

Groupe de Travail

QUALITÉ DES DONNÉES

cahier
pratique

n°
1

APPROCHES TACTIQUES ET STRATÉGIQUES SUR LA QUALITÉ DES DONNÉES



ISACA®

AFAI | PARIS-FRANCE CHAPTER



1^{ère} édition • Septembre 2021

Groupe de Travail

QUALITÉ DES DONNÉES

cahier
pratique

1
n°

APPROCHES TACTIQUES ET STRATÉGIQUES SUR LA QUALITÉ DES DONNÉES

1^{ère} édition • Septembre 2021

TABLE DES MATIÈRES

INTRODUCTION	p. 4
PRÉSENTATION de l'ISACA-AFAI et de DAMA France	p. 6
Liste des participants au Groupe de Travail DATA QUALITY	p. 9
1/ La qualité des données : qu'est-ce que c'est et qu'est-ce que ce n'est pas ?	p. 10
A/ Risques, coûts et bénéfices: définition des termes	p. 11
B/ Les risques, coûts et bénéfices associés à la qualité des données	p. 12
2/ Comment faciliter l'approbation d'un projet ou programme de Qualité des Données ?	p. 14
A/ Définition du ROI (Return On Investment)	p. 15
B/ Pourquoi calculer un ROI dans le cadre d'un projet de Qualité des Données	p. 15
C/ Comment calculer un ROI ?	p. 16
D/ Éléments intangibles en support au ROI pour l'approbation du Projet	p. 16
3/ Études de cas	p. 18
• Cas n°1 : mise en place d'une première démarche de qualité des données adossée au programme « client »	p. 18
• Cas n°2 : Approche pragmatique d'une revue globale de la qualité des données en R&D	p. 20
• Cas n°3 : comment engager une démarche globale sur la Data Quality avec une approche « lean »?	p. 22
• Cas n°4 : Les quatre leviers pour fiabiliser les données	p. 24
• Cas n°5 : la Data Quality comme levier pour optimiser le capital de la banque	p. 26
• Cas n°6 : la qualité des données chez Enedis : vers une démarche plus systématique	p. 28
4/ CONCLUSION	p. 32
• Alignement stratégique et Business case : Proposition de synthèse	p. 32

POURQUOI UN GROUPE DE TRAVAIL SUR LA QUALITÉ DES DONNÉES?

Pour la grande majorité des entreprises de taille significative, il ne fait aujourd'hui plus aucun doute que les **données** constituent un **actif majeur**.

Pour autant, de nombreuses initiatives autour des données rencontrent d'importantes difficultés lors de leur mise en œuvre, en particulier en lien avec la qualité des données.

Forts de ce constat, **DAMA France** et l'**ISACA-AFAI** ont recueilli de nombreuses demandes d'entreprises sur la gestion de la qualité des données, avec une relative faible maturité générale sur le sujet. Les deux associations se sont alors **rapprochées** pour **promouvoir et développer les avancées** sur ce sujet en constituant un **Groupe de Travail**.

Le **Groupe de Travail « Data Quality »**, à la composition diversifiée, a pour objet d'éclairer et de **fournir une aide concrète** aux entreprises sur le **sujet de la gestion de la qualité des données**, à travers des **livrables et des illustrations concrètes et pratiques**.

Les restitutions des travaux prendront des formes diverses, notamment des **cahiers pratiques** sur des thématiques ciblées.

Marcel Lee – Président DAMA France
Vincent Manière – Président ISACA-AFAI

APPROCHES TACTIQUES ET STRATÉGIQUES SUR LA QUALITÉ DES DONNÉES

La gestion de la **qualité des données** présente cette particularité d'induire des **moyens et des coûts bien visibles** dans les organisations (exemples: moyens humains, moyens technologiques), alors que les **impacts liés à la non-qualité de données sont souvent difficiles à évaluer**.

On peut citer par exemple les coûts de remédiation manuelle par les opérationnels, l'insatisfaction généralisée des consommateurs ou les impacts sur la qualité du service perçue par le client.

Cette **dualité défavorable** pour les praticiens de la qualité des données et globalement pour leurs organisations fait qu'il est souvent difficile d'engager des **initiatives sur la qualité des données**, de manière **tactique**, et à plus forte raison dans une approche **stratégique** et systématique.

Le **cahier pratique n°1** a pour ambition d'apporter un éclairage sur ces enjeux, selon une **méthode illustrative**, avec des **cas d'études concrets et réels d'entreprises** qui ont été confrontées à différents contextes et ont réussi à apporter une ou plusieurs solutions face à ces situations.

PRÉSENTATION DES ASSOCIATIONS



AFAI | PARIS-FRANCE CHAPTER

L'ISACA-AFAI est l'association de référence des métiers des systèmes d'information.

Elle est le chapitre français de l'ISACA, association internationale qui regroupe plus de **140 000 membres** dans **180 pays**, elle rassemble aujourd'hui près d'un millier de professionnels en France.

Sa vocation est d'aider les organisations à mettre le système d'information au service de leur stratégie et de leur développement, à travers trois axes :

- Contribuer à la réflexion, via des groupes de travail et des événements, sur les sujets clés de la gouvernance du numérique et de la sécurité des SI.
- Favoriser l'utilisation en France des référentiels et certifications de l'ISACA.
- Former les professionnels, en les préparant aux certifications professionnelles internationales délivrées par l'ISACA autour de la gouvernance et de la maîtrise des SI : COBIT, ITAF, Val IT, Risk IT, BMIS, mais aussi à la gouvernance et au management des SI.

L'ISACA-AFAI contribue ainsi à la création de valeur pour les métiers, à améliorer les compétences et la performance de l'informatique, ainsi qu'à maîtriser les risques liés au système d'information.

L'ISACA-AFAI membre du collectif Femmes@Numérique

Le 16 mai, à l'occasion de la conférence sur le sujet avec la participation du Cigref, membre fondateur de Femmes@Numériques, Bertrand Maguet, vice-président de l'ISACA-AFAI a signé la charte engageant l'association dans le collectif Femmes@Numérique. Déjà partie prenante dans SheLeadsTech, l'ISACA-AFAI réaffirme son engagement pour la féminisation des métiers du numérique, alors que la présence des femmes semble régresser dans ce secteur.

NOTRE MISSION VOUS AIDER À BÉNÉFICIER DE TOUTES LES POSSIBILITÉS DU NUMÉRIQUE
NOTRE PROMESSE LA CONFIANCE AU SERVICE DE L'INNOVATION TECHNOLOGIQUE

isaca-afai.fr



DAMA (Data Management Association) International est une association mondiale créée en 1980 à but non-lucratif, regroupant des professionnels du Data Management, dont de nombreux auteurs d'ouvrages techniques, conférenciers, et autres référents de l'industrie.

Le but de DAMA International est de promouvoir la compréhension, le développement, l'adoption et la pratique du Data Management au sein des organisations, qu'elles soient publiques, privées, ou associatives.

DAMA International se positionne comme une ressource essentielle pour les professionnels du Data Management et de la Gestion de l'Information, afin de les aider à approfondir leurs connaissances et compétences, notamment via

- la mise à jour et la publication du Data Management Body of Knowledge (DMBOK), ouvrage de référence du domaine,
- le développement d'un programme de certification «Certified Data Management Professional» (CDMP),
- la parution d'études de cas,
- le partage d'expériences et de bonnes pratiques entre professionnels,
- l'organisation de conférences dédiées au Data Management.

Aujourd'hui, l'association regroupe 70 chapitres locaux sur les cinq continents, et compte plus de 20 000 membres à travers le monde.

DAMA France, créé en 2014, est le chapitre français de l'association DAMA International. Comme DAMA International, DAMA France est une association à but non-lucratif, regroupant des professionnels du Data Management en France.

Les objectifs de DAMA France sont de :

- accompagner, favoriser l'acculturation et la montée en compétence des acteurs du Data Management en France,
- innover et participer à l'élaboration et promouvoir les meilleures pratiques du Data Management en France,
- fédérer les initiatives françaises dans le domaine du Data Management,
- créer des réseaux locaux de rencontres de professionnels du Data Management, susciter l'échange entre praticiens.

Aujourd'hui, DAMA France compte plus de 150 adhérents, dont plus de 40 entreprises, et :

- organise de manière trimestrielle un Club des Chief Data Officers des organisations françaises (plus de 30 sociétés membres),
- anime des webinars ou afterworks mensuels sur le Data Management,
- propose du contenu dédié au Data Management sur son site Internet.

dama-france.org



LISTE DES PARTICIPANTS AU GROUPE DE TRAVAIL DATA QUALITY

(par ordre alphabétique)

Antoine Mantelier – BPCE

Aude de Touchet – Enedis

Bernard Ruysen – AXA's Corporate Center

Christophe Bonnefoux – BNP Paribas Asset Management

Christophe Hesters – Safran

Florian Lambert – iQo

François-Benjamin Remazeilles – Nespresso, VP Organization & Logistics DAMA France

Eric Fonta – Mazars

Ghislaine Doukhan – Safran

Hélène Naudon – Groupe La Poste

Isabelle Pires – BNP Paribas Asset Management

Juliette Bechdolf – AXA's Corporate Center

Karina Razafimahenina - Etudiante Université Paris-Dauphine – PSL

Luc Aubin – iQo

Marcel Lee – Pramana, Président de DAMA France

Mathieu Minault – Société Générale, VP Marketing DAMA France

Olivier Dardenne – Safran

Olivier Tia Tiong Fat – Société Générale

Philippe Rideau – Société Générale

Sacha Koskas – Etudiant Université Paris-Dauphine – PSL

Sandra Holgado – Enedis

Sarah Benabdelhadi – Etudiante Université Paris-Dauphine – PSL

Simon Lebeau – Danone Nutricia Research

Thibaut Fasquel – Caisse des Dépôts

Valérie Houssin – Caisse des Dépôts

Vincent Givaudan – Pramana, VP Education DAMA France

Vincent Manière – VMA Conseils, Président de l'ISACA-AFAI

Virginie Cuvilliez - RSM France, Board Member – Education Chair ISACA-AFAI



LA QUALITÉ DES DONNÉES : QU'EST-CE QUE C'EST ET QU'EST-CE QUE CE N'EST PAS ?

La qualité des données est un terme qui se réfère à un ensemble de caractéristiques associées avec ce qu'on appelle une donnée de "bonne qualité" (DAMA-DMBOK). On considère ainsi qu'une donnée est de "bonne qualité" lorsqu'elle répond aux attentes et aux besoins de ses consommateurs. Elle est considérée de "mauvaise qualité" lorsqu'elle n'est pas adéquate aux usages qu'on veut en faire. La qualité des données est donc dépendante du contexte et des besoins des consommateurs de ces données (DAMA-DMBOK).

Si des données de "mauvaise qualité" engendrent des impacts négatifs, il conviendra d'envisager un programme ou un projet de mise en qualité des données. Quel type de retour sur investissement une entreprise peut-elle alors espérer en lançant un tel projet ou programme ?

A/ Risques, coûts et bénéfices: définition des termes

Risques: Un risque est souvent exprimé en termes de combinaison des conséquences d'un événement (incluant des changements de circonstances) et de sa vraisemblance.

ISO/IEC 73

Dans le cadre de la qualité des données et de notre étude, un risque représente l'effet négatif potentiel, d'une qualité moindre des données, sur un résultat escompté ou une situation. Ainsi lorsque l'on parle de risque en qualité des données, on fait référence aux risques de non-qualité.

Coûts : Le coût peut être défini comme "une ressource sacrifiée et/ou délaissée pour atteindre un objectif spécifique (ici un niveau de qualité de données), ou comme l'impact monétaire d'une action menée ou d'une décision prise, ou son absence."

M.J Eppler, Helfert M., 2004.

Bénéfices : Ce terme fait référence à la conséquence / au résultat d'une action dont la nature et la valeur (exprimée de différentes manières) sont considérées comme avantageuses pour l'entreprise.

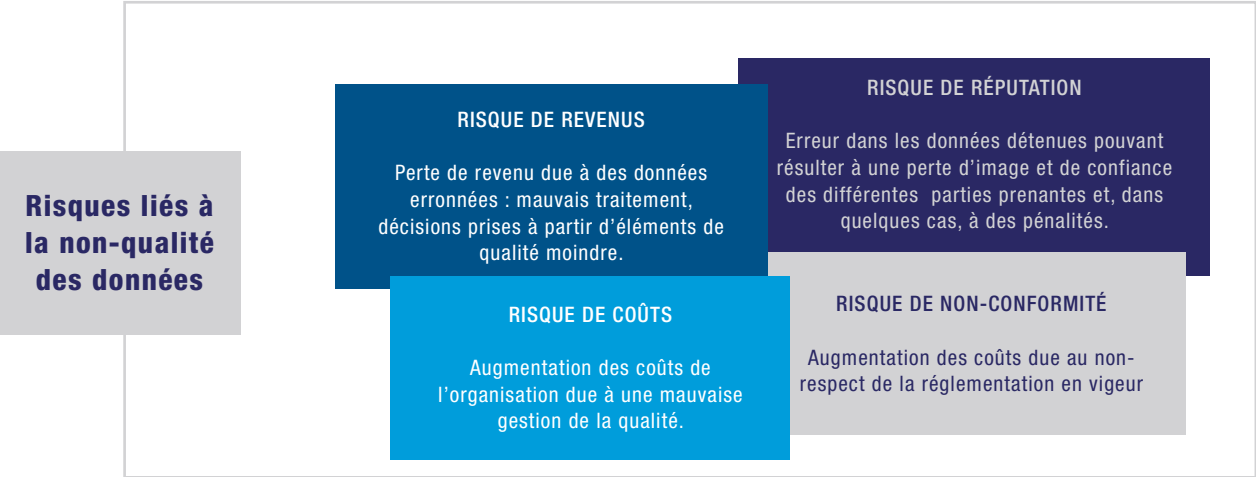
ISACA-AFAI

B/ Les risques, coûts et bénéfices associés à la qualité des données

L'évaluation des coûts et risques de la non-qualité des données n'est pas tâche facile dans la mesure où beaucoup de ces facteurs n'ont pas de liens immédiats avec la mauvaise qualité des données et ses effets (coûts indirects).

Dans le cadre de ce groupe de travail, et dans un souci de simplification, la convention suivante a été adoptée: les coûts sont considérés comme avérés tandis que les risques représentent les conséquences négatives non avérées de la non-qualité des données.

Les risques liés à la non-qualité des données:



Exemples non exhaustifs des différents risques identifiés:

- Risque de revenus**
- Impact stratégique: une insuffisance de qualité peut occasionner entre 15 et 25% de pertes du CA (Chiffre d’Affaires).
 - Impact sur les activités opérationnelles : une mauvaise qualité des données peut mener à des erreurs opérationnelles telles que la livraison de quantités / produits erronés aux clients, ou le rappel de produits défectueux moins efficaces induisant une perte de CA.

- Risque de réputation**
- Impact d’une erreur de gestion telle qu’une base de données clients non qualitative dans le cadre d’une campagne marketing, les messages distribués aux clients peuvent porter atteinte à l’image de l’organisation.

- Risque de coûts**
- Impact sur les activités opérationnelles : une mauvaise qualité des données peut mener à des erreurs opérationnelles telles le rappel de produits défectueux moins efficaces menant à des dédommagements - en plus des coûts de personnel associés à leur gestion.

- Risque de non-conformité**
- Impact du non-respect de la réglementation: une entreprise peut faire l’objet d’amendes, et même subir une perte ou l’arrêt momentané ou définitif d’une ou plusieurs activités

Coûts liés à la qualité des données

COÛT INDUITS PAR LA MAUVAISE QUALITÉ DES DONNÉES		COÛTS INDUITS PAR LA DÉMARCHÉ QUALITÉ	
Directs	Indirects	Récurrents	Mise en place
Vérification et retraitement	Ressources Humaines	Prévention	Réparation
Dédommagement	Renoncement	Maintenance	Projet
Stockage de données obsolètes		Détection	
Non-conformité (amendes, éléments fiscaux)			

Les bénéfices liés à la qualité des données:

Les bénéfices liés à la qualité des données peuvent être aussi bien tangibles qu’intangibles, et à effet immédiat ou direct, comme à plus long terme ou indirect. Tandis que les éléments quantitatifs représentent un élément tangible et convaincant dans l’élaboration d’un business case, la valeur qualitative des bénéfices reste néanmoins non-négligeable et à exposer, bien que son évaluation soit moins instinctive.

On note qu’une notion de probabilité peut s’ajouter aux bénéfices attendus qui dès lors s’apparentent plus à une opportunité pour l’organisation.

Enfin, les éléments de risque ou de coûts liés à la non-qualité des données décrits dans les paragraphes précédents sont autant d’éléments à considérer pour en faire potentiellement des bénéfices et exposer un projet ou un programme de mise en qualité plus attractif.

Bénéfices liés à la qualité des données

	LES BÉNÉFICES TANGIBLES	LES BÉNÉFICES INTANGIBLES
	Gains réalisés et réduction des coûts*	Satisfaction client
DIRECTS	Chiffres d’Affaire additionnel	Réduction des risques
		Optimisation des processus / Efficacité opérationnelle
		Conformité réglementaire
INDIRECTS	Coûts futurs évités**	Chiffre d’Affaire
		Performance IT
		Gouvernance

* Coûts directs
** Coûts indirects



PROGRAMME QUALITÉ ET ROI »

COMMENT FACILITER L'APPROBATION D'UN PROJET OU PROGRAMME DE QUALITÉ DES DONNÉES ?

A/ Définition du ROI (Return On Investment)

Le retour sur investissement (ROI) est une mesure de performance utilisée pour évaluer ou comparer l'efficacité et la rentabilité d'un ou plusieurs investissements par rapport à leur coût initial. Cette technique est essentiellement utilisée en gestion de projet permettant une estimation des bénéfices potentiels à tirer d'un investissement.

B/ Pourquoi calculer un ROI dans le cadre d'un projet de Qualité des Données

Le calcul du ROI est souvent demandé pour un projet de Qualité des Données. Pour autant, il n'est pas un élément toujours indispensable ni suffisant au bon engagement d'un projet de Qualité des Données, pour deux raisons principales :

- D'une part, parce que la justification d'un projet de Qualité des Données repose sur des estimations et interprétations qui peuvent varier d'une personne à l'autre, avec de surcroît de nombreux coûts et bénéfices indirects, comme beaucoup de facteurs de l'ordre de l'intangible;
- D'autre part, parce qu'en raison de cette complexité, tout calcul de ROI pourra être remis en question si les décideurs ne sont pas déjà convaincus du bien fondé de la démarche et ne sont pas alignés avec les enjeux sous-jacents exposés avec le projet de Qualité des Données.

Cependant, deux cas de figure peuvent justifier plus sereinement le recours au calcul du ROI pour un projet de Qualité des Données :

- Lorsque l'on doit répondre aux attentes du management requérant des données chiffrées et tangibles - auquel cas, nous nous accorderons sur les hypothèses d'estimation et de valorisation des coûts et bénéfices;
- Lorsque le projet concerne une activité métier très précise où la donnée et son usage sont connus
- Par exemple, sur la mise en qualité des dossiers clients dans une entreprise: si on observe que 30% des dossiers clients sont incomplets et que leur retraitement nécessite 10 min de consolidation en plus par dossier incomplet. Alors, le coût de la mauvaise qualité des données équivaut à ces 10 min (ETP et rémunération) multipliées par le nombre de dossiers incomplets sur un temps donné. Le ROI peut se calculer par [(coûts évités ETP – coûts mise en qualité) / coûts mise en qualité]



C/ Comment calculer un ROI ?

La formule “classique” du ROI est : (gain de l’investissement – coût de l’investissement) / coût de l’investissement.

Les variables de cette formule font partie des éléments liés aux coûts et aux bénéfices exposés ci-avant, et qui peuvent être difficiles à estimer dans le cadre d’un projet de Qualité des Données. La méthode préconisée est de se rapprocher au mieux d’une activité métier objectivement quantifiable avec des métriques simples à évaluer et opposables (ou a minima, sur lesquelles un consensus se dégage) pour proposer un ROI fiable.

Voici à titre indicatif les étapes courantes pour calculer un ROI au plus proche de la réalité :

- 1.Délimiter le périmètre (plus le périmètre sera large et diffus, plus le calcul sera complexe et non opposable)
- 2.Associer le défaut de qualité à des métriques pour en évaluer un bénéfice “en creux”
 - a.Volume de données
 - b.Nombre de collaborateurs impactés
 - c.Nombre de processus impactés
 - d.etc.
3. Estimer les bénéfices propres, non nécessairement induits par la “non-qualité”
4. Définir le coût de l’investissement
5. Calculer le ratio Gains de l’investissement / Bénéfices

D/ Éléments intangibles en support au ROI pour l’approbation du Projet

Le ROI étant un indicateur financier, il ne prend en compte que des intrants qui relèvent du tangible. Afin de faciliter l’approbation d’un projet de Qualité des Données, il existe d’autres éléments, d’une dimension plus globale, qu’il est important de considérer et d’exposer.

Certains inducteurs de risques ou de bénéfices non tangibles sont ainsi d’excellentes manières de mettre en avant des projets de mise en qualité de données.

On peut citer notamment :

- La réduction des risques
- Un meilleur positionnement dans l’écosystème
- La dotation d’un avantage concurrentiel
- L’amélioration des processus décisionnels
- Un meilleur time-to-market
- Le développement de nouveaux cas d’usages qui va permettre d’élargir le périmètre d’action du projet et donc les bénéfices escomptés

Le tableau suivant résume dans sa partie gauche les éléments chiffrables du ROI, et dans sa partie droite les éléments non chiffrables relevant de l’intangible.

ROI qualité des données				
	LES ÉLÉMENTS TANGIBLES		LES ÉLÉMENTS INTANGIBLES	
COÛTS	RÉPARTITION	PRÉVENTION		
	PROJET	MAINTENANCE		
		DÉTECTION		
BÉNÉFICES	GAINS RÉALISÉS ET RÉDUCTION DES COÛTS		RÉDUCTION DES RISQUES	
		Ressources humaines	Satisfaction client	Chiffre d’Affaire
		Vérification et retraitement	Gouvernance	Optimisation des processus / Efficacité opérationnelle
		Stockage de données obsolètes		Conformité réglementaire
	COÛTS futurs évités		Performance IT	
		Dédommagement		
	Chiffre d’affaire additionnel			
ROI	Différence des éléments tangibles : Coûts avérés		Risques résiduels	

DAMA - DMBOK. 2nd Edition. DAMA International. Technics Publications, 2017

ISACA AFAI

Eppler Martin J. Helfert M., 2004. *A framework for the classification of data quality costs and analysis of their progression*, *Proceedings of the Ninth International Conference on information quality*.

Panoho K., 2019. Council Post : *The age of analytics and the importance of data*, Forbes, Octobre 2019

Marcel, C., 2018. *Cost Of Quality*. [online] Altran.com.

> https://www.altran.com/as-content/uploads/sites/4/2018/02/altran_cost-of-quality-whitepaper.pdf

Thomas C. Redman, David Sammon & Tadhg Nagle, 2009. *Only 3% of data meets data basics Companies’s Data*, Harvard Business School, Septembre 2009

Thomas C. Redman, 2017. *Seizing Opportunity in Data Quality*, MIT Sloan, Novembre 2017

CAS N°1 : MISE EN PLACE D'UNE PREMIÈRE DÉMARCHE DE QUALITÉ DES DONNÉES ADOSSÉE AU PROGRAMME « CLIENT »

LE CONTEXTE D'ENTREPRISE

Ce groupe d'environ 6 000 collaborateurs reçoit des clients particuliers et entreprises dans une cinquantaine d'établissements, répartis sur ses 2 métiers (secteurs de l'hébergement, restauration, divertissement).

Avec le lancement d'un programme de digitalisation et l'accélération de l'utilisation des données clients, deux initiatives majeures ont été lancées en parallèle :

- La refonte de la base des clients uniques ;
- La mise en place d'un programme de fidélité avec points et passage de niveaux.

Ces initiatives représentent un investissement important, avec une attente forte en progression d'activité.

L'HISTORIQUE SUR LA DATA

Il n'y avait pas d'initiative labellisée « Qualité de données », mais des difficultés étaient identifiées sur la base de données clients, avec notamment :

- Un dédoublement par construction imparfait car les deux métiers ne collectent pas les mêmes données d'identification ;
- Des délais d'alimentation jugés longs et ne répondant plus aux attentes des consommateurs de la donnée (temps de passage des interfaces après la fin de la journée business, soit à l'issue de la nuit) ;
- Un manque de discipline dans la saisie des données à l'origine (exhaustivité, validité).

Par ailleurs, le système de fidélité reposait sur une chaîne totalement automatisée, depuis la génération d'une donnée captée sur une machine jusqu'au calcul des points de fidélité. Ce traitement bénéficiait donc d'une confiance a priori et d'une présomption d'exactitude qui n'appelait pas de contrôle particulier.

LES SOLUTIONS ET INITIATIVES MISES EN PLACE

Un plan d'actions a été construit et un dispositif a été proposé et accepté pour une durée d'un an, prolongée par la suite.

Un poste de Data Quality Manager a été ouvert au sein du Pôle Data de la DSI, pour orchestrer la démarche, autour des missions suivantes :

- Instruire les cas d'anomalie, en travaillant avec le métier consommateur essentiellement et les équipes de la DSI ;
- Proposer les corrections de données ;
- Proposer des modifications des SI ;
- Mettre en place une dynamique orientée contrôle avec une analyse des flux par les risques de non-qualité.

RÉSULTATS OBTENUS

Conformément à l'objectif premier de cette initiative orientée actions, une baisse des anomalies a pu être constatée et reconnue par les utilisateurs de la donnée.

D'autre part, un certain nombre de dispositifs ont pu être mis en œuvre, dont à titre d'exemples :

- des contrôles automatisés ;
- des états d'exception envoyés aux responsables des exploitations pour correction des saisies ;
- des reportings / classements des établissements sur la complétude des saisies.

Enfin, les problématiques liées à la qualité des données et les rôles et responsabilités des acteurs ont pu être partagés au sein de l'organisation.

Fort de ces résultats, ce dispositif a été reconduit à l'issue de la première année.



CAS N°2 : APPROCHE PRAGMATIQUE D'UNE REVUE GLOBALE DE LA QUALITÉ DES DONNÉES EN R&D

LE CONTEXTE D'ENTREPRISE

Danone Nutricia Research, ce sont près de 1500 hommes et femmes répartis sur 6 centres. Nous construisons chaque jour des ponts entre science et nutrition, tout en s'adaptant aux habitudes culturelles et alimentaires des différentes régions du monde et en s'engageant pour améliorer la qualité nutritionnelle et l'impact environnemental des produits afin d'apporter un monde en bonne santé.

L'HISTORIQUE SUR LA DATA

Dans la recherche scientifique, la génération de données reste un enjeu majeur, demandant du temps et des efforts financiers. Une étude clinique et des expérimentations en laboratoire peuvent produire quelques mégabits de données pour des millions d'euros, et ceci sur plusieurs années. C'est pourquoi il est important d'en assurer le maximum de valeur ajoutée, et de sécuriser certains aspects liés à la collecte de données. L'importance et la rigueur liées à cette dernière ont longtemps été sous-estimées et fut source de tension dans certains de nos projets. En conséquence, nous avons décidé de lancer une initiative de qualité de données sur le périmètre R&D de Danone Nutricia Research afin d'assurer la pérennité de nos données et de garantir robustesse et confiance dans les résultats des analyses effectuées.

Les sources sont multiples et en grand nombre du fait de leur nature et il était important d'évaluer leur qualité dans un premier temps avec une vue globale et quantifiable.

LES SOLUTIONS ET INITIATIVES MISES EN PLACE

Afin d'assurer une approche ciblée et pragmatique, nous avons mis en place une mesure afin d'effectuer une analyse de risques liés à la qualité sur nos sources de données provenant de plusieurs partenaires internes ou externes. Pour chaque type de source de données, un questionnaire est créé permettant une analyse de risque quantitative. Des questions techniques sur les méthodes utilisées, et d'autres généralistes sur les processus existants sont ciblées et scorées en fonction du degré d'occurrence du risque et de son impact pour chaque étape du cycle de vie de la donnée. Ces questions sont de types fermées et non libres. A chaque question, un score est obtenu en fonction de la réponse : plus le score est élevé, plus la réponse engendre un risque avec un fort impact ou/et avec une grande occurrence. Une pondération pour chaque question est aussi définie avec la même logique.

Basé sur le questionnaire, un interview est alors effectué par un data steward, avec les différents intervenants du cycle de vie de la donnée, pour chaque source qui obtiendra donc un score de risque final.

RÉSULTATS OBTENUS

Cette cartographie des risques nous a permis de faire remonter un certain nombre d'alertes pour chaque source qui a été un élément appuyant la demande de lancement d'une initiative qualité devant une équipe managériale. Cela a aussi permis d'inculquer une culture de la donnée dans les équipes et chez nos partenaires.

Une fois accordé, le plan de déploiement de cette initiative est défini en fonction du degré d'importance de chaque source supportant la stratégie groupe et du score de risque obtenu lors de l'audit à travers les questionnaires.



CAS N°3 : COMMENT ENGAGER UNE DÉMARCHE GLOBALE SUR LA DATA QUALITY AVEC UNE APPROCHE « LEAN »?

LE CONTEXTE D'ENTREPRISE

Avec plus de 3 000 collaborateurs et une présence dans 37 pays, BNP Paribas Asset Management (BNPP AM) est un des acteurs majeurs de la gestion d'actifs au niveau international et se démarque notamment par sa gamme de fonds intégrant des critères ESG(1). BNPP AM compte 10 fonctions métiers décentralisées et dispose d'un parc de plus de 200 applications informatiques.

En janvier 2018, face à des réglementations de plus en plus exigeantes à l'égard des données (MIFID, RGPD, BCBS-239, ...), BNPP AM a créé un Data Office. Depuis lors, des initiatives fortes ont été engagées pour mettre en place les fondations de Data Management au sein de l'organisation. L'approche sur les données est affirmée avec le sigle « DGU » : une donnée n'existe que si elle est Définie, Gouvernée et Utile. Le choix de la gouvernance s'est porté sur un modèle décentralisé au sein des métiers, qui ont bénéficié d'une acculturation soutenue à travers un social game, des formations et des guides opérationnels.

Début 2020, l'entreprise disposait de :

- « Data Assets » (50 données stratégiques), une cartographie globale des objets métiers (« Enterprise Information Model »), un dictionnaire de données qui liste plus de 6 500 données brutes décrites par 14 attributs (métadonnées);
- Une communauté Data de 520 membres actifs qui, aujourd'hui, ont en plus de leurs objectifs professionnels RH métiers, au moins un objectif Data;
- Un process de Data Management by design avec un assessment Data de tous les nouveaux projets informatiques.

Dans la continuité de ces initiatives fortes, la mise en qualité industrielle des données est le prochain challenge à relever.

L'HISTORIQUE SUR LA DATA

Les sujets de qualité de données étaient traités de manières réactive (traitement ponctuel d'incidents liés à la qualité de la donnée, réponse à des réglementations). Aussi, depuis 2020, le Data Office a engagé une démarche plus globale et continue sur la Data Quality. La première étape ayant consisté à recenser de manière factuelle les irritants Data Quality de l'organisation, puis à proposer une manière simple de les adresser.

LES SOLUTIONS ET INITIATIVES MISES EN PLACE

1. Construction d'un Data Quality "use case pipeline"

Des interviews ont été menées auprès de plusieurs équipes pour comprendre de manière transversale leurs irritants en matière de qualité des données et les conséquences de ces disqualités.

Le plus grand nombre d'irritants ont été recueillis auprès des fonctions transverses (conformité, contrôles et risques, finance) et pour les besoins de reporting car leur utilisation de la donnée se fait généralement en bout de chaîne et nécessite de croiser plusieurs sources de données. Puis un travail de simplification a été opéré intégrant l'identification des impacts, des parties prenantes (producteurs et consommateurs de la donnée ainsi que tous les intervenants dans le cycle de vie de la donnée) et, le cas échéant, les premiers éléments d'analyse des causes.

2. Mise en place d'un premier groupe de disqualité de données via des ateliers ad-hoc (pas de structure projet lourde)

L'approche adoptée était de monter des ateliers ad-hoc, sans structure projet lourde, avec comme objectifs de :

- Favoriser les échanges entre consommateurs et producteurs de la donnée et notamment de :
 - Faire prendre conscience aux producteurs des impacts des données erronées sur les utilisateurs finaux,
 - Permettre aux producteurs d'expliquer les difficultés qu'ils rencontrent à fournir une donnée correcte.
- Identifier les différents leviers pour améliorer la qualité de la donnée

Ces leviers peuvent prendre la forme de modification de processus, développements informatiques, clarification de responsabilité et gouvernance de la donnée, enrichissements des plans de contrôles nécessaires, ...

Le livrable final est de proposer un plan d'action conjoint (incluant solutions tactiques et long terme) et d'adresser ces actions aux équipes concernées.

Dans la continuité de ces initiatives fortes, la mise en qualité industrielle des données est le prochain challenge à relever.

RÉSULTATS OBTENUS ET PERSPECTIVES

Le coup d'essai a été transformé car il a su démontrer, sur la base de cas précis, qu'il était possible d'adresser des problèmes de disqualités de manière simple avec les avantages suivants :

- Disposer d'un plan d'action clair dans une organisation très décentralisée et ne nécessitant pas systématiquement l'intervention de l'IT
- Diminuer les coûts d'investigation par rapport à une task force traditionnelle
- Réduire les délais de résolution
- Atténuer le « bruit » autour de la disqualité de la donnée
- Identifier les besoins en formations opérationnelles, de change, et les personnes clé dans la résolution du use case

L'approche constitue en soi un excellent levier pour « pousser » un programme plus global sur la Data Quality, notamment pour améliorer les capacités de monitoring des données contrôlées.

C'est d'ailleurs la suite donnée à cette initiative qui va être complétée par la mise en place d'un outil de contrôle de qualité de données industrialisé et intégrant des process de remédiation au fil de l'eau adressés aux responsables métiers de la donnée.

Ce dispositif permettra par ailleurs de prendre en charge les besoins de contrôles de données qui auront été identifiés par les différents use cases.



CAS N°4 : LES QUATRE LEVIERS POUR FIABILISER LES DONNÉES

LE CONTEXTE D'ENTREPRISE

La Caisse Des Dépôts, pôle financier public, remplit des missions d'intérêt général en appui des politiques publiques. Gestionnaire de confiance pour le compte de tiers, investisseur et prêteur de long terme, elle facilite la réalisation de projets innovants, durables et solidaires, et participe au développement économique du pays.

Son intervention est organisée autour de 5 grands métiers regroupant 125 000 collaborateurs : les retraites et la formation professionnelle, les gestions d'actifs, le suivi des filiales et des participations, le financement des entreprises et l'accompagnement des acteurs territoriaux dans l'élaboration et le déploiement de leurs projets d'avenir.

En 2018, afin de servir au mieux les missions qui lui sont confiées, la Caisse Des Dépôts souhaite devenir « data-driven » au travers de son Plan de Transformation Numérique, c'est-à-dire d'évoluer en tant qu'organisation à même de capitaliser sur les données pour piloter et soutenir ses activités.

Ceci se traduit notamment par le déploiement d'un portefeuille de cas d'usage data définis, priorisés et pilotés par le Data Office – instance opérationnelle de gouvernance décentralisée dans le métier regroupant les acteurs métiers et SI.

LA PROBLÉMATIQUE

Certains des cas d'usage ont rencontré des problèmes d'industrialisation, freinant la transformation, en raison des difficultés suivantes :

- Un manque d'information sur les données (définition, transformation appliquée le long du cycle de vie, etc.) ne permettant pas de les contextualiser et d'estimer leur capacité à répondre aux attentes du cas d'usage ;
- Des responsabilités sur les données non allouées ou dispersées entre plusieurs acteurs ;
- Une qualité des données qui n'est pas actualisée mais constatée qualitativement et des actions de remédiation non pérennes.

LES SOLUTIONS ET INITIATIVES MISES EN PLACE

Un **cadre de gouvernance** est en cours de définition et de déploiement afin de garantir la maîtrise et l'exploitabilité des données en agissant notamment sur la qualité des données au travers de 4 leviers :

1. Organisation

Une organisation décentralisée dans les métiers, au plus près des enjeux opérationnels, et portée par le déploiement progressif de deux rôles clés :

- Le Data Owner est responsable de la gouvernance des données dans son domaine métier
- Le Data Steward assure le suivi, le contrôle et la mise en qualité des données

2. Normes

La définition de standards qui ont pour ambition d'être appliqués à l'échelle de l'organisation. Exemple : les dimensions de la qualité pour qualifier les mesures de qualité, qualification de donnée dite « clé », etc.

3. Processus

La mise en place d'un workflow entre les différentes parties prenantes permettant d'inscrire les travaux de qualité dans un cadre d'amélioration continue

4. Socle Technologique

Le déploiement d'un catalogue de données et d'un outil de data quality pour :

- Opérationnaliser le cadre de gouvernance
- Diminuer l'effort manuel, pérenniser les actions et ainsi renforcer l'engagement des acteurs
- Mettre à disposition les services (Profiling, Data Lineage, etc.) et livrables nécessaires (glossaire métier d'entreprise, Tableau de bord Qualité).

MISE EN ŒUVRE ET SUITE

Trois cas d'usage pris dans 3 différents métiers ont servi de pilote pour définir une première version du cadre de gouvernance et l'instancier. Ces cas d'usage ont été construits et choisis par opportunisme – ils reprennent des besoins en termes de gouvernance et de qualité des données de projets déjà lancés de construction de décisionnels.

Ceux-ci ont été menés de façon agile sur une période courte (3 mois) et sur un périmètre restreint de données (~ 150) afin de :

- Mettre en place un glossaire de données et un catalogue de données au travers de la définition d'un metamodel commun permettant de partager à tout collaborateur la connaissance sur ces données (le sens, les parties prenantes – en particulier le Data Owner et le(s) Data Steward(s), les niveaux de sécurité et de confidentialité, la taxonomie associée, etc.)
- Implémenter le profiling de données et des indicateurs de qualité alimentant un Tableau de bord ainsi qu'un système de « Change Request » permettant de remonter les problèmes de qualité et de rapprocher les différents acteurs.

Par la suite, il est prévu de :

- Présenter ce cadre de Gouvernance dans les différents Data Offices et mettre en place un véritable plan de communication
- Etendre le périmètre de données considéré par ces premiers cas d'usage
- Appliquer ce cadre de gouvernance à d'autres cas d'usage et le faire évoluer de façon itérative selon les besoins et problématiques rencontrés



CAS N°5 : LA DATA QUALITY COMME LEVIER POUR OPTIMISER LE CAPITAL DE LA BANQUE

LE CONTEXTE

Un Data Office qui entre dans une nouvelle phase

Au sein du Groupe Société Générale, la Ligne métier AFMO regroupe les activités de banque de détail en Afrique. La banque est présente sur ce continent depuis plus de 100 ans, soutenant notamment les PME locales et développant de nouvelles offres de paiement mobile. Dans le Bassin méditerranéen, le Groupe est présent notamment au Maroc (depuis 1913), en Algérie (1999) et en Tunisie (2002). Ce dispositif compte 665 agences et plus de 2 millions de clients. Société Générale est également la première banque internationale en Afrique de l'Ouest, avec une présence dans 16 pays avec des positions locales historiques fortes, notamment en Côte d'Ivoire, au Sénégal et au Cameroun. Sur l'année 2020, la zone affiche des encours de crédits à 6,6 milliards d'euros et 8,6 milliards d'euros de dépôts.

La filière Data est née sous l'impulsion de la réglementation BCBS 239, qui poussait les grands établissements bancaires à améliorer la gouvernance autour de la qualité des données et les systèmes d'information servant aux reportings des risques bancaires. Au total, 5 Data Offices comptant une vingtaine de collaborateurs sont répartis dans toute la zone (1 à Paris, 2 sur le Bassin Méditerranéen et 2 sur l'Afrique subsaharienne). Ainsi, sur plusieurs années, d'importants investissements auront été réalisés, pour déployer un entrepôt de données mutualisé alimenté par chacune des implantations africaines.

En 2019, l'enjeu réglementaire vis-à-vis de BCBS239, à l'origine de la création de la filière est rempli. Le dispositif de Data Quality est désormais opérationnel et traite les anomalies en mode run.

Mais le champ des anomalies à traiter est vaste. Dans une nouvelle phase, le Data Office fait donc évoluer sa pratique de la Data Quality pour mieux répondre aux besoins stratégiques du métier.

Quels enseignements pouvons-nous tirer de cette démarche ?

4 FACTEURS CLEFS DE SUCCÈS

1. Guider ses actions à l'aide de sujets stratégiques pour la banque

Le point de départ est la priorité stratégique de la banque autour de l'optimisation de son capital. Or une mauvaise intégration des données liées aux garanties augmente le besoin en capital.

2. Partir de l'existant, plutôt que de la feuille blanche

Plutôt que de démarrer un nouveau projet, un stream Data Quality fut ajouté à un projet existant de meilleure prise en compte des garanties, permettant de capitaliser sur la gouvernance déjà existante et les ressources du projet. La gouvernance Data s'enrichit progressivement, avec l'onboarding des différents acteurs clefs dans les réunions mensuelles de revue de Data Quality dédiées à chaque entité.

3. Être à l'écoute des différents intervenants de la chaîne pour mieux travailler ensemble

Pour résoudre les problèmes de qualité de données, il est important de développer l'adhésion de tous les acteurs, en particulier les producteurs de données placés en début de chaîne. Une bonne implication passe par une écoute active, le souci d'expliquer le contexte, et faire le lien entre leurs actions au quotidien et l'impact sur les systèmes aval. Des documents pédagogiques furent réalisés et présentés au fil de l'eau aux nouveaux participants rejoignant les ateliers de travail de Data Quality.

4. Capitaliser sur toutes les compétences du Data Office pour une qualité durable

Pour une qualité des données qui s'améliore dans le temps, il faut capitaliser sur toutes les fonctions du Data Office: Data Architecture (décrire finement les chaînes applicatives et lancer des projets d'amélioration), Data Design (documenter les transformations, corriger les mappings de données), Business Intelligence / Data Science (développer des indicateurs avec Python et PowerBI partagés mensuellement avec le métier).

QUEL BILAN ?

Cette démarche pragmatique a permis à la filière Data d'apporter une contribution à un volet stratégique pour la banque, avec un baisse du nombre de corrections hors systèmes et des systèmes d'information améliorés.

La gouvernance en place a été enrichie, avec une plus grande participation aux comités mensuels. Le Data Office a pu démontrer son savoir-faire, ce qui amène le métier à émettre de nouvelles demandes.



CAS N°6 : LA QUALITÉ DES DONNÉES CHEZ ENEDIS : VERS UNE DÉMARCHE PLUS SYSTÉMATIQUE

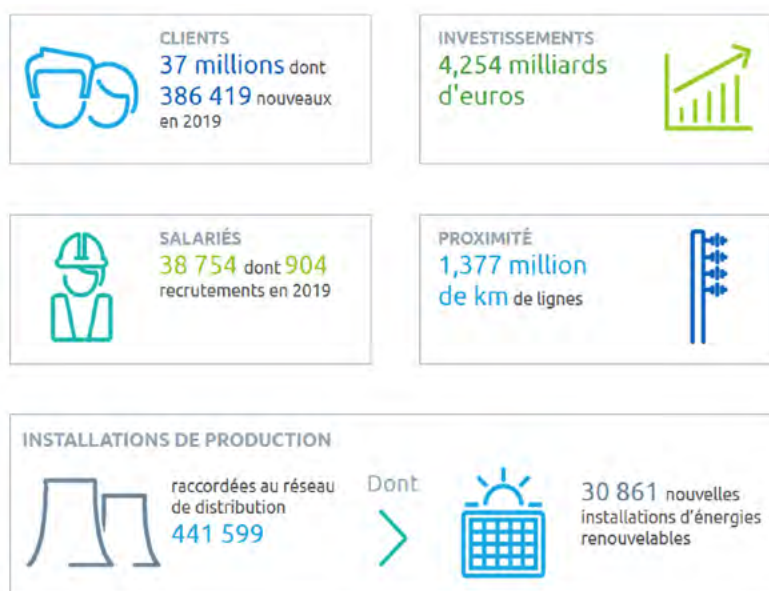
ENEDIS, GESTIONNAIRE DE RÉSEAU PUBLIC

DE DISTRIBUTION D'ÉLECTRICITÉ

MAJORITAIRE EN FRANCE

Depuis sa création en 2008, Enedis est le gestionnaire du réseau public de distribution d'électricité sur 95 % du territoire français continental. L'entreprise de service public a pour principale mission de développer et d'exploiter le réseau de distribution (situé entre le réseau de transport et les 37 millions de clients finaux). Elle gère équitablement le réseau de distribution d'électricité et réalise toutes les interventions techniques, raccordement, dépannage, relevé de compteur... grâce à ses 38 700 salariés.

CHIFFRES CLÉS



ENEDIS, UN OPÉRATEUR DE DONNÉES

AU SERVICE DE LA TRANSITION

ÉCOLOGIQUE

Enedis bénéficie d'un positionnement unique pour développer son rôle de gestionnaire de système de distribution au service de la transition écologique, en s'appuyant pleinement sur une activité d'opérateur de données, avec un double enjeu de protection et d'ouverture des données. Parmi celles-ci figurent entre autres les données précises de production et de consommation d'électricité collectées notamment grâce aux compteurs communicants comme Linky. Ces données contribuent à mieux piloter le réseau et à y insérer les énergies renouvelables, à accompagner les clients dans l'optimisation de leurs consommations, à faire mieux connaître aux collectivités le potentiel énergétique de leur territoire, à développer l'autoconsommation, à lutter contre la précarité énergétique ou bien encore à favoriser les nouveaux usages tels que la mobilité électrique, les flexibilités, etc.

Cette évolution implique pour l'Entreprise et ses agents de changer la manière de considérer et gérer les données : la donnée est dorénavant considérée comme un actif à part entière, avec ses implications en termes d'obligations, de gouvernance, d'organisation et de moyens.

UN CADRE DE GOUVERNANCE

DE LA DONNÉE DÉFINI EN 2017

Un des actes forts qui montre qu'Enedis a pris conscience de ces enjeux et s'est engagée de manière volontaire dans la transformation numérique, a été la signature du premier cadre de gouvernance de la donnée en 2017 afin de structurer, homogénéiser et coordonner les activités de gestion de la donnée en accord avec la stratégie de l'Entreprise.

Cette gouvernance repose sur quatre piliers :

- un référentiel (ou catalogue) de données et de services de données,
- des politiques, règles et procédures,
- des rôles et responsabilités,
- des mécanismes de suivi et de contrôles.

ENEDIS, D'UNE DÉMARCHE OPPORTUNISTE À UNE APPROCHE SYSTÉMATIQUE : METTRE EN PLACE UNE GOUVERNANCE DE LA QUALITÉ DES DONNÉES

De nombreux projets qualité des données dans les métiers et les systèmes d'informations (SI)

La qualité des données est un sujet que tous les métiers d'Enedis se sont appropriés, à plus ou moins grande échelle, en fonction de leurs besoins et de leurs contraintes. Les initiatives peuvent aller du mode opératoire de mise à jour des coordonnées clients à des projets ambitieux de traçabilité du matériel installé.

La plupart des SI disposent de processus de contrôle de qualité des données et d'indicateurs.

Mais malgré tout, un certain nombre d'irritants demeurent.

En effet, les projets de qualité des données sont souvent mis en place après un incident ou une prise de conscience d'une faiblesse, par exemple réglementaire. En découlent des actions ponctuelles, par réaction, avec une méthodologie réinventée à chaque fois et une faible capitalisation. La qualité des données d'Enedis ne faisait pas l'objet d'un suivi global et d'une priorisation transverse à l'Entreprise.

Fort de ce constat, le département Stratégie et Gouvernance des Données de la Direction du Numérique, a donc été chargé de définir une ambition d'Entreprise qui permettrait d'évaluer les périmètres à risques, de définir des règles communes, de mutualiser méthodes, outils et indicateurs, mais aussi de mieux intégrer la qualité « by design » dans les projets, qu'ils soient SI ou métiers.

DEUX APPROCHES MÉTHODOLOGIQUES ENVISAGÉES POUR UNE VISION COMMUNE

Plusieurs scénarios ont été envisagés pour finalement retenir deux approches : une première portée par le plus haut niveau de l'Entreprise et une seconde se basant plutôt sur les retours d'expériences terrain.

1. Piloter le suivi de la qualité des données essentielles de l'Entreprise

Le premier scénario consiste à identifier et piloter la qualité des données clefs de l'Entreprise. Il s'agit d'élaborer une liste réduite de données clefs (plus ou moins une vingtaine), de leur associer un niveau de qualité juste, de faire valider le tout par le Comité Exécutif et de suivre tout cela régulièrement lors de ce Comité. L'avantage de cette méthode est double : bénéficier d'un portage stratégique de la qualité des données et sécuriser les données essentielles grâce à un livrable simple et opérationnel.

Le criblage des données essentielles s'appuie à la fois sur le catalogue des données en place et les responsabilités associées, sur l'expérience des experts et l'avis des dirigeants.

Ainsi, en associant, vision objective et subjective, technique et métier, Enedis pense atteindre toutes les cibles ayant un œil averti à ce sujet et permettant d'établir cette liste.

2. De cas concret(s) à une politique

Une seconde approche vise, d'une part, à agréger les retours d'expériences (REX) de projets qualité des données passés ou en cours dans l'Entreprise, afin de synthétiser les leviers et les écueils, de compiler les méthodes employées, de partager les outils et indicateurs, et d'autre part, à mettre en relation les experts, voire de dégager, lorsque cela est possible et pertinent, des pistes de mutualisation. La combinaison de ces REX et du recueil des irritants et besoins qui demeurent permet de définir les principes d'une politique opérationnelle et utile aux services des métiers et des régions.

3. Le choix de combiner les deux approches pour porter une vision globale

Fin 2020, il a été décidé de mixer les deux approches pour en maximiser les bénéfices et créer un groupe de travail (GT) représentant chaque direction nationale et quelques directions régionales. Cette méthodologie a été approuvée par le Comité de Gouvernance des données d'Enedis.

Il est prévu que le premier semestre 2021 soit dédié à la collecte d'informations, se basant sur des résultats d'enquêtes, des interviews de grands projets métiers nationaux et d'experts identifiés grâce aux membres du GT. Le second semestre devrait permettre de consolider les résultats et de proposer la liste des données clefs avec ses indicateurs au COMEX. S'en suivra l'élaboration de cadres méthodologiques, voire de traitement d'irritants repérés.

4. Conclusion

Enedis compte sur cette approche duale et sur le niveau de maturité de l'Entreprise dans sa gouvernance de la data pour aboutir à des livrables et résultats concrets, utiles et applicables. Le succès de la démarche sera notamment grandement conditionné par l'appropriation, l'engagement et la collaboration de tous les acteurs de l'entreprise.



ALIGNEMENT STRATÉGIQUE ET BUSINESS CASE : PROPOSITION DE SYNTHÈSE

À l'aune des éléments exposés ci-dessus, à savoir d'une part, la théorie (le triptyque coûts - bénéfices - risques entrant dans un calcul de ROI), et la pratique (à travers les cas d'étude réels exposés), que pouvons-nous conclure quant à l'approche à adopter pour démarrer des initiatives qualité, ou bien engager une démarche plus systématique et pilotée ?

1.L'estimation formelle d'un ROI sur la qualité des données reste compliquée

Contrairement à un projet classique de création de produit ou service, la qualité des données a cette particularité que la partie « bénéfices » est souvent difficile à estimer.

Les bénéfices comportent en effet :

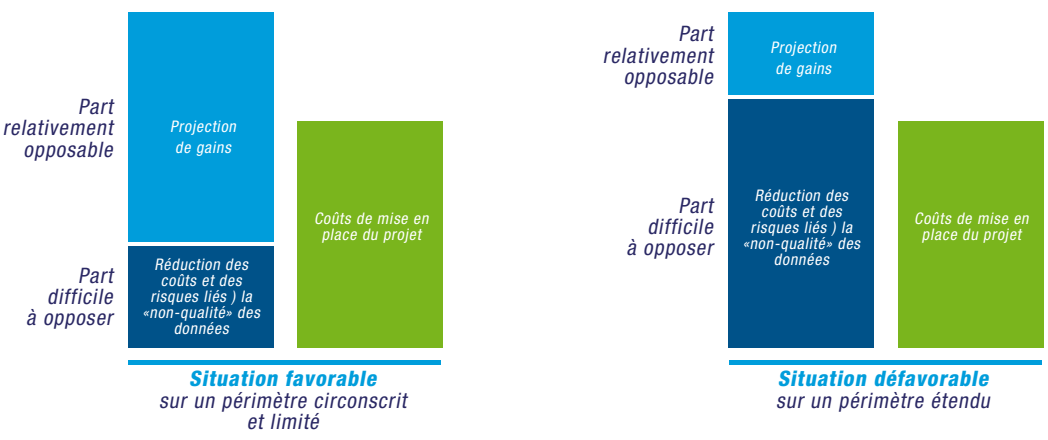
- Une part en lien avec des projections de gains
- Une part issue de la réduction des coûts et des risques liés à l'impact actuel de la « non-qualité » des données

La première part est « classique » et alignée à ce qu'on peut observer pour tout projet de mise en place d'un produit ou d'une application. Par exemple, on fera une projection de gain de CA réalisé grâce à une meilleure efficacité des campagnes marketing en liaison avec des données de meilleure qualité.

La seconde part est parfois très difficile à estimer en fonction du périmètre (ex : temps perdu lié à la remédiation locale des données, ou risque de mauvaise prise de décision, ou encore perte de confiance du client), et nécessite de prendre des hypothèses de calcul souvent non opposables.

Ainsi, le ROI présentera la plupart du temps un coût bien identifié, et des bénéfices difficiles à justifier.

On note par ailleurs que ce déséquilibre sera d'autant plus important que le périmètre sera étendu, tant en termes de données qu'au niveau organisationnel.



Nota Bene : la situation décrite à droite du schéma (situation défavorable) peut aussi parfois s'observer dans des périmètres circonscrits et limités.

La conséquence de cette dualité est que **tout ROI risque d'être contesté si l'interlocuteur n'est pas déjà convaincu par le bien-fondé de l'initiative.**

C'est la raison pour laquelle nous pouvons voir qu'aucune des études de cas décrites ne fait appel formellement à un calcul du ROI, ou du moins, de manière exclusive.

Ceci étant dit, le calcul du ROI reste une bonne pratique et un exercice intéressant et souvent exigé, notamment parce qu'il permet de se poser les bonnes questions quant à l'estimation des bénéfices (liés aux gains comme aux impacts de « non-qualité ») et qu'il va faciliter l'exposition de la situation actuelle et de ses impacts aux décideurs.

2. La prise de conscience collective et la conviction : premier déclencheur des initiatives qualité

Le ROI étant une piste a priori souvent insuffisante pour justifier d'une initiative sur la qualité des données, que reste-t-il comme approches possibles ?

Les cas d'études exposés le montrent de manière unanime, la clé réside d'abord dans l'instauration d'une culture de la « qualité des données », à **savoir la prise de conscience mutuelle** et collective, et le **développement d'une conviction**. Ceci peut se faire à plusieurs niveaux :

- **Autour d'un projet**, entre les acteurs de ce projet, en particulier entre les bénéficiaires métiers, la maîtrise d'ouvrage et la maîtrise d'œuvre (ou équivalent si en mode agile). Dans ce cas, le point clé consiste à identifier de manière précise les problèmes opérationnels engendrés par la non-qualité, les risques que cela fait courir au projet et la réalisation des bénéfices escomptés, et enfin, de bien délimiter les causes racines à ces problèmes. La limite de l'exercice étant que certains agents de la non-qualité seront hors périmètre du projet et ne pourront être adressés dans le cadre circonscrit ;

- **Autour des acteurs du cycle de vie de la donnée** : en termes de qualité des données, on observe souvent que les consommateurs de la donnée ont peu d'opportunités pour exprimer une insatisfaction sur la qualité et les conséquences que cela engendre. Par ailleurs, même si les problèmes sont exprimés, ils restent souvent étrangers aux producteurs de la donnée ou bien aux acteurs influant sur le niveau de qualité de la donnée – ceux-ci étant eux-mêmes non conscients des risques qu'ils engendrent par leurs actions. La recommandation qui peut être faite est d'abord de donner une voix aux consommateurs de données afin qu'ils puissent exprimer les impacts de la non-qualité – soit par des interviews, ou encore soit par un questionnaire mettant en lumière les impacts et les risques par exemple. Ensuite, il convient de provoquer le dialogue entre les producteurs de la donnée, tous les acteurs influant sur la qualité le long du cycle de vie de la donnée et les différents consommateurs, ceci afin de créer une conscience individuelle de l'autre, et donc une conscience collective du problème. De cette manière pourra émerger une collaboration entre les parties prenantes et il sera alors plus facile de justifier l'initiative à plus haut niveau.

- **Autour d'initiatives passées mal ou insuffisamment abouties** : beaucoup de projets de développement de nouveaux services ou produits, ou encore d'industrialisation de cas d'usage de Data Science n'aboutissent pas, ou de manière insatisfaisante en raison de la non-qualité des données. Et force est de constater que même une action « coup de poing » est insuffisante dans beaucoup de cas. Cette frustration, si elle est exploitée de manière intelligente, à savoir en mettant en lumière le sujet de non-qualité et ses causes racines, peut se transformer en un puissant vecteur pour lancer des initiatives de gestion de qualité, et même de manière systématique et pilotée.

- **Autour de données avec des finalités critiques** : dans toute entreprise ou organisation, certaines données revêtent un caractère plus important que d'autres : on peut les appeler données clés ou critiques. Mais si ces données sont plus importantes que d'autres, c'est parce qu'elles sont destinées à un usage ou plusieurs usages particuliers, souvent en exposition à des décideurs (niveau COMEX) ou des parties prenantes externes (ex : régulateur). La clé dans ce cas est non pas de juste mesurer la qualité des données clés élémentaires (souvent granulaires) et de les exposer, mais de mesurer la qualité des données résultantes de l'usage de ces données clés élémentaires. On pourra par exemple mesurer et exposer la qualité des KPI suivis régulièrement en COMEX. Cette mesure et cette exposition engendreront alors soit un confort, soit une appréhension, voire une inquiétude par rapport à la confiance que les parties prenantes peuvent avoir dans les chiffres, ce qui va déclencher cette prise de conscience recherchée et potentiellement le lancement d'initiatives qualité.

3. S'adosser à des initiatives existantes ou lancer une démarche dédiée : une question de maturité

Les différents cas d'étude le montrent : il existe une grande diversité de contextes et de situations, et on peut tantôt observer des initiatives qualité lancées de manière tactique, et tantôt à un niveau plus systématique et stratégique.

Lorsque l'entreprise est relativement peu mature, l'approche la plus réaliste est de s'adosser à un projet, en mettant en évidence la nécessité d'adresser un pan de travail dédié à la qualité des données. Le financement de ce pan est alors relativement simple car tout simplement pris en charge par le projet : le business case de l'initiative qualité est souvent implicite et embarqué dans le business case global du projet.

Vient ensuite le cas des initiatives dédiées mais non adossées à des projets – qu'on va appeler cas d'usage qualité. Les études de cas montrent qu'il est possible de le faire, pourvu que la prise de conscience soit au rendez-vous et que le ou les métiers bénéficiaires soient bien convaincus des bénéfices qui peuvent en être tirés. Dans ce cas, l'initiative sera certainement financée en majeure partie par le ou les métiers bénéficiaires du cas d'usage, avec le support d'acteurs transverses et à différents maillons de la chaîne de production / consommation de la donnée. On observe ce

genre de cas d'usage qualité dans les organisations un peu plus matures, car une structure à minima de gouvernance doit souvent être en place pour pouvoir engager des ressources sur les maillons « amont » du cycle de vie de la donnée.

Enfin, pour finir, comment aller plus loin et engager une démarche plus stratégique et systématique ?

La question est loin d'être évidente, car on observe souvent des entreprises où malgré la multiplication des cas d'usage qualité, le sujet est loin d'avoir une portée globale, avec une systématisation des règles et standards, un pilotage systématique, et des règles d'engagement claires. La raison principale est soit que l'entreprise n'est pas encore assez mature, soit elle n'a pas conscience en quoi avoir une démarche plus stratégique et systématique peut lui être bénéfique. Car après tout, elle observe que les problèmes qualité sont résolus au fur et à mesure : dans ce cas, pourquoi faire l'effort supplémentaire ? Encore une fois, la clé réside dans la prise de conscience, cette fois-ci un peu moins collective, mais plutôt au plus haut niveau dans l'entreprise. Gouverner la qualité des données n'est pas la même chose que la gérer. Elle recouvre une dimension supplémentaire associée à l'alignement avec la stratégie de l'entreprise, la mise en place de priorités bien établies, l'allocation de ressources optimisées et mutualisées, et un pilotage au global. De ce fait, lorsque la qualité des données est gouvernée, et non simplement gérée, l'entreprise peut s'attendre à une meilleure efficacité et efficience dans l'ensemble des activités de gestion de qualité. Il conviendra d'évangéliser sur ces aspects auprès des décideurs pour parvenir à ce stade « supérieur ».

Groupe de Travail

QUALITÉ DES DONNÉES

cahier
pratique

n°
1

APPROCHES TACTIQUES ET STRATÉGIQUES SUR LA QUALITÉ DES DONNÉES



DAMA
FRANCE



ISACA

AFAI | PARIS-FRANCE CHAPTER



Groupe de Travail

QUALITÉ DES DONNÉES

cahier
pratique

n°
1

**APPROCHES TACTIQUES
ET STRATÉGIQUES SUR
LA QUALITÉ DES DONNÉES**



Première édition Publiée à Paris le 20 septembre 2021
par ISACA-AFAI & DAMA France

Direction éditoriale : Virginie Cuvilliez, Marcel Lee, Vincent Manière et Mathieu Minault
Conception graphique : Agence Imedia Conseil
Crédits photographiques Shutterstock et AdobeStock

© 2021 - Tous droits réservés

Groupe de Travail

QUALITÉ DES DONNÉES



ISACA®

AFAI | PARIS-FRANCE CHAPTER

DAMA
FRANCE