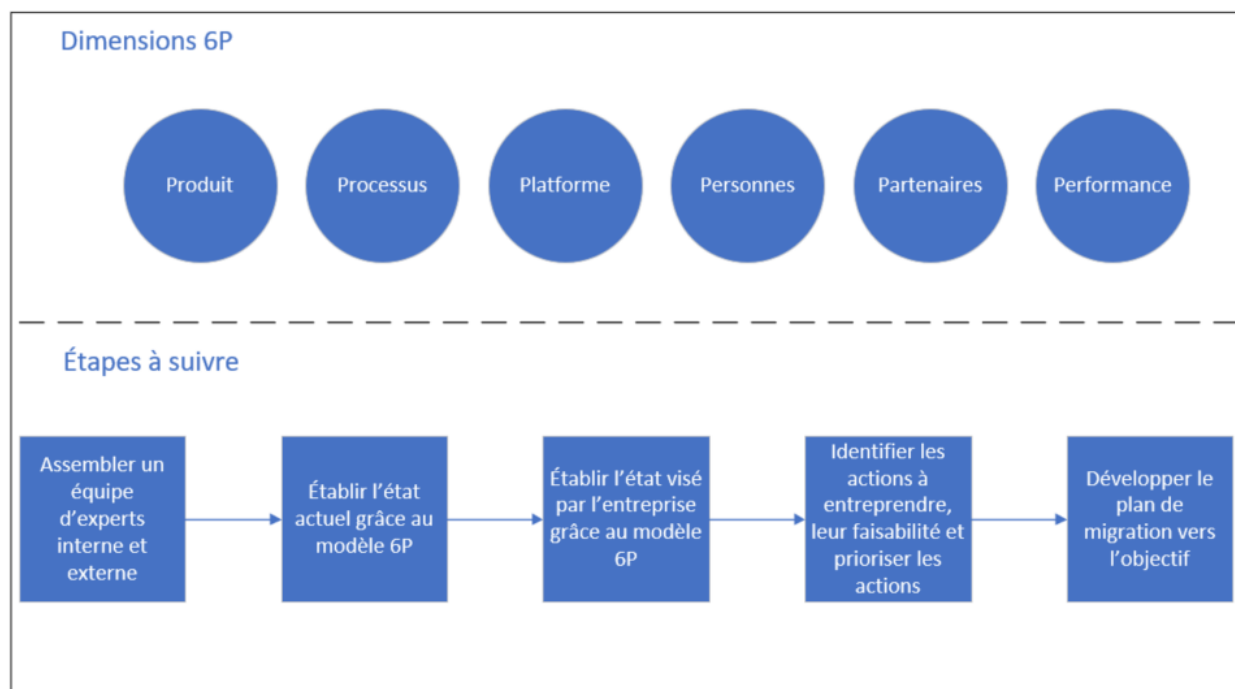


FIGURE 2.2 Méthode 6P-Traduite de [Spaltini et al., 2022]

sont ; le produit étudié, les processus entrant en jeu dans la chaîne de valeur du produit, les plateformes informatiques, les personnes clés, les partenaires d'affaires et les performances de l'entreprise. La méthode consiste à établir avec une équipe d'experts le portrait de l'entreprise en ce qui concerne la numérisation pour chacune des dimensions. Ensuite, il faut décrire les standards que l'entreprise désire atteindre. Pour combler cet écart, des projets sont définis et un plan est mis en place.

2.4.2.2 Méthodes *10 steps*

Le livre de McGilvray [2021], "Executing Data Quality Projects" présente lui aussi une méthode pour établir des objectifs d'amélioration. Ce livre a pour objectif l'amélioration de la qualité des données générées et utilisées par une entreprise. Les projets mentionnés dans ce livre s'inscrivent dans une volonté d'intégrer les outils de l'industrie 4.0, plus précisément celles de l'utilisation des données.

Dans ce livre, l'autrice présente des tableaux utilisés pour avoir une vision des manques en termes de données. Ces manques vont être des futurs projets sur lesquelles travailler. Une copie de ce tableau est présentée avec le tableau 2.2.

Dans le tableau 2.2 un groupe d'étude inscrit les besoins de l'entreprise en données. Certains

TABLEAU 2.2 Besoins QdD-Traduit de [McGilvray, 2021]

Numéro	Besoin de l'entreprise	Problème de QdD	parties prenantes des besoins				Objectifs du projet
			Processus	Personnes	Technologie	Données	
1							
2							
3							

de ces besoins ne sont pas du tout alimentés, d'autres ont des manques. La colonne "problème de Qualité des Données (QdD)" permet d'expliquer les manques en termes de qualité des données pour subvenir au besoin. Les parties prenantes expliquent les processus, personnes, technologie et données prises en compte dans le projet. Les objectifs du projet sont déterminés pour combler les besoins de l'entreprise. Au-delà du plan d'action des outils pour définir le cadre qui support les données de qualité existe. La section suivante introduit un outil pour décrire la structure qui soutient la qualité des données.

2.4.3 Structure pour la qualité de l'information

Dans le livre de McGilvray [2021], une structure pour la qualité d'information est présentée. La priorité est placée sur les données et leur gestion au sein de l'entreprise. C'est une méthode qui décrit la structure des données. Contrairement aux autres MM, le but n'est pas de donner une note, mais plus d'avoir une vue d'ensemble du cycle de vie de données. Le tableau 2.3 résume cette méthode. Lorsqu'un projet de valorisation de données prend place, ce tableau permet d'avoir une vue d'ensemble sur les données qui seront utilisées.

TABLEAU 2.3 Méthode POSMAD-traduit de [McGilvray, 2021]

Structure pour la qualité de l'information		cycle de vie de l'information					
		Planifier	Obtenir	Stocker et partager	Maintenir	Appliquer	Effacer
composants clés	Données (quoi)						
	Processus (comment)						
	Personnes (qui)						
	Technologie (comment)						

Le tableau 2.3 décrit l'information en utilisant deux axes. Le premier axe constitue les composants clés, soit la donnée qui constitue l'information, les processus qui interviendront pour traiter l'information, les personnes-ressources qui gèrent l'information et pour finir la technologie utilisée. Le deuxième axe représente le cycle de vie. C'est une description chronologique qui suit l'information de sa planification jusqu'à sa suppression. L'utilisation du tableau 2.3 permet de comprendre la structure des données au sein de l'entreprise. De plus, cette compréhension mène à l'identification des manques ce qui permet de planifier ce qui doit être consolidé pour améliorer la qualité des informations.

Ce tableau est présenté comme un outil de visualisation pour l'environnement complexe des données. Ce visuel permet de réaliser trois choses, le diagnostic de problème, la planification pour un nouveau projet, la communication des besoins pour de la donnée de qualité. Cet outil est utilisé dans la méthode des 10 étapes pour des données de qualité présentées dans le livre [McGilvray, 2021]. Cette méthodologie qui explique le cheminement d'un projet de récupération de données utilise des outils tels que le POSMAD pour visualiser le cadre dans lequel des données de qualité sont créées.

2.5 Conclusion

La revue de littérature a pu mettre en évidence l'importance que l'industrie 4.0 est en train de prendre. Il est aussi à noter que le terme industrie 4.0 renvoie à un regroupement de nouvelles technologies utilisées dans l'industrie. La maîtrise de ce nouveau paradigme de production passe par la maîtrise des leviers technologiques. Cela implique de s'équiper en machines capables de soutenir ces technologies. Il faut aussi s'équiper en logiciels de gestion de base de données et former les employés. Les progiciels se démocratisent pour la gestion des ressources, de la production et des clients. Tous cela implique le développement de nouvelles compétences à l'interne. Plusieurs études portent sur les étapes d'implémentations et les meilleures stratégies. Toutefois, un travail d'adaptation est nécessaire pour les PME. Les PME ont un ensemble de problématiques qui rend l'adoption de l'industrie 4.0 difficile. Il y a encore un manque d'outils pour permettre aux PME d'avoir accès aux nouvelles technologies.

Dans ce nouveau paradigme de production, la gestion des données fait partie des clés du succès. L'industrie 4.0 est bâtie sur la capacité de faire circuler les données et les valoriser. C'est la maîtrise des données qui permet de suivre, de contrôler d'optimiser et enfin d'automatiser les systèmes de productions. Comme présenter dans la revue de littérature ce sont ces quatre niveaux des systèmes de production qu'une entreprise vise en investissant dans l'industrie 4.0. Il est donc nécessaire d'avoir une maîtrise et une compréhension des données. Une entreprise génère toute sorte de données, mais ne maîtrise pas forcément leur qualité. Une entreprise a

besoin de comprendre ce que la qualité des données veut dire. Plusieurs études qui se sont penchées sur cette question montrent une pluralité dans le sens du mot qualité. Un besoin de décortiquer la qualité en plusieurs caractéristiques apparaît. Ces caractéristiques servent à mieux juger les données. De plus, les travaux faits à ce sujet montrent que la donnée ne peut pas être évaluée sans être confrontée à un contexte. Quatre dimensions sont utilisées pour juger la qualité des données. Ces dimensions englobent les caractéristiques présentées à la section 2.3.2. Il y a les caractéristiques intrinsèque, contextuelle, d’accessibilité et de représentation de la donnée.

Une fois le lien entre les caractéristiques et la qualité des données faite, un moyen de l’évaluer est nécessaire. Un outil se détache du lot, les KPIs. Ce sont des indicateurs de performance utilisés dans les entreprises pour évaluer et suivre des métriques importantes. Traditionnellement, les KPIs étaient utilisés pour juger les capacités financières et la performance de productions. Depuis, l’utilité des KPIs s’est diversifiée vers d’autres sphères de l’entreprise comme la santé et sécurité au travail, la rétention d’employé et la transmission des connaissances. Les KPIs sont un bon outil d’évaluation, car ils permettent de prendre en compte différentes caractéristiques. La littérature présente un exemple d’utilisation des KPIs pour juger de la qualité des données dans le domaine médical, mais cette voie reste inexplorée pour les données issues du plancher de production industrielle. Il y a donc un intérêt à explorer cette avenue. Le développement de nouveaux KPIs est aussi un sujet sous-exploité. Des méthodes pour choisir les KPIs les plus pertinents existent, mais développer des KPIs est bien moins courants.

Dans un projet de transition numérique, plusieurs étapes sont à couvrir. La compréhension du point de départ d’une entreprise est cruciale. Cette partie est couverte par la revue de littérature. Même dans le cas particulier d’une PME la revue de littérature adapte les outils d’analyse. Formuler un plan d’action est aussi couvert. Les manques commencent à apparaître au moment de parler des données industrielles. Les articles sur le traitement des données une fois dans des bases de données sont abondants dans la littérature, analyser leurs qualités lors de la récupération est moins courant. Le concept de qualité d’une donnée a déjà été défini. Des caractéristiques sont établies et l’importance des données de bonne qualité a été prouvée. Mais il manque cependant un lien entre cette théorie et la pratique du terrain. Comment définir les besoins de qualité des données pour un projet de transition numérique? Ensuite, comment vérifier que les besoins sont satisfaits? En somme, quelle méthodologie suivre pour évaluer la qualité des données industrielles dans le cadre d’un projet de transformation numérique?

Ce mémoire présentera la problématique plus en détail dans le chapitre 3, et expliquera les objectifs à atteindre. Le chapitre 4 présente la méthodologie proposée pour évaluer la qualité

des données. La méthodologie consistera à comprendre les besoins spécifiques en qualité de données pour un projet précis, puis à évaluer le système de récupération de données pour valider que les besoins sont atteints. Au sein de cette méthodologie des KPIs sont utilisés pour l'évaluation des données. Une méthode de développement simplifiée des KPIs est présentée pour répondre aux besoins des PME. Dans le chapitre 5, des cas d'étude issus d'une collaboration avec un partenaire industriel seront présentés. Ces cas d'étude mettent l'accent sur l'installation de capteurs et la réflexion faite pour atteindre des objectifs de qualité des données. Enfin une conclusion résumant les objectifs atteints conclura le travail.

CHAPITRE 3 PROBLÉMATIQUE

3.1 Retour sur la revue de littérature

La revue de littérature présente en profondeur les différents aspects de la transition numérique. Des concepts tels que l'industrie 4.0 et le SM sont définis. Le nouveau paradigme de l'industrie est porté par des technologies présentées dans la revue de littérature. Ces technologies sont : les données massives, l'IOT, l'intelligence artificielle, les systèmes cyber-physiques, la communication entre machines, les machines autonomes, l'infonuagique, la cybersécurité, la simulation et la réalité augmentée. Un des points saillants relevé au cours de cette revue est l'importance des données industrielles pour le fonctionnement de ces technologies.

Les données ainsi que leurs analyses sont devenues des ressources cruciales pour la transformation numérique. Cette façon de voir les données leur donne une nouvelle dimension. En effet, pendant longtemps les données étaient stockées dans le but d'être archivées, et étaient peu valorisées. Maintenant, avec le développement des technologies de l'industrie 4.0 cette valorisation est accessible aux entreprises. Les données ne sont plus des objets informatiques fixes et stockées mais une ressource évolutives. Il en ressort ainsi l'importance de gérer la récupération et le stockage des données en vue de leur valorisation. La qualité d'exécution de ces deux étapes, la récupération et le stockage, sont critiques pour minimiser le temps perdu lors de la fouille des bases de données. La revue de littérature explique que le nettoyage des bases de données et la correction d'erreurs liées à la récupération des données prend plus de temps que l'analyse de ces données.

Cette gestion des données passe par le fait d'identifier la qualité des données récupérées. La revue de littérature amène des éléments de compréhension sur le concept de qualité d'une donnée. La qualité d'une donnée est contextuelle et dépend de l'application dans laquelle la donnée sera utilisée. La qualité d'une donnée est définie grâce à plusieurs caractéristiques qui couvrent une variété d'éléments qui décrivent cette donnée. Une liste de caractéristiques a été présentée dans la revue de littérature.

Un outil d'évaluation de ces caractéristiques est nécessaire pour juger la pertinence de l'utilisation d'une donnée dans un projet de transformation numérique. La revue de littérature montre que les KPIs sont utilisés dans l'industrie pour évaluer la performance d'une entreprise. Ces KPIs peuvent aussi être utilisés pour évaluer la qualité d'une donnée. Une telle pratique a été observée dans le domaine médicale.

Enfin, la définition d'une feuille de route pour mener à bien des projets de transformations

numériques est un sujet qui est bien documenté. La revue de littérature montre que plusieurs méthodes existent. Dans les projets de transformation numérique qui impliquent la récupération de données, des feuilles de route existent pour définir les besoins en données, comment les récupérer, les stocker et enfin les gérer.

3.2 Problématique

La revue de littérature présente des outils pour mener à bien des projets de transformation numérique. Des feuilles de route sont proposées pour récupérer les données et les transformer en informations utiles. Des caractéristiques sont définies pour apprécier la qualité des données. Cependant, les feuilles de route ne prennent pas en compte la qualité des données. Ce manque est présent à plusieurs niveaux, tels que la définition de la qualité des données à récupérer ainsi que le suivi de cette qualité. Ce qui manque c'est l'intégration d'une étape pour évaluer la qualité d'une donnée lors du lancement et du suivi d'un projet de transformation numérique. Ainsi la problématique suivante se dégage :

"Comment développer une méthodologie qui permet : 1) de définir les besoins en qualité des données d'un projet de transformation numérique, 2) d'évaluer cette qualité et 3) d'en faire le suivi?"

Ce mémoire veut faire le lien entre les pratiques de gestion des projets de transformation numérique et la recherche autour de la qualité des données. Une méthodologie est développée pour répondre aux trois questions posées plus haut et apporter une aide décisionnelle lors du développement des projets de transformation numérique.

3.3 Solution proposée

Il est à noter que la méthodologie développée se place dans le contexte d'un projet d'amélioration de la qualité des données. Ainsi la définition des informations désirées et le projet de récupération de données sont des étapes préalables. Ces étapes sont documentées dans plusieurs travaux de recherche (présentées dans l'état de l'art - Section 2.4.2). La méthodologie développée définit la qualité des données à récupérer nécessaire pour l'obtention des informations désirées. Cette méthodologie peut être appliquée pour le lancement d'un nouveau projet d'installation de capteur ou pour évaluer la qualité des données récupérées par un capteur existant.

Le diagramme (Figure 3.1) illustre la méthodologie développée pour répondre à la problématique posée plus haut.

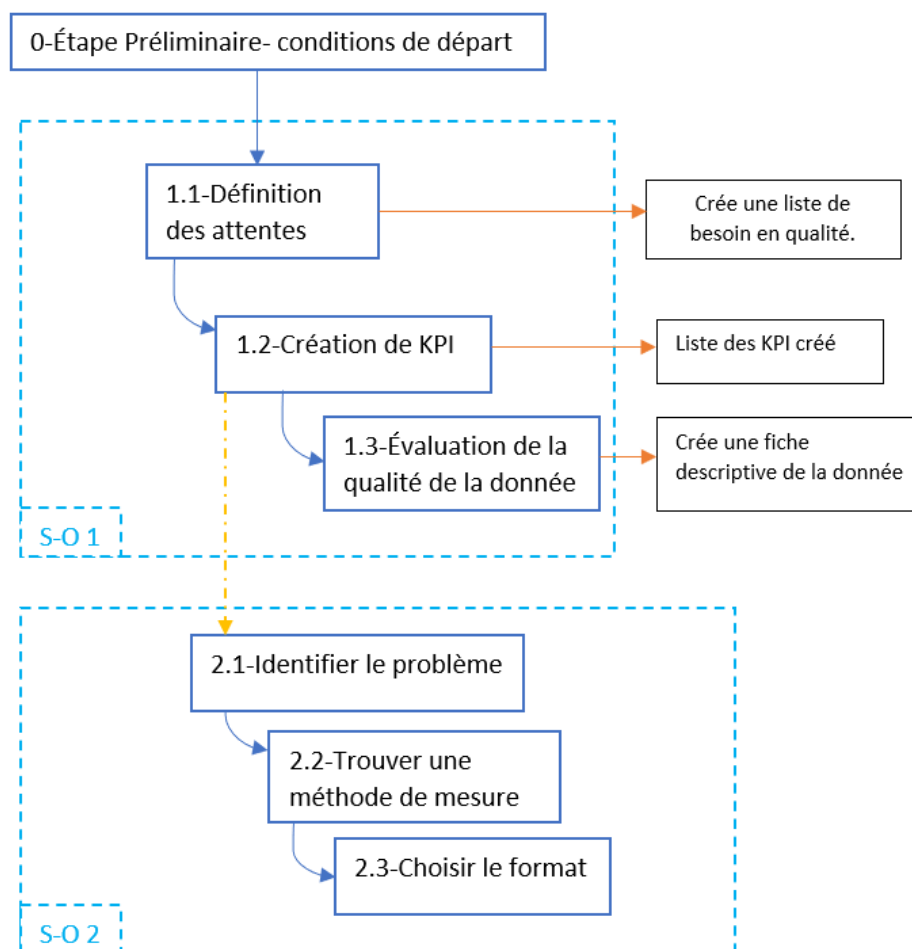


FIGURE 3.1 Méthodologie d'évaluation de la qualité des données

La figure 3.1 décrit les étapes à suivre pour évaluer la qualité d'une donnée pour répondre aux attentes d'un projet de transformation numérique. Deux sous objectifs permettent de répondre à la problématique. Le sous objectif 1 (S.O.1) vise à créer une méthodologie pour évaluer la qualité des données. Le sous objectif 2 (S.O.2) vise à créer une méthodologie pour développer des KPIs. Une étape préliminaire qui décrit les conditions préalables pour le lancement de la démarche est incluse.

Étape préliminaire - Conditions de départ Pour lancer un projet il faut faire une évaluation de l'état initial. Cette évaluation permet de mettre en évidence les points forts et les points faibles de l'entreprise. La finalité de cette étape est de créer une liste des manques et des besoins de l'entreprise.

Après l'évaluation de l'état initiale, il faut lister et sélectionner les projets à lancer. Le projet sélectionné doit être bien défini et documenté. Cela permet de préparer un

budget ainsi qu'un plan.

Étape 1 - Définition des attentes qualité : Dans le cadre d'une valorisation de données, il faut être conscient de la qualité des données utilisées. Cette étape est utilisée pour définir les standards attendus des données pour le projet sélectionné au préalable. L'objectif est d'avoir une réflexion sur ce qui est attendu des données. La qualité d'une donnée est liée à sa capacité à être utilisée dans un projet de valorisation de données. Les attentes en qualité d'un projet serviront à évaluer le niveau de satisfaction. À la fin de cette étape une liste des besoins en qualité des données est établie.

Étape 2 - Création de KPI : Une fois la liste des besoins créée il faut utiliser des outils de jugement pour évaluer la qualité de la donnée. Les KPIs seront utilisés à cet effet. Une méthodologie de création de KPIs en trois étapes est élaborée. Cette méthodologie est illustrée dans la figure 3.1 (S.O.2).

Étape 3 - Évaluation de la qualité de la donnée : Cette étape permet de produire une fiche descriptive des données analysées en utilisant les KPI mis en place.

Le chapitre suivant revient sur ces étapes et détaille leur application. Chaque sous-objectif est décortiqué pour avoir une feuille de route exploitable. Par la suite, le chapitre 5 mettra en application cette méthodologie.

CHAPITRE 4 MÉTHODOLOGIE D'ÉVALUATION DE LA QUALITÉ DES DONNÉES

Dans ce chapitre, la méthodologie d'évaluation de la qualité des données récupérées, présentée au chapitre précédent à la figure 3.1, est approfondie.

Comme relaté dans la revue de littérature, il y a une tendance à investir dans des projets de gestion de données. Ces projets participent à l'évolution de l'entreprise et la rendent compétitive. Cependant, ces projets sont coûteux et les résultats mettent du temps à être observés. Il y a dans la littérature des démarches qui aident à la planification de ces projets. Néanmoins, les démarches existantes ne prennent pas en considération la qualité des données. La méthodologie développée au cours de ce mémoire vient définir les besoins en qualité des données et offre des outils pour évaluer la qualité des données récupérées.

Au sein d'un projet de valorisation de données, plusieurs sous-projets de nature différente sont entrepris. Ce mémoire se focalise sur les projets visant à faire remonter de la donnée à partir du plancher de production dans les manufactures. Ces projets peuvent porter sur l'installation de capteurs, l'utilisation de technologies comme le code à barres ou la RFID, ou la récolte de données grâce au personnel sur le plancher de production. Les sources de données industrielles ont été présentées dans la section 2.2.3 de la revue de littérature.

Les données remontées vont servir dans des études pour faire le suivi de tout le processus de la production, faire des prévisions afférentes et ainsi prendre des décisions informées. Pour que ces études soient efficaces, la donnée utilisée doit être maîtrisée. Pour ce faire, avoir une compréhension de la qualité de la donnée utilisée est primordial. La qualité d'une donnée est contextuelle, elle dépend des besoins d'un projet. Pour évaluer cette qualité, une étape préliminaire est nécessaire pour définir le projet dans lequel les données sont utilisées.

4.1 Étapes préliminaires - Conditions de départ

Évaluer la situation initiale est une étape primordiale pour toute entreprise qui veut se lancer dans un projet de passage au numérique. Cette étape permet de voir la réalité de l'entreprise en face. Quels sont les points forts et les lacunes de l'entreprise ? Cette connaissance permet d'établir un plan d'action qui tentera de couvrir plusieurs enjeux auxquels l'entreprise est confrontée et de lister les projets à mettre en avant.

Plusieurs méthodes sont déjà présentées dans la littérature. L'étude ne vise pas à créer ou modifier l'une de ces méthodes. Le mémoire se concentre sur la partie de définition d'une

qualité de la donnée récupérée. Cependant l'évaluation de la situation initiale reste une étape primordiale. Elle est le point de départ d'un ensemble de projets. Ici une méthode est choisie comme méthode de référence, car elle est en adéquation avec la philosophie du mémoire et se concentre sur l'aspect de la donnée. Il reste à la discrétion de chaque gestionnaire de projet de choisir la méthode qui correspond le plus à ses besoins.

Ce mémoire se base sur le livre "Executing data quality projects" [McGilvray, 2021] comme canevas de départ. Ce livre détaille comment planifier un projet de transformation numérique en 10 étapes. Cependant, il traite peu de la problématique de la qualité des données. Ce mémoire veut compléter ces étapes en ajoutant une méthodologie pour définir les besoins en qualité des données et évaluer la qualité des données récupérées. Pour évaluer la situation initiale de l'entreprise, il est donc logique de partir avec une méthodologie présente dans le livre : la méthode POMAD [McGilvray, 2021]. Cette méthode consiste à donner une vue d'ensemble sur l'information disponible pour comprendre une problématique particulière de l'entreprise. Pour se faire, le tableau 2.3 qui donne une compréhension exhaustive du cycle de vie d'une information est rempli. Cette compréhension est atteinte en posant les questions "Quoi ?" "Comment ?" et "Qui ?" à chaque étape du cycle de vie de la donnée. Ces étapes sont : Planifier(P), Obtenir(O), Stocker et partager(S), Maintenir(M), Appliquer(A), Effacer(E). Une fois ce tableau rempli, il est possible de voir où les lacunes en information se situent.

Une fois la compréhension de la situation initiale établie, des projets pour améliorer la capacité de l'entreprise à obtenir des informations peuvent être menés. Comme vue précédemment, l'industrie 4.0 rassemble plusieurs leviers technologiques qui permettent la transition vers le numérique. À cette étape, une entreprise est invitée à définir un plan d'action pour palier aux lacunes en information détectées plus haut.

Dans la section 2.4.2 de la revue de littérature, deux méthodes de définition de plan d'action pour la transformation numérique sont présentées. Pour les besoins du mémoire, une méthode sera choisie comme méthode de référence. La méthode est tirée du livre "Executing data quality projects" [McGilvray, 2021]. Cette méthode a été présentée durant la revue de littérature. Avec la connaissance des lacunes en information de l'entreprise définie grâce au POSMAD, un deuxième tableau de besoin en donnée est construit grâce à la méthode choisie. Dans ce tableau un lien entre l'information désirée et les données nécessaires pour l'obtenir est établi. Le tableau 2.2 nous permet d'établir une liste de projets pour obtenir des données permettant d'acquérir l'information voulue.

Le remplissage du tableau 2.2 se fait par itération pour réussir à affiner la compréhension du problème. En reprenant une partie du tableau POSMAD on définit l'information voulue. Cette information réfère à un besoin d'entreprise qui est aussi inscrit. Les problématiques

liées aux données nécessaires pour répondre à ce besoin commencent à apparaître. Plus la définition du besoin de l'entreprise et des problématiques liées à la donnée est précise, plus des projets d'amélioration peuvent être définis avec précision. La dernière colonne du tableau 2.2 liste les objectifs des projets à mener. C'est sur cette liste des objectifs de projet que la méthodologie développée se base pour identifier le besoin en qualité des données. Le mémoire se focalise sur les projets visant à récupérer des données du plancher de production.

4.2 Étape 1- Définition des attentes qualité

Comme vue précédemment, l'étape préliminaire permet de définir les projets de récupération de données à étudier. Ces projets permettent d'extraire des données du plancher de production. Ces données extraites sont ensuite valorisées pour obtenir des informations sur l'entreprise. Cela permet une meilleure prise de décision au sein de l'entreprise.

Un projet de récupération des données sur la chaîne de production vient avec un nombre de décisions à prendre. Pour un spécialiste, il est simple d'établir ses besoins en termes d'instruments de collectes de données. Mais dans une PME ces profils d'expert en transformation numérique sont rares. C'est pourquoi il est important pour une PME d'être outillée afin de mener à bien ces projets de transformation numérique. En effet, savoir comment les informations circulent dans une entreprise est important. Pour prendre des décisions, il est intéressant de savoir sur quelles données sont basées les informations utilisées. En s'inspirant de la liste des critères de qualité établie dans la section 2.3.1, une check liste de questions à poser est développée dans ce mémoire afin d'aider les PME à comprendre leur besoins en données. Ces questions sont présentées sous la forme d'un tableau 4.1. Ce tableau est pensé pour une utilisation facile lors du lancement d'un projet. Les conclusions tirées grâce aux réponses collectées peuvent être directement inscrites dans la dernière colonne "décision prise". Ainsi cet outil est développé comme un outil d'aide à la décision.

Le tableau compile les questions à se poser dans sept catégories d'étude afin de faciliter une prise de décision. La précision acceptable cherche à définir le standard de précision et fidélité que la donnée doit avoir pour réussir le projet d'analyse de données. Sous cette catégorie il faut en premier lieu définir l'importance d'une donnée à récupérer pour le projet. Cette importance est définie par l'impact de la donnée sur le résultat de l'analyse. Plus l'impact est grand plus la précision de la donnée et sa capacité à décrire la réalité est un enjeu sensible. La précision fait référence aux tolérances acceptables. La fidélité rend compte de la capacité d'une donnée à décrire la réalité physique.

La gamme des instruments est une catégorie qui aide à définir la technologie qui sera utilisée

TABLEAU 4.1 Tableau de prise de décision

Décision	Où trouver l'information	Décision prise
Précision acceptable	➤ Importance de la donnée à récupérer ➤ Précision du capteur ○ Qu'elles sont les tolérances acceptables pour l'étude? ○ La précision du capteur correspond-elle aux besoins? ○ Y a-t-il de la maintenance à faire pour garder un niveau de précision? ➤ Fidélité avec la réalité ○ Le capteur représente-t-il fidèlement la réalité? ○ Peut-on calculer l'écart entre la réalité et la donnée mesurée? ○ L'erreur introduite est-elle incluse dans l'intervalle d'acceptation?	
Gamme des instruments de collecte	➤ Le budget défini ➤ Grade à respecter ▪ Le milieu industriel impose-t-il des normes à respecter? ▪ L'installation est-elle aux normes? ➤ Retour sur la précision acceptable	
Formation du personnel	➤ Automatisation ▪ La récupération de données est-elle automatique? ▪ Si non, combien de personnes sont incluses dans ce processus? ▪ Les étapes sont-elles complexes? ➤ Base de données ▪ Les données sont-elles compatibles avec la base de données? ▪ Sont-elles facilement accessibles? ▪ La récupération a-t-elle besoin de manutention ?	
Maintenance des instruments de collecte	➤ Pannes possibles ▪ Peut-on énumérer les risques de pannes? ▪ Quelle est la fréquence de pannes? ▪ La panne est-elle détectable? ▪ Combien de temps pour réparer la panne?	
Fréquence	➤ Fréquence de récupération ▪ Quelle est la fréquence maximale permise? ▪ Quelle est la fréquence nécessaire pour l'étude? ▪ La fréquence est-elle modifiable?	
Méthode d'installation du système de collecte	➤ Risques de contamination ▪ L'installation respecte-t-elle l'intégrité du produit? ▪ Quels sont les risques de contamination? ➤ Quantité d'information ▪ Peut-on obtenir plusieurs informations avec la même installation? ▪ Cela complique-t-il le projet?	
Règles de sauvegarde des données	➤ Fréquence de récupération ▪ Quelle est la fréquence maximale permise? ▪ Quelle est la fréquence nécessaire pour l'étude? ▪ La fréquence est-elle modifiable? ➤ Base de données ▪ Les données sont-elles compatibles avec la base de données? ▪ Sont-elles facilement accessibles? ▪ La récupération a-t-elle besoin de manutention ?	

pour la récupération des données. Les outils sur le marché ont différents standards de précision et choisir une gamme demande au préalable d'avoir rempli la catégorie "Précision acceptable". La gamme des instruments va influencer grandement le budget du projet. Il faut donc définir un budget durant cette étape et essayer de s'y tenir.

La formation du personnel indique si la récupération des données nécessite du personnel compétent ou pas. Si la récupération est automatique, aucune main d'oeuvre n'est nécessaire pour la saisie. Si du personnel doit interagir avec le système de récupération, il faut estimer si une formation sera nécessaire et à quel point cela peut être une source d'erreur. Cette

catégorie étudie aussi la facilité d'accéder aux données dans la base de données.

Dans la catégorie maintenance, il est question des pannes possibles. Un système de récupération a besoin de maintenance préventive pour éviter une perte de données due à une panne. Il faut estimer la fréquence des pannes, si elles sont facilement détectables ainsi que la difficulté de les réparer. L'environnement dans lequel le système de récupération se trouve va influencer la fréquence des pannes. Un milieu avec des vibrations ou de la poussière demande plus d'effort de maintenance pour éviter les écarts de mesure.

La fréquence de récupération de la donnée doit être définie de telle sorte à ne pas manquer des événements intéressants à étudier. Le système de récupération aura des limitations sur la rapidité à prendre des mesures. Il faut vérifier si la vitesse de prise de mesure est suffisante pour obtenir une information complète. Certaines mesures ont besoin d'être prises en continu d'autre demandent à être prise à l'occurrence d'un événement. Il faut vérifier si le système de récupération permet d'avoir le mode de récupération désiré.

La catégorie méthode d'installation permet de savoir si des précautions particulières dues au milieu d'installation sont à prendre en compte. Par exemple les industries alimentaires ont des normes précises à respecter pour éviter les contaminations des produits alimentaires. La méthode d'installation pose aussi la question de la versatilité du système de récupération. Il est intéressant de savoir si plusieurs données peuvent être récupérées par la même installation.

Les règles de sauvegarde reprennent les mêmes questions que pour la fréquence et la formation du personnel. Mais ce qui est recherché est de définir comment programmer la récupération. Quand le système le permet, il faut programmer la récupération pour avoir la fréquence désirée et une méthode d'enregistrement des données qui facilitent l'utilisation des données.

Une personne qui veut lancer ou comprendre les besoins d'un projet d'exploitations des données doit utiliser ce tableau. Son utilisateur prend conscience des besoins en qualité de données de son projet. Passer à travers ces questions une à une permet de ne rien oublier en termes d'attente en qualité des données. Il est possible que les questions posées ne s'adaptent pas à toutes les manufactures. D'autres contextes industriels peuvent amener d'autres questions spécifiques. Le modèle actuel doit être considéré comme un canevas de base et évoluer pour s'adapter aux besoins. Ce tableau fait le lien avec les caractéristiques et les dimensions à prendre en compte pour la qualité des données vues dans la revue de littérature. Le tableau 4.2 présente quelles caractéristiques ont été vérifiées en posant les questions d'aide à la décision. Les questions du tableau 4.1 ont été pensées de telles sortes à passer en revue les caractéristiques de qualité des données. Par exemple, les questions de la catégorie "Précision acceptable" viennent vérifier :

1. La précision intrinsèque du système de récupération ;
2. La crédibilité aux données en identifiant leur justesse ;
3. La pertinence d'utiliser ces données en insistant sur leurs capacités à représenter le réel ;
4. La capacité à représenter le réel ;
5. La réputation de la source des données en identifiant le type de la source et la précision qu'elle permet d'obtenir ;
6. La capacité de comprendre comment les données informent sur le réel ;
7. La valeur ajoutée des données récupérées en s'assurant qu'elles aident à la compréhension du contexte de production.

Dans le tableau 4.2 l'axe des abscisses retranscrit la première colonne du tableau 4.1. L'axe des abscisses fait allusion à des catégories de questions à poser. l'axe des ordonnées liste les caractéristiques de qualité. La colonne méthode d'installation ne traite aucun des caractéristiques de qualité. Cette liste de questions sert à traiter le contexte industriel et les normes à respecter. Cette liste de question établie dans le tableau 4.1 s'inspire de la dimension contextuelle de la qualité des données. La traçabilité n'est pas traitée par les questions du tableau, 4.1 mais elle sera traitée dans l'étape 3 par la création de méta-données. La sécurité est hors du champ d'études de ce mémoire.

Cette étape de la méthodologie développée dans ce mémoire est importante. En effet, elle répond à la première composante de la problématique qui est de définir les besoins en qualité de la donnée à récupérer. Le résultat de cette étape est une liste de besoin en qualité de la donnée. Grâce à cette liste, il est possible d'entamer l'évaluation des données récupérées. La suite du travail consiste à créer un moyen d'évaluer la qualité de la donnée et de la comparer aux besoins en qualité définis ci-dessus.

4.3 Étape 2- Développement de KPIs

Une fois les besoins en qualité définis, un moyen pour valider qu'ils sont atteints est nécessaire. Pour ce faire des KPIs sont mis en place pour évaluer la qualité de la donnée récupérée. Ces KPI servent à identifier les performances de nos données récupérées et indiquer si elles sont toujours aux normes identifiées ci-dessus. Leur mise en place est un exercice intéressant pour l'entreprise, car elle mène à comprendre les enjeux d'amélioration continue de la qualité. En effet, si on décide de suivre la performance de certaines données c'est pour pouvoir à terme les améliorer pour continuer à correspondre aux besoins en qualité qui seront amenés à évoluer.

TABLEAU 4.2 Décision et critères de qualité

	Précision acceptable	Gamme des instruments de collecte	Formation du personnels	Maintenance des instruments de collecte	Fréquence	Méthode d'installation du système de collecte	Règles de sauvegarde
Accessibilité							x
Précision	x	x					
Quantité de données					x		x
Disponibilité							x
Crédibilité	x	x	x	x			
Données complètes				x	x		x
Consistance		x		x			
Pertinence	x						
Traçabilité							
Variance		x		x			
Représentation de la réalité	x	x					
Facilité d'utilisation			x				x
Objectivité			x				
Réputation	x	x					
Sécurité							
Données compréhensible	x						x
Valeur ajoutée	x						
Erreurs		x		x			
Marge d'erreur		x		x			

Une fois les besoins en qualité établis, il faut les transformer en métriques. Comme vu dans l'état de l'art, il n'existe pas une banque de KPIs qui décrivent les données industrielles. Des travaux ont été menés pour les données utilisées dans le domaine médical [Stausberg et al., 2019]. Ces KPIs permettent de mesurer la qualité d'une donnée utilisée dans l'organisation des hôpitaux. Ces données ayant un impact majeur sur le service rendu aux patients sont importantes à implémenter. Une mauvaise gestion de ces données ou une incompréhension de leurs significations peut avoir un impact grave sur la vie des patients, d'où l'importance d'utiliser des KPIs pour évaluer la qualité de ces données. Il devrait en être de même dans le cadre industriel. Ce travail propose de développer des KPIs pour évaluer la donnée industrielle. Ces KPIs qui seront mis en place sont très généraux et peuvent s'appliquer à la majorité des cas de récupération de données. Pour réussir à créer des KPIs, une méthode est développée ci-dessous. Il est à noter que des installations particulières peuvent nécessiter la création de nouveaux KPIs.

4.3.1 Méthode de développement de KPIs

Les installations techniques dans une PME sont souvent peu équipées en technologie de communication [Zhou et al., 2020]. Cela se traduit par une récupération de données manuelle ou inexistante. Ce manque en technologie du numérique empêche un bon suivi de la production. Quand une entreprise lance un projet d'analyse de données, s'équiper en machineries de nouvelles générations est très coûteux. Une entreprise à la possibilité de modifier ses équipements au moindre coût et d'installer ses propres capteurs. Cette démarche est d'autant plus intéressante qu'il existe plusieurs offres de capteurs industriels disponibles.

Certains projets de recherche et développement mènent les entreprises à avoir un besoin très spécifique en données. Cette situation peut mener à l'installation d'un système de récupération de données unique. Qu'il soit artisanal ou produit en partenariat avec un spécialiste, le système aura des attributs particuliers. Les données récupérées par ce système seront elles aussi uniques et particulières. La qualité de ces données particulières doit être maîtrisée d'où l'intérêt de créer des KPIs spécifiques.

Une méthode est développée pour créer de nouveau KPIs. Elle s'inspire de la méthode de développement des KPIs vu à la section 2.3.3 par [Zheng, 2018]. La méthode de [Zheng, 2018] insiste sur l'importance de définir les informations prioritaires avant de créer des KPIs. Dans ce mémoire, ceci est effectué lors de l'étape 1 "Définition des attentes qualité". [Zheng, 2018] utilise le guide PMBoK pour créer des KPIs. Vu que nos projets visent la qualité des données du plancher de production, les dix aires de connaissances du PMBoK ne sont pas utiles. C'est pourquoi dans ce travail la création des KPIs ne prend pas en compte le PMBoK.

Le développement des KPIs est décrit ci-dessous.

4.3.1.1 Identifier le problème

Avec la liste des besoins en qualité des données obtenues à l'étape 1, il est possible d'anticiper les défis qui seront rencontrés pour atteindre ces besoins. Chaque système de récupération de données rencontrera des défis pour obtenir la qualité des données voulue. Chaque défi se compose d'un ou plusieurs problèmes qui empêchent l'obtention d'une donnée de qualité. L'identification de ces problèmes est une étape primordiale de l'évaluation de la qualité des données à récupérer. Les KPIs qui seront développés vont juger si les problèmes de qualité ont été résolus. Un KPI a besoin de se rattacher à un élément factuel et mesurable pour être opérationnel. Il faut donc décortiquer le problème jusqu'à arriver à un élément mesurable. Il est conseillé d'utiliser la méthode des "5 pourquoi" pour trouver la cause racine du problème de qualité. La méthode des "5 pourquoi" est une méthode développée par Taiichi Ohno, fréquemment utilisée dans le cadre des projets d'amélioration de la qualité. Elle consiste de partir d'un problème général et de poser la question pourquoi 5 fois afin de découvrir les causes profondes du problème. Ci-dessous sont listés des problèmes potentiels :

- Le système est mal utilisé par l'humain ;
- Le système tombe souvent en panne ;
- Le système a besoin d'être calibré souvent ;
- Le système a un mécanisme qui engendre un manque de précision.

Pour illustrer cette méthode de développement de KPI, le problème lié au manque de précision et la création des KPIs "Exactitude et précision" seront pris en exemple. Lors d'une prise de mesure, l'exactitude et la précision sont deux qualités importantes. L'exactitude est la capacité du système de récupération d'obtenir une information correcte et proche de la réalité. La précision représente la capacité d'un système à obtenir de valeurs avec peu de variances. Dans le langage commun, ces deux métriques se confondent facilement. Créer un KPI autour de ces définitions est intuitif. Le souhait est toujours d'avoir une information la plus proche de la réalité le plus souvent. Pour comprendre la méthode présentée à la section précédente, elle sera appliquée ici.

Premièrement on identifie le problème. Pour réaliser cette étape, il est conseillé de dérouler un fil de questionnement qui décortique le problème.

1. Pour quelle situation la précision est-elle nécessaire ?
 - (a) **Mesurer une valeur** : est-ce un problème de précision ou d'exactitude ?
 - i. **Problème de précision du capteur** : la tolérance du capteur est-elle acceptable ?

- ii. **Problème d'exactitude** : la mesure représente-t-elle avec fidélité la réalité ?
- (b) **Comptabiliser des événements** : le nombre d'événements mesurés correspond-il aux nombres d'événements produits ?

L'enquête reste très générale, elle doit être approfondie pour chaque cas particulier. Ceci est une ébauche pour expliquer les étapes à suivre.

4.3.1.2 Trouver une méthode de mesure

Normalement une entreprise se base sur des données qu'elle estime fiables pour alimenter des KPIs. Ici, le but est de créer des KPIs qui jugeront la qualité des données récupérées. Il faut être astucieux pour trouver une source de données qui permettront de juger ces données récupérées. Pour cela il faut trouver une méthode de mesure qui permet de vérifier que la cause du problème de qualité identifiée grâce à la méthode des "5 pourquoi" est résolue. Cela peut être une indication donnée par le fabricant d'un capteur. Aussi, dans le cas d'un système de récupération qui nécessite l'intervention d'un ouvrier, la solution pourrait être d'avoir une période d'essai pour déterminer les performances du système. Les solutions sont nombreuses, mais une réflexion rigoureuse doit être menée pour établir la méthode de mesure. Elle doit être répliquable pour évaluer l'amélioration. Elle doit aussi être objective.

Pour illustrer cette étape de la méthode, l'exemple du KPI "Exactitude et précision" est repris. Après avoir défini les causes racines des problèmes de qualité, il faut définir une méthode de mesure. Ici, le but est de définir comment mesurer les KPI et quelle métrique utiliser. Précédemment, 3 cas de figure ont été identifiés. Chacun aura une méthode d'évaluation propre.

- **Précision** : La tolérance intrinsèque de l'outil de mesure utilisé est un bon indicateur de la précision de la donnée ;
- **Exactitude** : Une étude du contexte d'installation et de la donnée qui peut être perdue ;
- **Comptabiliser des événements** : Définir le nombre d'événements non détectés sur une période. Mettre en place un temps d'essai.

Il faut comprendre que chaque situation est unique. Mettre en place des méthodes de mesure doit être fait en prenant en compte la situation. Cette étude doit être menée au cas par cas.

4.3.1.3 Choisir le format

Une fois le problème identifié et la méthode de mesure choisie, il faut penser au format que prendra le KPI. L'un des critères pour voir un bon KPI est de comprendre très vite le message

qu'il véhicule. Il est donc important de trouver un format explicite et compréhensible. Des exemples de format des métriques sont :

1. Pourcentage
2. Niveau de performance
3. Booléen

Une fois cela fait, le KPI est prêt à être utilisé. Les sous-sections suivantes donneront des exemples de KPI. Les KPI présentés sont assez généraux et considérés comme applicables à plusieurs situations. Le KPI "Exactitude et précision" sera utilisé pour illustrer cette démarche.

Une fois la façon de mesurer est établie, il faut définir le format que va prendre le KPI. En reprenant les trois cas de figure :

- **Précision** : Rapport entre la tolérance du capteur et la plage des valeurs mesurables ;
- **Exactitude** : Pourcentage de perte sur les informations du réel ;
- **Comptabiliser les événements** : Pourcentage des événements ratés.

La figure 4.1 résume la démarche. Il peut être observé trois niveaux de précision. Ces niveaux reprennent l'étape d'identification du problème. Le premier cercle représente le problème initial. Les deux cercles "mesurer une valeur" et "comptabiliser des événements" sont des niveaux de précision supérieurs. Les carrés blancs identifient les valeurs qui sont à évaluer, soit de gauche à droite la précision, l'exactitude et la comptabilisation des événements. Les carrés bleus sont les formats de présentation de ces valeurs.

4.3.2 Autre exemple de KPI développé

Cette section décrit des exemples de KPI développés grâce à la méthode proposée. Ces exemples sont moins détaillés que "Exactitude et précision", mais sont utilisés par la suite dans l'étude de cas.

4.3.2.1 KPI-Fréquence

La fréquence fait référence aux rythmes auxquels les informations sont récupérées et enregistrées. En effet, généralement les mesures prises sont évolutives. Chacune change à une vitesse particulière. Ces changements sont souvent la clé de l'analyse que l'on veut effectuer. Ces changements sont des événements qui peuvent influencer la production. Un changement de masse, un réglage qui change, une vitesse qui est modifiée sont des événements qui peuvent être utilisés pour faire des corrélations. L'objectif d'un capteur est de pouvoir établir avec précision le moment d'occurrence de ces événements.

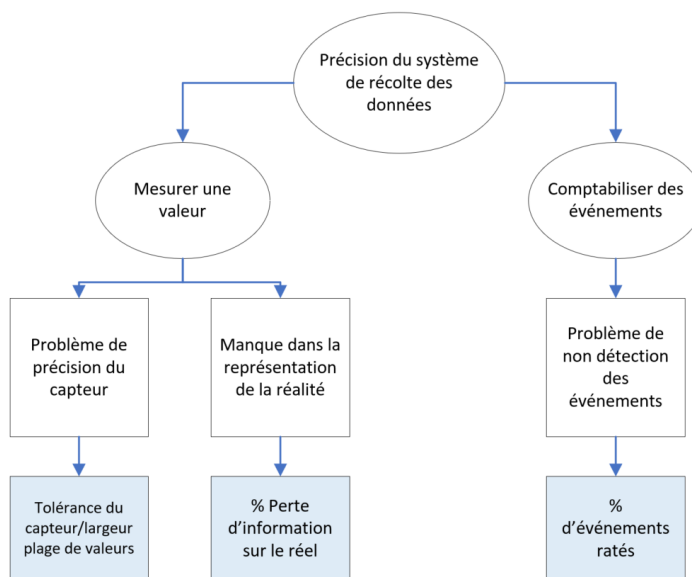


FIGURE 4.1 KPI- Exactitude et précision

Pour réussir à faire cela, il faut pouvoir estimer la fréquence d'occurrence de l'événement à mesurer. Le problème qui peut survenir serait d'avoir une fréquence de récupération trop basse comparée à la fréquence d'occurrence de l'événement. Si les événements se produisent à une fréquence élevée, il faut que la récupération des données puisse suivre. Le format du KPI peut être un booléen qui annonce si la fréquence de récupération est suffisante ou pas. Le tableau 4.3 illustre le KPI.

TABLEAU 4.3 KPI-Fréquence

Fréquence	
F: Fréquence de l'événement	
f: Fréquence de capture	
OUI	$F < f$
NON	$F > f$

Deux variables sont définies dans le tableau 4.3. "F" la fréquence de l'événement à observer, "f" la fréquence de capture de données. Avec un jeu de couleur, on peut mettre en évidence si la fréquence est adéquate. Cela permet de réagir et de voir quelles mesures peuvent être prises. Bien sûr, il peut y avoir des marges de tolérance, néanmoins c'est une question importante à se poser et à suivre.

4.3.3 KPI-Manutention

Une PME qui décide d'étendre ses capacités de récupération de données peut être confrontée à des questions technologiques. En effet l'automatisation des processus de récupération est disponible. Les coûts d'automatisation par contre peuvent être élevés. Prendre une décision sur le coût d'installation d'un projet de récupération de données est important. Cet investissement peut changer la façon de faire des ouvriers. Plus on veut récupérer de données, plus on multiplie les processus au sein de l'entreprise. Plus ces processus sont automatisés, moins on accapare des ressources humaines. Dans une situation de pénurie de main d'oeuvre, l'automatisation est un enjeu capital.

De plus, un projet de récupération est évolutif. Le processus peut être revu selon l'importance de la donnée que l'entreprise récupère. Si une période d'essai est nécessaire, avoir une solution temporaire qui demande de la manutention peut être envisagée. Cela permet d'évaluer l'importance d'une donnée avant d'investir une somme importante dans de l'équipement. Mais pour qu'une solution temporaire ne se transforme pas en une solution permanente, il est important de mettre en avant ces désavantages.

Le KPI qui décrit le degré de manutention nécessaire permet de faire ce suivi. En plus du temps de manutention, l'erreur humaine est une cause de manque de qualité dans la donnée. Un système de récupération de données difficile à opérer et qui nécessite beaucoup d'étapes peut être une cause de baisse de qualité. La figure 4.4 décrit une échelle d'appréciation du niveau de manipulation manuelle. Plus le score obtenu est élevé, plus le processus de récupération est automatique et apprécié.

TABLEAU 4.4 KPI-Manutention

Manutention	
Note /4	Description
4	Automatique
3	Intéraction minime
2	Nécessite une formation
1	Complètement manuelle

4.3.4 KPI-Évolution

Pour une entreprise qui se lance dans le passage au numérique, il peut être difficile d'investir de grandes sommes dès le début. Le sondage mené par le gouvernement québécois [Fortin, 2019] montre qu'une grande partie des entreprises ne croit pas encore à l'impact de la numérisation. Une approche itérative peut être adoptée pour réduire l'appréhension de départ. Pour ce faire des étapes d'investissement sont mises en place. Cela permet de vérifier l'apport qu'une donnée récupérée peut avoir et décider de continuer ou non à l'extraire.

Pour pouvoir faire ce suivi, il est intéressant d'estimer la facilité avec laquelle une installation peut être améliorée.

Dans certains cas, une installation de capteur nécessite des modifications irréversibles dans la chaîne de production. Ce genre d'installation est difficile à modifier, il faut donc être sûr de son choix.

Dans d'autres cas, l'intervention d'un spécialiste peut être requise pour faire évoluer une installation. Que ce soit d'un point de vue informatique ou physique, certaines installations peuvent évoluer avec des solutions simples.

D'autres sont des ajouts faciles et nécessitent peu de modifications. Modifier ces installations est accessible. Dans d'autres cas, une même entreprise offre plusieurs gammes d'un même produit, voire des extensions qui peuvent être installées. Dans ce cas les évolutions possibles sont faciles à obtenir.

Un indicateur est introduit pour évaluer cette facilité à faire évoluer un système de récupération. Il est imaginé comme une échelle d'appréciation. Plus le score est élevé, plus il est facile de faire évoluer le système. La figure 4.5 illustre ce KPI. Cette facilité d'évolution montre aussi une facilité d'adapter la qualité de l'information récupérée. En effet plus une entreprise devient avancée dans la numérisation, plus ces attentes en termes de qualité de la donnée vont être élevées.

4.4 Étape 3- Évaluation de la qualité de la donnée

Cette partie est la conclusion du travail effectué en amont. En effet, la problématique du mémoire est de définir une façon d'évaluer si la donnée récupérée a les standards de qualité nécessaire pour un projet d'analyse de données. Pour se faire, les sections précédentes décrivent une méthode pour arriver à cette évaluation. Dans le but de maintenir une compréhension des données récoltées sur la ligne de production, il faut mettre en place un document qui décrit la récupération. Grâce à ce document, l'évaluation faite sera enregistrée et consul-

TABLEAU 4.5 KPI-Évolution de l'installation

Capacité évolutive	
Note /3	Description
3	Facile: achat d'une extension
2	Moyen: programmation ou bricolage impliqué
1	Difficile: n'évolue pas ou peu, gros investissement

table facilement. De plus d'autres informations liées à la récupération des données pourront être enregistrées.

Dans un environnement où les données sont récupérées de façon standard, comme pour les données météorologiques par exemple, comprendre la source des données est facile. Dans le cas où la récupération des données n'est pas standardisée, c'est plus compliqué. Les projets d'industrie 4.0 se font à l'interne. Ces projets visent à faire évoluer le plancher de production et à numériser au mieux les processus. Il y a donc plusieurs sources différentes de données qui ne font pas partie des équipements de base. Il y a le besoin pour une industrie d'expliquer les processus de récupération de données.

La disponibilité de cette information aide les utilisateurs à comprendre les données à leur disposition. Cette compréhension permet principalement deux choses, la première est d'analyser la donnée en sachant sa provenance, la deuxième est d'avoir une meilleure vue de comment améliorer le système ou le modifier en cas de besoin.

Pour fournir cette information, le concept de méta-données est exploré. Selon l'Office québécois de la langue française, une métadonnée est une "Donnée qui renseigne sur la nature de certaines autres données dans le but d'en faciliter la compréhension et la gestion." [québécois de la langue française, 2020]. Une métadonnée est implémentée pour suivre les données récupérées. Cette métadonnée décrit le processus de récupération et donne du contexte.

En s'inspirant des problématiques vues plus haut on dégage certaines informations qui peuvent suivre la donnée. Il faut aussi comprendre que les processus de récupération peuvent varier de tel sorte que la description doit s'adapter. Cependant, des informations communes sont utiles à la compréhension des données. Le tableau 4.6 présente un gabarit de fiche descriptive incluant des informations générales. L'étude de cas explore comment il peut être adapté à des utilisations plus spécifiques. Le document prendra la forme d'un "read me".

TABLEAU 4.6 Gabarit des méta-données

Titre		
Description	Utilisation et sens de la donnée:	
Dates	Date de la première mise en place:	Date de la dernière évaluation
Personnes	Nom de la personne responsable de l'installation:	Nom de la personne responsable de la maintenance:
	Listes de personnes ayant accès à/utilisant la donnée:	
Technologie	Processus de récupération:	Technologie:
	Schéma ou dessin aidant à la compréhension (si nécessaire)	
Évaluation	Liste des KPI utilisés:	Commentaires relatifs au KPI

Pour décrire au mieux la donnée, on y inclut un ensemble d'information comme décrit dans le tableau 4.6. Ce document doit être accessible à toute personne qui veut utiliser les données. De cette façon, une économie de temps est réalisée en facilitant la compréhension des données. Bien évidemment, la figure est un prototype qui est évolutif et des informations supplémentaires peuvent y être ajoutées.

Ce prototype se divise en cinq sous partie :

Description : Cette partie est réservée pour décrire l'idée derrière la récupération des données. Quelles étaient l'intention de départ et la signification de ces données ;

Dates : Ici on indique le moment où le système de récupération a été installé et la date de la dernière mise à jour du système. Cela permet de voir si le système est à jour et s'il est encore assez récent ;

Personnes : Une liste des personnes d'intérêt est gardée pour savoir qui contacter en cas de besoin ;

Technologie : Cette partie décrit la technologie utilisée ainsi que le processus en place pour récupérer la donnée. Cette partie décrit le contexte dans lequel la donnée est récupérée. Des précisions sur la nature informatique des données peuvent être incluses ;

Évaluation : Les indicateurs de performance de la donnée sont listés et commentés. L'évaluation de la donnée est enregistrée ici en se servant des KPIs.

Une fois cette partie faite toute personne ayant besoin de la données peut s'y référer. Ce gabarit de description de la donnée permet aussi de faciliter les boucles d'itération pour améliorer la qualité de l'information. En effet, les KPIs permettent d'évaluer la donnée. Cette évaluation sera stockée dans ce gabarit qui permet d'être une référence pour décider si une amélioration du processus de récupération est nécessaire.

Pour résumer, l'étape préliminaire sert à comprendre les besoins de l'entreprise et établir un plan d'action. Par la suite, pour définir les besoins spécifiques pour chaque projet, un

gabarit de questions à poser est présenté. Ce gabarit sert à guider la réflexion sur les besoins qualité des données récupérées. Des KPI sont développés pour évaluer la qualité des données récupérées. Enfin, il faut enregistrer les informations obtenues sur la qualité des données récupérées, un gabarit de méta-données est présenté. Ce gabarit sert à maintenir un suivi de la qualité des données, mais aussi aide à la prise en main des données par un nouvel utilisateur.

Le chapitre suivant présente l'étude de cas réalisée au cours de cette maîtrise. Le partenaire industriel est présenté ainsi que le projet de récupération de données. L'étude de cas illustre comment la méthodologie d'évaluation de la qualité de données récupérées a été utilisée pour récupérer des données efficacement.

CHAPITRE 5 ÉTUDES DE CAS

Dans ce chapitre, la méthodologie développée précédemment va être utilisée en vue d'assurer la qualité des données récupérées sur le plancher de production. La maîtrise a été effectuée en partenariat avec un industriel. Le travail sur le plancher de production a pu inspirer les solutions présentées dans ce mémoire. L'entreprise dans laquelle le partenariat a été mené est en pleine expansion de ses capacités numériques. Ces changements se font à grande vitesse et en collaboration avec l'université. C'est donc un terrain propice pour développer des outils et des méthodes qui aident à suivre les projets de numérisation.

Durant la maîtrise, trois projets de transformation numériques ont été retenus pour ce mémoire. Ces trois travaux consistaient à de la récupération de données de qualité par divers moyens dans le but d'avoir un meilleur suivi sur la chaîne de production. Chacun de ces travaux permet de tester une partie complémentaire de la théorie développée. Il est à noter que l'un des projets a nécessité le développement d'un nouveau système de récupération de données.

Le partenaire industriel est présenté, ensuite le projet global dans lequel s'insère ce travail est présenté. Pour finir, les trois projets menés au sein de l'entreprise seront utilisés pour illustrer la méthodologie développée au chapitre 4. L'étape préliminaire de la méthodologie est menée au préalable et donc ne sera pas présentée au cours de cette étude de cas.

5.1 Contexte industriel

5.1.1 Présentation du partenaire

Le projet de recherche s'inscrit dans une volonté d'intégrer des connaissances académiques sur les technologies reliées à l'industrie 4.0 directement dans l'industrie québécoise. Pour ce faire, un partenariat avec un industriel qui cherche à investir dans les technologies 4.0 a été entrepris pour développer et intégrer des outils de gestion de données. Dans ce contexte, le projet de qualité des données récupéré est né, comme étape préalable à tout investissement dans ce domaine.

L'entreprise partenaire est une PME qui travaille dans le domaine agroalimentaire, plus précisément dans la production de farine biologique à partir de grain de blé canadien. La production d'aliment biologique obéit à certains principes de production écoresponsable [Québec, 2021] :

- prohibition d'utiliser des engrais chimiques ;

- prohibition d'utiliser des pesticides ;
- prohibition des Organisme Génétiquement Modifié (OGM) ;
- prohibition des additifs durant la transformation alimentaire.

Par ailleurs, le partenaire est dans un processus de modernisation en investissant dans les technologies 4.0. Il est donc indéniable que ces choix de production stricte et de modernisation s'accompagnent de certains défis auxquels le partenaire est confronté. Ces enjeux sont liés à la production elle-même mais aussi au fait que le partenaire est une PME aux moyens limités. En effet, le processus de transformation doit être maîtrisé pour déjouer le peu de marge de manoeuvre. Comme vu précédemment, les technologies reliées à l'industrie 4.0 sont peu adaptées aux entreprises de petite taille et demandent le déploiement de moyens financiers volumineux et une main d'œuvre spécialisée pas toujours disponibles en PME.

La production de farine est un processus continu, et entièrement automatisée du stockage des matières premières à l'emballage. Une partie de l'emballage ainsi que l'entreposage de la farine restent encore manuelles. L'entreprise possède un système informatique de suivi de la production. Certains capteurs sont présents sur la chaîne de production pour faire le suivi. Cependant, les données récupérées sont peu nombreuses et certains systèmes de récupération des données sont défaillants. Cela mène à une difficulté pour effectuer des projets d'analyse de données et plus particulièrement l'entreprise n'a pas de modèles de prédiction pour sa production. L'efficacité de la production dépend essentiellement du savoir-faire du chef meunier. Cela fait de lui une ressource critique indispensable à la production et une source d'information importante.

5.1.2 Projet d'amélioration de la farine

Dans ce contexte, où l'efficacité de la production dépend des connaissances d'une seule personne, l'entreprise veut entreprendre des projets pour numériser son processus. Cela permet une prise de décision assistée numériquement. En ayant ce système en place, les risques d'erreurs et surtout de perte en productivité si le chef meunier n'était pas présent diminuent. À cet effet, un projet de numérisation est en cours avec la participation de plusieurs étudiants aux cycles supérieurs. Deux projets étaient lancés :

1. Projet de prédiction de la qualité de la farine : mené par un doctorant, le projet vise à utiliser les données récoltées pour modéliser la qualité de la farine. Ceci est fait pour trouver les réglages qui maximisent la qualité de la farine.
2. Projet pour installer un CPS : mené par un étudiant à la maîtrise, ce projet a mis en place une feuille de route pour installer un CPS dans une PME.

Dans la continuité, le projet de définition de la qualité des données récupérée viendra com-

pléter ces deux projets. Au cours de ce projet, trois exemples de collecte de données ont été étudiés.

1. Capteurs de serrage des rouleaux de broyage du blé.
2. Capteurs de niveau pour les silos de blé.
3. Système d'enregistrement des palettes finies.

Le choix des projets présentés a été fait de manière à couvrir l'intégralité des étapes proposées dans la méthodologie détaillée dans le chapitre 4. Ces projets ont été menés pour appuyer le besoin de suivi du processus de production. Ainsi l'installation des capteurs au niveau de la chaîne de production permet d'avoir les données nécessaires qui serviront à la prédiction de la qualité de la farine.

Dans la suite de ce chapitre, le travail effectué dans l'entreprise en termes de récupération de données sera présenté. À cela s'ajoutera une analyse de la qualité des données récupérées. Ce chapitre est utilisé pour valider la méthodologie mise en place. Pour rappel, les étapes qui seront illustrées dans ce chapitre sont :

1. Définition des attentes qualité
2. Développement des KPI
3. Évaluation de la qualité de la donnée

Le schéma 3.1 de la méthodologie d'évaluation de la qualité des données récupérée sert de guide pour le reste de cette section.

5.2 Projet 1 : Mesure de l'écartement des rouleaux de broyage

Lors de la transformation du blé en farine, le grain passe par plusieurs étapes. Le schéma 5.1 illustre le processus de façon simplifiée. Les étapes de validation et test sont omises. Le schéma indique les grandes étapes de transformation du blé. La figure 5.1 contient huit cases

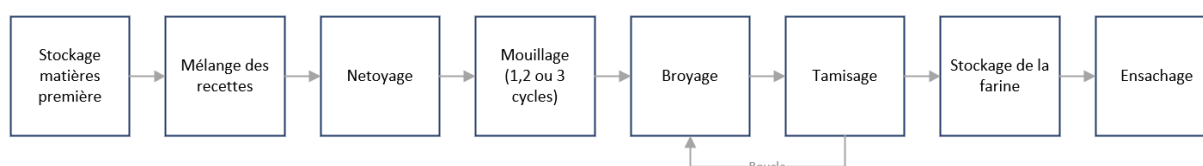


FIGURE 5.1 Processus simplifié

qui représentent les étapes de transformation du blé. Ce schéma couvre les étapes du stockage

de la matière première à l'ensachage de la farine. Une boucle est visible entre le tamisage et le broyage. Cette boucle se répète un nombre différent de fois selon le résultat final souhaité.

L'une des étapes qui défini le type de farine produite est le broyage. L'article [Alfin, 2007] explique l'importance du "break release" ou en français le pourcentage de broya obtenu. Le but est d'extraire un maximum d'albumen du grain de blé tout en tamisant le son. La figure 5.2 présente les différentes parties du grain de blé de façon simplifiée [Feillet, 2000]. La farine blanche est une farine qui est composée uniquement d'albumen. Pour séparer l'albumen du son, le blé passe par des étapes de broyages et de tamisages. La force avec laquelle le grain est broyé va influencer sur la taille des particules et ainsi sur la composition de la farine obtenue. Le "break release" est le calcul du pourcentage de matière qui passe à travers le tamis. Cette valeur a de l'importance pour les meuniers, elle leur donne un aperçu de la qualité de la farine qui sera obtenue.

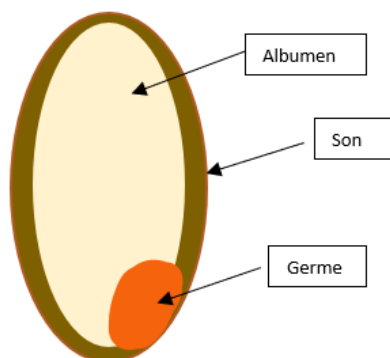


FIGURE 5.2 Anatomie du grain de blé

Chaque type de blé a des caractéristiques différentes. Le blé réagit différemment à la mouture selon ses caractéristiques et les conditions de broyage. Un bon meunier sait prendre en considération ces différents paramètres pour obtenir une farine de qualité. Le projet d'installer des capteurs pour le serrage des rouleaux de broyage permettra d'établir des liens entre le serrage des rouleaux et la qualité de la farine. Ces liens seront établis par analyse des données récoltées. Ils permettront au meunier d'ajuster son réglage de serrage par rapport au grain de blé qui sera utilisé.

Les rouleaux de broyage fonctionnent en compressant le grain de blé jusqu'à craquer la coquille de son pour extraire l'albumen. Les rouleaux de broyages consistent en deux cylindres métallique dont l'écartement est ajustable. Les grains de blé passent à travers les deux cylindres et se font broyer. Plus l'écartement entre les deux cylindre est petit plus le blé se fait

moudre finement. Pour avoir la mouture de la taille voulue, le blé passe successivement entre plusieurs paires de cylindres dont l'écartement devient de plus en plus petit. La séquence d'écartement des cylindres par laquelle le grain passe détermine la finesse et la composition de la farine. Le savoir-faire du meunier lui permet par un processus d'essais/erreurs de déterminer la meilleurs séquences pour obtenir le produit désiré.

Les deux cylindres ont une distance qui les sépare ajustable. Le réglage de l'écartement des cylindres se fait en ajustant deux poignées d'ajustement se trouvant aux deux extrémités des cylindres mobiles. Chaque poignée d'ajustement ajuste l'écartement d'un côté du cylindre. Cela permet d'avoir un écartement progressif sur la longueur des deux cylindres. Par exemple si le côté droit est serré à son maximum et le côté gauche à son minimum, l'écart entre les deux cylindres ira en grandissant de la gauche vers la droite. De même, pour avoir un écartement constant les deux manivelles doivent être ajustées de la même façon. La figure 5.3 montre le mécanisme d'une poignée.

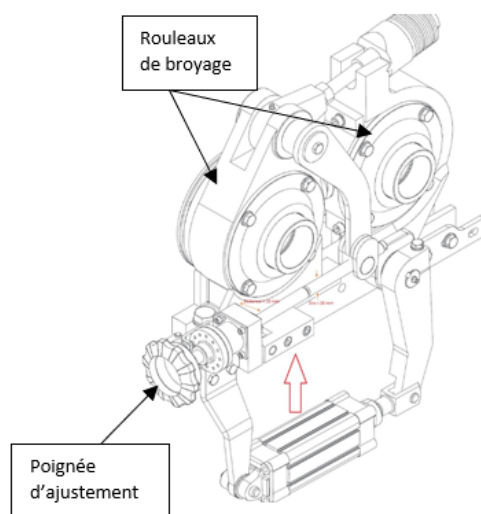


FIGURE 5.3 Poignée d'ajustement des rouleaux de broyage

La figure est un plan qui isole une des extrémités des rouleaux. La poignée d'ajustement est connectée aux rouleaux de broyages. Il suffit de tourner la poignée dans le sens horaire pour serrer les rouleaux. L'objectif du projet est de documenter les réglages des rouleaux. Pour se faire un système de récupération sur mesure a été développé. Ce système de récupération imite le nombre de tours effectués par la poignée d'ajustement grâce à une hélice. Le nombre de tours fait par l'hélice est répertorié à l'aide d'un capteur infrarouge. La photo 5.4 montre l'installation de l'hélice imprimée en 3D.

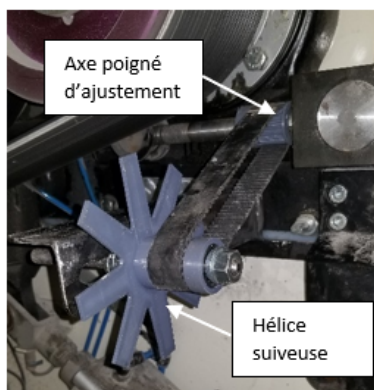


FIGURE 5.4 Hélice du capteur infrarouge. Par Abdellah Aboukhar

Le capteur utilisé est installé de sorte à suivre les réglages que le meunier fait. Les rouleaux ne sont pas munis de capteurs indiquant leurs serrages. Il est difficile d'utiliser des capteurs optiques dans des milieux poussiéreux. L'appareil ne peut donc pas être installé près des rouleaux de broyage. De plus, il faut éviter d'entrer en contact avec le produit alimentaire. Le capteur est installé de telle sorte à enregistrer le nombre de tours que l'opérateur effectue avec la poignée de serrage. Cela donne un suivi sur les ajustements faits selon le blé utilisé, les conditions de broyage et le type de farine souhaitée. Permettre un enregistrement automatique de ces prises de décisions évite la perte d'information. Cela revient à interviewer le meunier sur son travail, mais sans lui faire perdre son temps.

Le système est composé de deux capteurs infrarouges et une hélice qui suit le mouvement de la poignée de réglage. Quand l'opérateur manipule la poignée de réglage, l'hélice tourne dans le même sens. En tournant, l'hélice vient activer les capteurs infrarouges qui enregistrent combien de tours d'hélice ont été réalisés. L'utilisation de deux capteurs permet de connaître le sens de l'hélice. Ces deux informations combinées permettent d'établir de combien de tours l'opérateur a serré ou desserré les rouleaux de broyages. L'image 5.4 montre comment le système a été monté. Les pièces ont été imprimées en 3D pour correspondre aux besoins spécifiques de notre cas.

La création de ce système de récupération a engendré beaucoup de recherches sur la qualité de la donnée qui sera récupérée. La méthodologie d'évaluation de la qualité de la donnée a été utilisée lors de la création de ce système de récupération. L'application de la méthodologie présentée au chapitre 4 est présentée ci-bas.

5.2.1 Étape 1-Définition des attentes qualité

Pour réaliser ce projet de telle sorte à correspondre aux besoins en qualité des données, il a fallu définir les besoins en qualité. Des discussions entre les parties prenantes ont été nécessaires. En effet l'installation d'un capteur nécessite l'expertise de plusieurs personnes. Le meunier est la personne référence pour décrire la façon de régler les rouleaux de broyage. Il est l'expert qui a les informations sur les fréquences de réglage, les corrections ainsi que l'amplitude des réglages. Sa connaissance des appareils a permis d'établir les possibilités d'installation. Vu que le capteur essaye de documenter la façon de travailler du meunier, l'installation doit s'adapter à ses habitudes. Le travail avec l'analyste de données a mis en évidence ses besoins en termes de données. Étant la personne qui utilisera les données, il faut récupérer les données les plus proches de ses besoins. Définir le format des données, où les enregistrer, leur niveau de précision est fait avec l'analyste. Pour définir une méthode de récupération, il a fallu mener une étude de marché. En effet, une exploration des différentes technologies disponibles aide à orienter le choix technologique. Pour réaliser l'étude de marché il a fallu contacter des professionnels en capteur et rechercher plusieurs marques et options différentes.

Pour mener à bien les entrevues avec les collaborateurs de ce projet, les questions proposées dans le tableau présenté dans la section 4 ont été utilisées. Après avoir récupéré les informations nécessaires au courant de plusieurs interactions, elles ont pu être compilé de façons intelligibles. Le tableau de prise de décision est complété pour le capteur de serrage des rouleaux de broyage. Ce tableau est présenté à la figure 5.1.

Comme vu lors de l'introduction de ce projet de récupération de données, l'étape de broyage joue un rôle important dans le processus de production. Il a donc été présumé que les données liées à cette étape du processus de production auront un impact sur la prédiction de la qualité de la farine. L'impact réel de cette donnée ne pourra être déterminer qu'une fois l'analyse de donnée faite. Une période d'essai est prévue pour déterminer l'importance de la donnée. La précision nécessaire pour donner une représentation fidèle des réglages a été discutée avec le meunier. Lors d'un réglage plusieurs tours de poignée d'ajustement sont réalisés. En général cela varie entre 1 et 3 tours. Il est jugé qu'une précision d'un cinquième de tours serait tolérable pour l'analyse. Cela se traduit par l'utilisation d'une hélice suiveuse possédant au moins 5 pales. La fidélité à la réalité est jugée bonne, car il y a une corrélation directe entre la distance entre les cylindres de broyage et le nombre de tours de poignée d'ajustement. Toutes ces informations sont inscrites dans la partie précision acceptable du tableau 5.1.

Dans la partie gamme des instruments se sont les questions liées au budget qui sont abordées. Vu qu'il est question d'une phase d'essai pour déterminer l'importance de la donnée, il

TABLEAU 5.1 Tableau de prise de décision pour le capteur serrage rouleaux

Décision	Où trouver l'information	Décision prise
Précision acceptable	➤ Importance de la donnée à récupérer ○ Importance estimé haute, car l'étape exerce une grande influence sur le processus de fabrication. ➤ Précision du capteur ○ Après discussion avec le meunier, les ajustements sont faits par rotation de la poignée de 1 à 3 tours. Avoir une précision de 1/5 de tour est suffisant. ○ Le système doit être remis à zéro de temps à autre pour être certain des valeurs. ➤ Fidélité avec la réalité ○ Oui la réalité est bien reproduite.	L'hélice du système doit posséder au moins 5 pales. Une remise à zéro doit être effectuée à des intervalles réguliers.
Gamme des instruments de collecte	➤ Le budget défini ○ Le budget final sera défini après une période d'essai pour valider les résultats. L'installation est expérimentale pour l'instant. ➤ Grade à respecter ○ Le voltage utilisé est acceptable selon les normes.	L'installation doit être résistante mais les capteurs utilisés peuvent être d'entrée de gamme. Pour une période d'essai la durabilité n'est pas importante.
Formation du personnel	➤ Automatisation ○ Complètement automatique. Ne change pas les habitudes des utilisateurs. ➤ Base de données ○ Les données doivent pouvoir être poussées dans le système en place.	L'utilisation d'un Arduino est une bonne méthode de récolte de données. Le mécanisme permet une récupération automatique.
Maintenance des instruments de collecte	➤ Panne possible ○ Le capteur brise/ La poussière s'accumule sur le capteur / La connexion se rompt ○ Détection de panne peut prendre du temps	Surveiller les pannes avec attention pendant la période d'essai.
Fréquence	➤ Fréquence de récupération ○ Pas de problème de fréquence. Les données sont récupérées plus souvent que les ajustements.	
Méthode d'installation du système de collecte	➤ Risque de contamination ○ Le capteur n'a pas de risque d'entrer en contact avec le produit. ➤ Quantité d'information ○ La date et l'heure peuvent être récupérées. Il est possible d'installer d'autres capteurs avec le Arduino	Installer un Arduino pour moduler les capteurs.
Règles de sauvegarde	➤ Fréquence de récupération ○ Les ajustements sont peu fréquents ➤ Base de données ○ Pas de manutention nécessaire.	Il faut nettoyer la base de données des données redondantes

est possible d'utiliser des capteurs d'entrée de gamme avant d'investir dans un équipement industriel coûteux. Cependant les broyeurs génèrent beaucoup de vibrations que le système de récupération de données doit supporter cette situation. De plus, le voltage impliqué dans l'utilisation du système doit respecter les normes de sécurité d'un milieu poussiéreux.

En terme de formation du personnel il n'y a pas un grand impact. Le système est fait pour s'intégrer automatiquement à la pratique des opérateurs. Cependant par soucis d'éviter un écart de mesures, il est conseillé de remettre le système à zéro périodiquement pour s'assurer du bon calibrage. La récupération des données se fait automatiquement et s'enregistre dans la base de données de l'entreprise. Un Arduino est choisi pour gérer les capteurs ainsi que la transmission des informations.

La maintenance est un sujet sensible. Le milieu d'installation est poussiéreux et contient beaucoup de vibrations. Les capteurs peuvent être amenés à perdre en visibilité à cause de la poussière. Des bris mécaniques sont possibles à cause des vibrations. L'Arduino peut connaître des problèmes de connexion au réseau. Les chances qu'une panne se produise sont importantes. L'idée de remettre à zéro le système de façon périodique sert de maintenance. Détecter ces failles est compliqué car les meuniers ne suivent pas une routine pour faire les réglages des rouleaux de broyage. Il faut profiter de la période d'essai pour établir les nécessités en terme de maintenance.

Lors de l'installation il faut s'assurer que le capteur n'entre pas en contact avec la farine produite. Toute forme de contamination est dangereuse. L'installation d'une hélice excentrée des rouleaux de broyage évite les contaminations. De plus, l'utilisation d'un Arduino donne la possibilité d'ajouter d'autre capteurs. En effet un Arduino est un processeur qui offre la possibilité de faire évoluer le système et d'ajouter des modules. Par exemple, une caméra thermique qui enregistre le réchauffement des rouleaux de broyage pourrait être ajoutée.

Les réglages effectués par le meunier étant peu fréquents, 2 ou 3 fois par course de production, il n'y a pas de problème lié à la fréquence de récupération. Les informations peuvent être poussées dans la base de données de façon automatique. Cependant, des règles pour effacer les données redondantes peuvent être mises en place.

5.2.2 Étape 2- développement des KPIs

Le tableau des besoins en qualité des données donne des indications sur les défis de qualité de cette donnée. Vu que la récupération est automatique et que l'installation est faite sur mesure, ces deux catégories ne posent pas de problème. Cependant, la précision doit être au niveau des attentes et permettre de valider qu'elle est suffisante. De plus, le milieu poussiéreux et vibrant cause des problèmes de maintenance. Il faut avoir un KPI qui suit le nombre de pannes. Les deux KPIs reliés à ce capteur restent assez basiques. Deux métriques se dégagent et sont définies dans des banques de KPIs :

1. La précision : Savoir si le système mis en place donne une information assez fine sur les réglages faits par l'opérateur.
2. Les pannes : Savoir si le nombre de pannes est trop élevé pour réussir à utiliser les informations

Pour la précision, le choix de la métrique sera une fraction de tour. L'hélice installée comporte 8 pales. Cela signifie que nous aurons une précision de 1/16 de tour. En effet, les capteurs comptabilisent le nombre de pales passées devant leurs détecteurs et enregistrent le sens de rotation. C'est le changement d'état des capteurs qui indique la rotation. on peut savoir

si les capteur sont sur une pale ou un creux. On a donc 16 possibilités que nous pouvons reconnaître.

Pour les pannes, il a été établi que la période actuelle était une période d'essai. Néanmoins, une étude du nombre de pannes est importante, car il donne aussi la possibilité d'améliorer la détection de ces dites pannes. Pour le calcul de ce KPI, il faut maintenir un journal des pannes. Cela permet de trouver la fréquence de leur occurrence. Cette fréquence aide aussi à prévoir la maintenance. De plus, prendre en note les raisons d'une panne offrira une meilleure compréhension de ce qui doit être fait pour les reconnaître et les éviter. Un KPIs doit être construit en sachant sur quoi se base la cause du problème pour réussir à le régler. Les sources de pannes peuvent être multiples, les lister avec des fréquences d'occurrence aidera à déterminer des actions à mener.

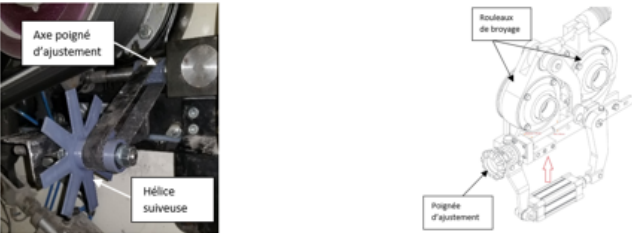
La phase de test sert à trouver les améliorations nécessaires à apporter. Elle sert aussi à évaluer l'importance de la donnée récupérée. Le système actuel sera amené à évoluer et certaines composantes seront à changer. Le KPI "Évolution" développé dans la section 4.3.4 est utilisé dans la section suivante pour évaluer la capacité du système à être modifié. Ainsi un utilisateur est capable d'évaluer la possibilité d'adapter le capteur à ses besoins.

5.2.3 Étape 3- Évaluation de la qualité de la donnée

Cette partie vise à synthétiser les informations développées au fil de l'étude pour évaluer la qualité de la donnée. Toutes les informations pertinentes sont inscrites dans le tableau des méta-données descriptives 5.2. La définition du système de récupération, les personnes qui l'utilisent et le maintiennent ainsi que les KPIs développés.

Dans ce cas particulier, le KPI sur la précision de l'installation est facile à évaluer. Les huit pales de l'hélice offrent une précision au 1/16 de tour. Cette précision est plus grande que le 1/5 mentionnée dans la sous-section des besoins. La précision est donc validée pour les tests à effectuer. Il se peut que ces besoins évoluent si le meunier change ses habitudes d'ajustement. Le design fait en impression 3D permet une évolution du système. Grâce au KPI "capacité évolutive" vu à la section 4, nous pouvons mettre une note de 2,5 pour la capacité évolutive. En effet, l'achat d'un meilleur capteur améliore les performances, et il est aussi possible de réimprimer un nouveau système facilement. Le KPI sur la fréquence des pannes est ajouté à la liste des KPIs utilisés mais n'a pas encore une méthode de calcul qui lui est propre.

TABLEAU 5.2 Méta-données remises à jour avec KPI.

Titre: Serrage des rouleaux de broyage		
Déscription	Utilisation et sens de la donnée: La donnée enregistrée indique le nombre de tours de serrage fait par un opérateur. Cette donnée est aussi traduite en mm d'écartement entre les rouleaux de mouture. Une poignée joue sur le serrage des rouleaux d'un seul côté.	
Dates	Date de la première mise en place: Juillet 2022	Date de la dernière évaluation
Personnes	Nom de la personne responsable de l'installation: -khelil Joudane -Abdellah Aboukhar	Nom de la personne responsable de la maintenance: -Abdellah Aboukhar
	Listes de personnes ayant accès à/utilisant la donnée: Khelil Joudane/ Loïc Parenin	
Technologie	Processus de récupération: Deux capteurs infrarouges détectent le nombre de tours d'ajustement fait par l'opérateur. Une hélice suit le nombre de tour qu'effectue la poignée et active la détection infrarouge.	Technologie: Description de la technologie utilisée: Un Arduino est utilisé avec deux capteurs infrarouges basiques. (Les capteurs devront être remplacés par un des capteurs plus adaptés à l'industrie après la période d'essai)
	Schéma ou dessin aidant à la compréhension (si nécessaire)	
Évaluation		
	Liste des KPI utilisés:	Commentaires relatifs au KPI
	Précision:	1/8 de tour - approuvé car plus précis de 1/5
	Capacité évolutive:	2,5/3
	Fréquence de panne:	à déterminer

5.3 Projet 2 : Mesure du niveau dans les silos

Ici, l'intérêt porte sur les silos d'entreposage de la matière première. En effet, comme vu dans la figure 5.1, le grain est d'abord entreposé, puis des mélanges de grains de différents silos sont réalisés pour respecter des recettes. Ces recettes sont faites pour équilibrer les caractéristiques des différents grains. Cela permet de développer une farine équilibrée et qui a des attributs constants malgré une matière première variable. Pour cela, le meunier va activer les vannes des silos pour que le blé s'écoule et se mélange. Pour gérer l'écoulement et les stocks, des équilibreurs de flux sont installés aux sorties des silos.

Deux problèmes apparaissent avec ce système :

1. les équilibreurs de flux ont des défauts qui les empêchent d'avoir un suivi précis. Cela donne une difficulté de suivi des niveaux de blé dans les silos ;
2. Les équilibreurs n'enregistrent pas les informations sur les entrées de blé. Les silos sont parfois remplis avant d'être totalement vides. Même s'ils sont remplis de grain qui se ressemblent, il y a des différences entre chaque arrivage. Les caractéristiques du

mélange de blé dans un silo sont actuellement une moyenne simple des caractéristiques des différents blés qui s'y trouvent.

Cela va créer des problèmes dans le respect des recettes et la création de recettes précises. Pour cela, le meunier en chef inspecte les silos grâce à un détecteur de distance laser. Cela lui permet de suivre le niveau des silos. Mais cela prend non seulement beaucoup de temps, mais aussi l'information n'est pas enregistrée dans la base de donnée. Pour cela, un projet d'installation de capteurs de niveaux a été entrepris. Ici, la méthodologie au complet ne sera pas présentée, mais certains KPI intéressants seront explorés. Quelques points sur la prise de décisions sont à relever. En effet, l'installation requiert de percer les silos, ce qui est une modification permanente de l'équipement. Cela a créé une réticence à l'installation, qui a été discutée.

L'un des problèmes soulevés lors de la prise de décision est la véracité de l'information qui sera obtenue. En effet, le niveau de grain n'est pas une surface plane. Lors de l'ajout des grains de blé, une dune se crée à l'endroit où les gains tombent. Lors du retrait des grains, un gouffre se crée à l'endroit où le grain sort. Cette différence de topologie sur la surface de grain du silo fait varier le volume de grain contenu. Le capteur qui a été sélectionné mesure le niveau à un point donné. Le fournisseur conseille de positionner le capteur de sorte à avoir un point de mesure qui varie peu. Ce point de mesure sera considéré comme la médiane des hauteurs. Cela pousse à avoir une réflexion sur l'erreur que cette vérité du terrain engendre sur nos mesures. Un KPI peut être développé pour rendre compte de cette erreur. Nous parlons ici d'un manque d'exactitude. L'installation est encore au début de ses tests, mais un protocole pour rechercher cette information est émis. Avoir une comparaison entre la mesure de l'appareil fixe et les mesures que le meunier prend avec différents points de mesures aide à calculer la différence entre le réel et la mesure.

Lors de l'exécution de ce projet deux catégories du tableau des besoins en qualité ont été en contradiction. La précision est difficile à atteindre à cause des phénomènes de dune et de creux dans le silo. La gamme des instruments peut varier grandement et le prix des capteurs peut varier de dizaines de milliers de dollars. Il a donc fallu trouver un équilibre entre le prix et la précision à obtenir.

Documenter cette information permet de prendre des précautions lors de l'utilisation de la donnée pour des analyses. Aussi cela permet de travailler à améliorer les mesures en y ajoutant des coefficients de correction. Le passage par la méthodologie présentée a permis de documenter les problèmes de cette solution. Développer des outils de mesure de qualité et un protocole pour réaliser ces mesures. Enregistrer toutes les étapes entreprises permet aux utilisateurs d'apprécier les données qui leur sont fournies. Cela augmente la confiance dans

les données.

5.4 Projet 3 : Suivi des palettes de farine

Le projet de suivi des palettes de farine est un projet qui a connu plusieurs embûches. Ce projet était pensé pour suivre les palettes de sac de farine dans l'entrepôt. Hors avant de se lancer dans ce projet qui allait coûter cher, l'entreprise a voulu prendre une période d'essai en utilisant des codes à barres. Cette période d'essai visait à évaluer le nombre d'informations que l'entreprise pouvait avoir en réalisant ce projet. Ce projet est intéressant dans l'utilisation du tableau de prise de décision. Ici seront présentées les deux sous-sections qui ont été cruciales. Le projet est encore en cours de développement.

En se basant sur le questionnaire développé dans la section 4 étape-1 certaines de ces questions vont être étudiées ici.

- Base de données :
 - Les données sont-elles compatibles avec la base de données ? Réponse : La base de données utilisée est rigide et ne permet pas de modifications. Certaines informations prises à la main par les opérateurs ne pouvaient pas être enregistrées.
 - Les données sont-elles facilement accessibles ? Réponse : L'accessibilité aux données n'est pas un problème. C'est l'accessibilité à la plate-forme d'enregistrement de la donnée qui est difficile. L'opérateur doit se reconnecter entre chaque enregistrement, cela alourdit la tâche.
 - Il y a-t-il une copie de sauvegarde ? Réponse : Oui la donnée s'enregistre sur l'ERP de l'entreprise.

Pour régler les problèmes d'enregistrements, des discussions ont été ouvertes avec le fournisseur de l'ERP. Il manque la possibilité d'enregistrer certaines informations sur la plate-forme d'entrée de données et la connexion à cette plate-forme est pénible. Des demandes de modifications ont été envoyées et le projet attend l'évolution de ces demandes. Cela indique une difficulté à faire évoluer le système en place. Le KPI évolution a une mauvaise note dans ce cas. Cette mauvaise performance du KPI est due à un manque de flexibilité de la part de l'ERP en place.

- Quantités d'informations
 - Peut-on obtenir plusieurs informations avec la même installation ? Réponse : Les données peuvent être scannées ou entrées à la main sur la plate-forme. Plusieurs informations sont enregistrées sur papier actuellement. Exemple : Numéro de la course, numéro de palette, le nom de l'opérateur, le nombre de rebuts, le silo de provenance. Enregistrer ces informations nécessite de créer des entrées spéciales

pour elles.

- Cela complique-t-il le projet ? Réponse : Oui, plus il y a d'information, plus l'opérateur doit prendre du temps pour enregistrer tout cela sur le système. De plus, si nous voulons enregistrer des informations spéciales il faut leur prévoir des entrées sur la plate-forme.

Il est possible de récolter les informations désirées, mais la complexité augmente. C'est donc une décision qui doit être prise sur les informations importantes à suivre pour ne pas compliquer le travail de l'opérateur. Une tâche qui comporte plusieurs étapes complexes crée des erreurs.

5.5 Résultats de l'étude de cas

Le premier exemple a permis une utilisation complète des outils développés. Passer par toutes les étapes et a aussi permis de tester la méthodologie. Cela a rendu un résultat satisfaisant en améliorant la rapidité de prise de décision grâce à l'étape 1. En effet, le gabarit du tableau de prise de décision aide à évaluer plus vite les enjeux d'un projet. Les deux dernières étapes permettent quant à elle de créer des KPI et de les utiliser pour évaluer les données. À la fin de ces deux étapes un document compilant toute les informations relatives au capteur est produit. Ce document de métadonnées permet à tout utilisateur d'avoir un contexte lors de l'utilisation des données en question. Le document de métadonnées donne une compréhension de la situation initiale lors d'une boucle d'amélioration.

Les deux exemples suivants ont servi à apporter des exemples complémentaires sur l'utilisation des outils développés. Dans l'exemple des capteurs de niveau dans les silo, la définition des problématiques liées à la précision de l'information guide le débat sur la gamme d'instruments à utiliser. La documentation des problématiques engendre des projets d'amélioration continue qui pousse à l'innovation. Dans l'exemple de suivi des sacs de farines, lister les problématiques de l'installation des RFID mène à découvrir qu'il faut adapter l'ERP pour accommoder les employés sur le terrain.

CHAPITRE 6 CONCLUSION ET RECOMMANDATIONS

Ce mémoire a permis d'explorer la notion de qualité des données récupérées en contexte industriel, plus précisément, en contexte des PME qui ont la volonté d'entrer dans l'ère du numérique. Pour résumer le contexte du mémoire, il faut rappeler que les PME ont plus de difficulté à adopter des technologies de l'industrie 4.0. Ces technologies sont plus facilement implémentées dans de grandes entreprises qui possèdent les moyens financiers et des connaissances techniques nécessaires. L'existence d'expert en informatique et en installation de technologie numérique aide à lancer des projets de numérisation. Dans des entreprises de petite taille, il manque des ressources humaines pour mener ce genre de projet. Pour rattraper ce retard, de plus en plus de projets de recherche sont lancés dans le domaine du numérique pour créer des outils faits pour les PME.

L'entreprise partenaire avait comme désir de prévoir la qualité de sa farine en s'appuyant sur de l'analyse de données. Il y avait donc un désir d'installer différents capteurs sur la chaîne de production. Cela permettait de remonter des données du plancher de production, mais les données étaient récoltées de façon plus ou moins conventionnelle. Il fallait valider leur qualité. C'est ce qui a inspiré ce mémoire.

La revue de littérature montre de sérieux manques dans l'évaluation de la qualité d'une donnée. Même s'il existe des méthodologies pour mener des projets pour améliorer la collecte et l'utilisation des données rien est dit sur l'évaluation de la qualité des données récoltées. Cela a mené à une réflexion sur la qualité des données et leurs capacités à répondre aux besoins d'un projet. La problématique qui est ressortie est : "Comment développer une méthodologie qui permet : 1) de définir les besoins en qualité des données d'un projet de transformation numérique, 2) d'évaluer cette qualité et 3) d'en faire le suivi ?"

6.1 Synthèse des travaux

Pour répondre à cette problématique, une méthodologie a été développée. Un livre a inspiré la démarche, "Executing data quality projects : Ten steps to quality data and trusted information" [McGilvray, 2021]. Ce livre présente une méthodologie pour mener à bien des projets de numérisation. Il insiste sur la nécessité de comprendre les besoins en information de l'entreprise pour lancer des projets pertinents. Avec la même logique, une méthodologie a été créée pour vérifier que la qualité d'une donnée récupérées. D'abord, il faut comprendre les besoins en qualité des données d'un projet. En effet, une donnée de qualité acceptable

est une donnée qui satisfait un besoin. Ensuite, il faut mettre en place des métriques et les utiliser pour évaluer la donnée. Enfin, il faut développer un moyen pour documenter l'analyse qualité qui est menée. Le travail d'évaluation doit être enregistré pour pouvoir faire évoluer l'entreprise.

L'étape préliminaire de la méthodologie inscrit la démarche d'évaluation de la qualité de données dans un projet global de numérisation. L'étape préliminaire reprend les étapes initiales du livre "Executing data quality projects : Ten steps to quality data and trusted information" [McGilvray, 2021] et évalue l'état initial de l'entreprise ainsi que ses besoins. C'est une évaluation globale de l'expertise de l'entreprise à utiliser ses données.

L'étape 1 a été développée pour définir les besoins d'un projet de valorisation de données particulier. Chaque projet de valorisation a besoin de données avec une certaine qualité. Il fallait donc créer un outil pour décortiquer les besoins d'un projet. Cet outil est un gabarit prenant la forme d'un tableau de prise de décision. Son application durant l'étude de cas a permis une compréhension facile des enjeux de chaque projet.

L'étape 2 permet de créer des KPI et les utiliser pour évaluer la qualité des données. Ici, la qualité est divisée en plusieurs dimensions pour couvrir tous les aspects de la donnée. L'évaluation est donc plus précise. Dans les cas d'étude, on montre que les notes servent à évaluer ce qui est à améliorer dans le système de récupération de données.

L'étape 3 est une étape qui permet de documenter le processus de récupération des données. Des informations importantes sont répertoriées pour permettre à l'utilisateur des données de comprendre le contexte de récupération. Ces informations complémentaires au set de données sont appelées méta-données. Les méta-données permettent aussi de compiler les évaluations des données récupérées et suivre ainsi leurs évolutions. Cette étape facilite les boucles d'améliorations, elle renseigne sur l'état initial.

6.2 Limitations de la solution proposée et solutions

La méthodologie mise en place a été testée sur plusieurs cas d'étude, mais dans une seule entreprise. L'entreprise en question produit de la farine biologique. D'autres industries auront certainement des besoins très différents. Il est difficile d'affirmer que la méthodologie sera facile à calquer dans d'autres industries. Cependant, elle peut être adaptée et modifiée. Son utilisation dans d'autres contextes permettra de voir la capacité de la méthodologie à s'adapter à d'autres besoins.

La méthodologie doit aussi être utilisée durant les installations des systèmes de récupération de données. Si les données sont déjà produites et utilisées par des analystes externes à l'en-

treprise, ils ne pourront pas juger de la qualité de la donnée. Cette méthodologie encourage les visites sur le terrain pour pouvoir juger de la donnée récupérée.

De plus, la méthodologie est développée dans une PME et sur un nombre réduit de capteurs. Une perspective de recherche peut consister à appliquer cette méthodologie à plus grande échelle. La gestion des métadonnées créées à l'étape 3 à plus grande échelle posera plusieurs problématiques. Un lien entre les systèmes de récupération de données et les métadonnées qui leur sont reliées doit être fait. Il faut donc penser à la structure que prendra un tel lien dans le système d'information. De plus, mettre en place une norme d'utilisation de cette méthodologie dans une entreprise doit être pensée. Ces questions d'implantation peuvent constituer un sujet de recherche futur.

RÉFÉRENCES

- Latif Al-Hakim. Information quality function deployment. In *Challenges of Managing Information Quality in Service Organizations*, pages 26–51, 2007. doi : 10.4018/978-1-59904-420-0.ch002.
- Farhan Alfin. Controlling wheat milling operation- break release. *Miller*, 119 :270–276, 2007. ISSN 0041-624x. doi : 10.1016/j.ultras.2007.03.002.
- Rawhi Alrae, Qassim Nasir, and Manar Abu Talib. Developing house of information quality framework for iot systems. *International Journal of System Assurance Engineering and Management*, 11(6) :1294–1313, 2020. doi : 10.1007/s13198-020-00989-6.
- Angelos Angelopoulos, Emmanouel T. Michailidis, Nikolaos Nomikos, Panagiotis Trakadas, Antonis Hatziefremidis, Stamatis Voliotis, and Theodore Zahariadis. Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects. *Sensors*, 20(1) : 1–34, 2022. ISSN 1424-8220. doi : 10.3390/s20010109.
- Eyad Ayoubi and Shadi Aljawarneh. Challenges and opportunities of adopting business intelligence in smes : Collaborative model. In *1st International Conference on Data Science, E-Learning and Information Systems*, article no. : 42, pages 1–5, 2018. ISBN 9781450365369. doi : 10.1145/3279996.3280038.
- Dan Braha. *Data mining for design and manufacturing : methods and applications*, volume 3. Springer Science & Business Media, 2013.
- Gouvernement du Canada. *Principales statistiques relatives aux petites entreprises*. 2020. URL https://www.ic.gc.ca/eic/site/061.nsf/fra/h_03126.htm.
- Qing Chang, Robert Gao, Yong Lei, Lihui Wang, and Changxu Wu. Cyber-physical systems in manufacturing and service systems. *Mathematical Problems in Engineering*, 2015 :2, 2015. ISSN 1024-123X. doi : 10.1155/2015/704213.
- Zoubida Chorfi, Abdelaziz Berrado, and Loubna Benabbou. Selection of key performance indicators for supply chain monitoring using mcda. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 1–6, 2021. doi : 10.1109/IEEM50564.2021.9672871.
- Shirley Coleman, Rainer Göb, Giuseppe Manco, Antonio Pievatolo, Xavier Tort-Martorell, and Marco Seabra Reis. How can smes benefit from big data? challenges and a path

- forward. *Quality and Reliability Engineering International*, 32(6) :2151–2164, 2016. ISSN 07488017. doi : 10.1002/qre.2008.
- Christophe Danjou, Robert Pellerin, and Louis Rivest. *Le passage au numérique : Industrie 4.0 : des pistes pour aborder l'ère du numérique et de la connectivité*. Centre francophone d'informatisation des organisations (CEFRIO), 2017.
- Anna De Carolis. A methodology to guide manufacturing companies towards digitalization. 2017.
- Nikolaos Efkarpidis, Andrija Goranovic, Chen-Wei Yang, Martin Geidl, Ingo Herbst, Stefan Wilker, and Thilo Sauter. A generic framework for the definition of key performance indicators for smart energy systems at different scales. *Energies*, 15(4) :1289, 2022. doi : <https://doi.org/10.3390/en15041289>.
- Dieter Etz, Thomas Frühwirth, and Wolfgang Kastner. Flexible safety systems for smart manufacturing. In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–4, 2020. doi : 10.1109/ETFA46521.2020.9211905.
- Pierre Feillet. *Le grain de blé : composition et utilisation*. Editions Quae, 2000.
- Charlotte Fortin. L'industrie 4.0 : Sondage auprès de pme québécoises du secteur manufacturier. pages 1–53, 2019.
- Maryam Ghasemaghaei. Does data analytics use improve firm decision making quality ? the role of knowledge sharing and data analytics competency. *Decision Support Systems*, 120 : 14–24, 2019. doi : 10.1016/j.dss.2019.03.004.
- Dale L Goodhue. Understanding user evaluations of information systems. *Management science*, 41(12) :1827–1844, 1995. doi : 10.1287/mnsc.41.12.1827.
- Stephanie Guerlain, Donald E. Brown, and C. Mastrangelo. Intelligent decision support systems. In *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics.*, pages 1934–1938, 2000. ISBN 0-7803-6583-6. doi : 10.1109/ICSMC.2000.886396.
- Mario Hermann, Tobias Pentek, Boris Otto, et al. Design principles for industrie 4.0 scenarios : a literature review. *Technische Universität Dortmund, Dortmund*, 45(1) :1–15, 2015.

- Senthil Kumar Jagatheesaperumal, Mohamed Rahouti, Kashif Ahmad, Ala Al-Fuqaha, and Mohsen Guizani. The duo of artificial intelligence and big data for industry 4.0 : Applications, techniques, challenges, and future research directions. *IEEE INTERNET OF THINGS JOURNAL*, 9(15) :12861–12885, 2022. doi : 10.1109/JIOT.2021.3139827.
- Iris Kirchen, Daniel Schütz, Jens Folmer, and Birgit Vogel-Heuser. Metrics for the evaluation of data quality of signal data in industrial processes. In *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, page 819–826, 2017. ISBN 978-1-5386-0837-1. doi : 10.1109/INDIN.2017.8104878.
- Sri Kolla, Meysam Minufekr, and Peter Plapper. Deriving essential components of lean and industry 4.0 assessment model for manufacturing smes. In *52nd CIRP Conference on Manufacturing Systems*, page 753–758, 2019. ISBN 2212-8271.
- Daria Larsson, R.M. Chandima Ratnayake, and Arne Gildseth. Development of a balanced score card for knowledge work in projectoriented engineering organization : Kpi prioritization using ahp. In *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 1–6, 2021. doi : 10.1109/IEEM50564.2021.9672871.
- Yang W. Lee, Diane M. Strong, Beverly K. Kahn, and Richard Y. Wang. Aimq : a methodology for information quality assessment. *Information & Management*, 40(2) :133–146, 2002. doi : 10.1016/S0378-7206(02)00043-5.
- Danette McGilvray. *Executing data quality projects : Ten steps to quality data and trusted information (TM)*. Academic Press, 2021. ISBN 0128180161.
- Dimitris Mourtzis, Anna-Maria Papatheodorou, and Sophia Fotia. Development of a key performance indicator assessment methodology and software tool for product-service system evaluation and decision-making support. *Journal of Computing and Information Science in Engineering*, 18(4) :1–13, 2018. doi : 10.1115/1.4040340.
- David Parmentier. *Key Performance Indicators : Developing, Implementing, and Using Winning KPIs*. Wiley, Hoboken, NJ, 3rd ed edition, 2015. ISBN 978-1-119-01984-8.
- Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. Data quality assessment. In *Communications of the ACM Volume 45 Issue 4*, page 211–218, 2002. doi : 10.1145/505248.506010.
- Michael Porter and James Heppelmann. How smart, connected products are transforming competition. *Harvard business review*, 92(11) :74–80, 2019.

- Gouvernement du Québec. Plan d'action en économie numérique. pages 1–78, 2016. ISBN 978-2-550-75695-8.
- Gouvernement du Québec. *Agriculture biologique*. <https://www.quebec.ca/agriculture-environnement-et-ressources-naturelles/agriculture/agriculture-biologique>, 2021.
- Office québécois de la langue française. *Métadonnée*. 2020.
- Brozzi R., D'amico R. D., Pasetti G. Monizza, Marcher C., Riedl M., and Matt D. T. Design of self-assessment tools to measure industry 4.0 readiness. a methodological approach for-craftsmanship smes. *15th IFIP International Conference on Product Lifecycle Management (PLM)*, 15 :566–578, 2018. doi : 10.1007/978-3-030-01614-2_52.
- Stuart Russell and Peter Norvig. *Intelligence artificielle*. Pearson, 3rd ed edition, 2010. ISBN 2744074554.
- Kaplan R. S. and Norton D. P. “measuring the strategic readiness of intangible assets. *Harvard Business Review*, 82(2) :52–63, 2004.
- Imad Saleh. Internet des objets (ido) : Concepts, enjeux, défis et perspectives. *Internet des objets*, 1(2) :19, 2018.
- Mark C Scott. *Reinspiring the corporation : the seven seminal paths to corporate greatness*. Wiley, 2000.
- Kaustav Sen. Does the measure of information quality influence survival bias? *International Journal of Quality & Reliability Management*, 18(9) :967–981, 2001. ISSN 265-671X. doi : 10.1108/02656710110407136.
- František Simetingera and Josef Basla. A pilot study : An assessment of manufacturing smes using a new industry 4.0 maturity model for manufacturing small- and middlesized enterprises (i4mmsme). *Procedia Computer Science*, 200 :1068–1077, 2022.
- Marco Spaltini, Federica Acerbia, Marta Pinzonea, and Marco Taischa Sergio Gusmerolia. Defining the roadmap towards industry 4.0 : The 6ps maturity model for manufacturing smes. In *29th CIRP Life Cycle Engineering Conference*, page 631–636, 2022. ISBN 2212-8271.
- J Stausberg, U Bauer, D Nasseh, R Pritzkeleit, CO Schmidt, T Schrader, and M Nonnemacher. Indicators of data quality : review and requirements from the perspective of networked medical research. In *GMS Medizinische Informatik, Biometrie und Epidemiologie*, pages 1–8, 2019. doi : 10.3205/mibe000199.

- Diane M Strong, Yang W Lee, and Richard Y Wang. Data quality in context. *Communications of the ACM*, 40(5) :103–110, 1997.
- Amr T. Sufian, Badr M. Abdullah, Muhammad Ateeq, Roderick Wah, and David Clements. A roadmap towards the smart factory. In *2019 12th International Conference on Developments in eSystems Engineering (DeSE)*, pages 1–6, 2019. doi : 10.1109/DeSE.2019.00182.
- Zhanybek Suleiman, Dinara Dikhanbayeva, Sabit Shaikholla, and Ali Turkeyilmaz. Readiness assessment of smes in transitional economies : Introduction of industry 4.0. In *The 2nd International Conference on Industrial Engineering and Industrial Management*, page 753–758, 2021. ISBN 978-1-4503-8914-3. doi : <https://doi.org/10.1145/3447432.3447434>.
- Hui Yie Teh, Andreas W. Kempa-Liehr, and Kevin I-Kai Wang. Sensor data quality : a systematic review. *Journal of Big Data*, 7(11) :1–49, 2020. doi : 0.1186/s40537-020-0285-1.
- Richard Y. Wang and Diane M. Strong. Beyond accuracy : What data quality means to data consumers. *Journal of Management Information Systems*, 12(4) :5–33, 1996. doi : 10.1080/07421222.1996.11518099.
- Dietmar Winkler, Alexander Korobeinykov, Petr Novák, Arndt Lüder, and Stefan Biffl. Big data needs and challenges in smart manufacturing : an industry-academia survey. In *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, pages 1–8, 2021. doi : 10.1109/ETFA45728.2021.9613600.
- Tangxiao Yuan, Kondo Hloindo Adjallah, Alexandre Sava, Huifen Wang, and Linyan Liu. Issues of intelligent data acquisition and quality for manufacturing decision-support in an industry 4.0 context. In *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems : Technology and Applications (IDAACS)*, volume 2, pages 1200–1205, 2021.
- Li Zheng. *Improving performance measurement of engineering projects : methods to develop indicators*. PhD thesis, Computer Science [cs]. Institut national des sciences appliquées de Toulouse, 2018.
- Junhong Zhou, Yu Wang, and Yong Quan Chua. Machine oee monitoring and analysis for a complex manufacturing environment. In *15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 1413–1418, 2020. doi : 10.1109/ICIEA48937.2020.9248351.