Qualité de l'information

Thierry Bontems, Sabine Goulin

Qualité de l'information

Thierry BONTEMS
Responsable qualité
UMR 5194 - PACTE
Centre National de la Recherche Scientifique – CNRS - Grenoble
thierry.bontems@umrpacte.fr

Sabine GOULIN
Directrice –
Délégation à l'Aide au Pilotage et à la Qualité
DAPEQ - Nancy
Université de Lorraine
sabine.goulin@univ-lorraine.fr

ABSTRACT –

La qualité de l'information est la clef de voûte de tout système d'aide à la décision, hors actuellement encore trop peu d'informations répondent à des critères de qualité.
L'objet de cet article est de proposer la mise en évidence d'un certain nombre de facteurs qualifiant la notion même d'information, ainsi que des dimensions permettant de définir la qualité de cette information. En complément, nous identifierons des modes de gouvernance permettant de garantir un niveau de « qualité de l'information » acceptable et les pistes d'améliorations associées.

Index terms
Qualité de l'information, Système d'information, pilotage, données, référentiel, interopérabilité, gouvernance

1. INTRODUCTION

Outil indispensable du pilotage de laboratoire de recherche ou de leur tutelle de rattachement, la qualité des données n'en est pas pour autant un enjeu majeur. Hors de cette qualité des données, dépend la prise de décision des actions de pilotages engagés. Dans cet article, nous mettrons en évidence les facteurs qui définissent l'information ainsi que les dimensions à explorer pour définir la « qualité de l'information » enfin nous proposerons des modes de gouvernance associés afin de garantir cette qualité de l'information.

2. QUATRE FACTEURS QUI QUALIFIENT L'INFORMATION[i]

La qualité des données repose à la fois sur des caractéristiques (complètes, fiables, pertinentes et à jour) mais aussi sur l'ensemble des processus qui permet de garantir ces caractéristiques. Le but est d'obtenir des données non doublonnées, sans omissions, exactes, sans variations superflues et conforment à la structure définie.
Comme dans toute démarche qualité, l'objectif de la démarche demeure la satisfaction du besoin du client. Les données sont dites de qualités si elles satisfont aux exigences de leurs utilisations. En d'autres termes, la qualité des données dépend autant de leur utilisation que de leur état. Pour certain une donnée de qualité ne comportera ni fautes d'orthographes, ni erreurs grammaticales, pour d'autre, ces mêmes erreurs permettrons de qualifier l'exactitude de la restitution d'un entretien. Pour satisfaire à l'utilisation prévue, les données doivent être exactes, opportunes, pertinentes, complètes, compréhensibles et dignes de confiance. Nous pouvons illustrer ce propos en déterminant 4 facteurs définissant cette information.

2.1 Accessibilité.

Le rôle des systèmes d'informations est prépondérant dans la gestion de l'information. Dit comme cela, ça parait être une évidence et pourtant les rôles et les responsabilités inférant aux SI sont souvent mal considérés.
Afin de garantir des données de qualité, il faut veiller à pourvoir garantir d'une part la **disponibilité des systèmes**, c'est-à-dire garantir que les éléments considérés sont accessibles au moment voulu, d'autre part, garantir la **disponibilité des transactions**, c'est-à-dire garantir la disponibilité des connexions et des flux entre les différents systèmes permettant l'exécution de requêtes entre eux. Enfin, garantir la mise en place des **privilèges et droits d'accès** permettant aux personnes autorisées et uniquement à ces personnes une possibilité d'accès à l'information recherchée.
Si effectivement l'accessibilité est capitale, l'utilité n'en est pas moins un facteur indispensable à l'usage de l'information

2.2 Utilité

La quantité d'informations disponibles aujourd'hui est déjà colossale et croît de manière exponentielle. Une information de qualité est une information qui se veut utile, à savoir tout d'abord issue d'une donnée **non-dupliqué**. Les données sont gérées au sein de plusieurs bases de données (Recherche, Formation, Finance, Ressources Humaines…), par plusieurs systèmes d'informations sous des identifiants différents, et donc sa vue n'est pas unifiée. L'**usage** des données est aussi important pour définir l'information dont on aura besoin. En effet cet usage permettra de prévoir quels supports et quels formats utiliser.
On entend par **opportunité** le fait que les données sont à jour au moment de leur utilisation. Cette propriété nécessite de travailler de façon permanente

sur l'actualisation des sources, l'actualisation des cibles ainsi que sur la volatilité des éléments qui composent l'information.
Troisième facteur qui permet de définir l'information, la crédibilité.

2.3 Crédibilité

Afin d'être qualifié de crédible, les données doivent être standardisées, c'est-à-dire permettre aux données de rentrer dans un cadre de codification et de vérifier que les valeurs sont correctes par rapport à un intervalle de répartition ou par rapport à un domaine. L'objectif est d'éviter qu'une information comme un nom de laboratoire soit codée dans un système d'information UMR4130 dans un autre système A4130, dans un troisième LBI et dans un quatrième laboratoire de biologie intracellulaire. La standardisation soit par modification de la métadonnée soit par la mise en œuvre d'EDI[ii] a pour vocation de garantir l'unicité de l'information. Cette propriété doit être en adéquation avec le besoin du client.
Autre fondamental, l'information doit être exacte. **L'exactitude** garantie que les données représentent la réalité, ou sont vérifiables à partir d'une source externe comme par exemple le code postal qui doit correspondre à la localité.
Les données doivent être **cohérentes**. La cohérence est la capacité pour un système à refléter sur la copie d'une donnée les modifications intervenues sur d'autres copies de cette donnée. Concrètement, si une donnée *d* est écrite sur un système *A* puis dupliqué sur un système *B* et *C*, toute modification de *d* sur *A,B* ou *C* fera l'objet d'une mise à jour sur les deux autres systèmes. Aujourd'hui la vérification de cohérence fait partie des fondamentaux de la gestion d'un système d'information.
Enfin, toutes les données nécessaires au besoin du pilotage sont à sa disposition, c'est-à-dire l'ensemble des données et des métadonnées[iii] toutes complétées. On parlera alors d**'intégralité des données,**

2.4 Interprétabilité

Une donnée doit être stockée et affichée dans un format sans ambiguïté et cohérent. On peut montrer comme exemple le problème d'Interprétabilité soulevé par le format de date entre Paris et New York. Une date affiché sur un écran français 12/11/20102 (jour / mois / année) doit être affiché chez son homologue américain 11/12/2012 (mois / jour / année). Cette notion d'Interprétabilité peut être à l'origine de nombreux problèmes liés aux traitements des données.
Les contrôles syntaxiques permettent de vérifier que les données saisies dans une zone prévue à cet effet respectent une syntaxe prédéfinie : la syntaxe porte sur le format des données saisies, l'organisation des éléments d'un champ, les séparateurs entre éléments, etc. Par exemple, on pourra vérifier qu'un champ DATE est bien de la forme AAAA, ou que les éléments d'un champ multivalué[iv] sont séparés par un « ; », etc.
Pour garantir l'interprétabilité des données, il faut tout d'abord garantir la **syntaxe** de ces données, c'est-à-dire l'ensemble des règles qui concourent à l'écriture des données. La norme ISO2382[v] à d'ailleurs pour objet de faciliter les échanges internationaux dans les systèmes de traitement de l'information. A cet effet, elle présente un ensemble bilingue de termes et de définitions ayant trait à des notions choisies dans ce domaine, et définit les relations pouvant exister entre les différentes notions. Les définitions ont été établies de manière à éviter les particularismes propres à une langue donnée, en vue de faciliter leur transposition dans les langues autres que celles ayant servi à la rédaction initiale.
Une fois la syntaxe normalisée, on se concentrera sur la notion de sémantique. En effet, la **sémantique** joue un rôle essentiel. On s'intéressera alors au sens de la donnée en complémentarité à la syntaxe. On peut dire qu'il y a la même différence entre syntaxe et sémantique qu'entre fond et forme. Le décryptage sémantique des données permet une optimisation et une efficacité accrue des recherches d'informations opérées par un utilisateur. La valeur ajoutée repose sur un mode de requête qui écarte les informations parasites (le bruit documentaire) et réduit simultanément les silences documentaires (les informations pertinentes existantes mais non rapportées). Les contrôles de structure permettent de vérifier que le document produit respecte la structuration de l'information définie dans les normes de description (présence d'un champ défini comme obligatoire, non redondance de champs, ordre des champs, ...). Autre élément essentiel à l'interprétabilité des données, le **contrôle des versions** d'une information, à la fois en tant que tel mais aussi dans sa diffusion. Les contrôles d'unicité permettent de s'assurer qu'une valeur permettant d'identifier de manière non équivoque un élément d'un corpus n'apparaît qu'une seule fois dans ce corpus (identifiant unique d'une notice, d'un dossier ou d'une image, par exemple). L'identifiant doit en outre répondre à une règle de nommage, précisée dans les *Systèmes descriptifs* correspondants.
Les nouvelles technologies de l'information et de la communication comme le cloud computing[vi] ainsi que tous les outils de connexion à l'information en mode nomade permettent aujourd'hui d'accéder à l'information en mode partagé, à distance en conservant la notion d'unicité de l'information.
Il en est de même pour la notion d'**alias** qui permettra de ne pas dupliquer l'information, mais de créer des liens entre les fichiers, et ceux afin de garantir l'unicité de l'information.
Enfin, garantir l'interprétabilité des données consiste également à en garantir l'**origine**.
Les données doivent avoir la qualité nécessaire pour supporter le type d'utilisation pour laquelle elles ont été générées. En d'autres termes, la demande de qualité est aussi importante sur les données nécessaires au pilotage, à la gestion, à l'évaluation

d'un risque que sur celles utilisées pour une opération d'enquêtes de masses.
Ces quatre facteurs posent les bases essentielles de ce que peut être l'information. Il faut maintenant se demander de quelles notions a-t-ont besoin pour élargir notre propos à la qualité de l'information et comment l'appliquer à une organisation de production de données scientifiques. La qualité du système d'information et donc de la prise de décision est directement liée à ces dispositifs.

L'établissement de ces critères qualifiant les données permettent d'établir six dimensions qui définissent la qualité des données.

3. SIX DIMENSIONS DEFINISSANT LA QUALITE DES DONNEES

3.1 Pertinence des données

On entendra par pertinence des données, la satisfaction des besoins clients. Pour être qualifiées de pertinentes, les données doivent **éclairer les utilisateurs** sur les questions les plus importantes à leurs yeux. Cette pertinence a un **caractère subjectif**, car elle dépend avant tout des divers besoins utilisateurs et des réponses que l'on pourra apporter compte tenu des contraintes en matière de ressources.
Pour assurer cette pertinence, il faut s'appuyer sur quatre processus : un *processus de liaison* avec les utilisateurs, un *processus d'examen* des programmes de recherche ainsi qu'un besoin de pilotage global, un *processus d'établissement des priorités* et enfin intégrer *le suivi de la performance.*

a. Suivi des besoins

L'organisme référent doit se doter d'un ensemble de mécanismes qui lui permette de rester au courant des besoins en matière d'informations actuels et futurs de ses principaux utilisateurs. Pour ce faire, on pourra par exemple :

- participer à la création et / ou participer à des comités consultatifs composés de spécialistes des principaux secteurs étudiés et des métiers concernés (Ressources Humaines, finance, formation.. ;)
- participer à la création d'observatoires
- participer à des rencontres avec les principales associations industrielles
- travailler sur les rétroactions formulées par les utilisateurs et/ou tirer des demandes de renseignements.

Ces dispositifs ont pour but de cerner les lacunes liées à la qualité de l'information (l'information nécessaire aux utilisateurs qui n'est pas disponible ou pas suffisamment fiable).

b. Examen du programme

Le suivi des besoins tel qu'explicité ci-dessus aura pour effet d'obtenir un retour des utilisateurs sur les programmes de recherche ou sur les objectifs stratégiques ainsi que sur les besoins pour les programmes futurs. Il faut cependant également examiner les programmes de manière explicite et sur une base régulière pour évaluer si ceux-ci répondent aux besoins des clients, non seulement afin de valider la pertinence des sujets traités, mais aussi pour ce qui concerne l'exactitude et la rapidité de diffusion des données produites.
Pour évaluer le programme et/ou le projet, on va donc rassembler les données issues de la phase précédente de suivi des besoins que l'on complétera par le rapport d'un expert indépendant par exemple. Les programmes peuvent également recueillir et évaluer périodiquement les commentaires qu'ils reçoivent et rédiger un rapport énonçant les changements possibles afin de répondre au mieux aux besoins des clients.

c. Etablissement des priorités

Il y a fort à penser que les demandes seront souvent pour ne pas dire toujours supérieures aux ressources disponibles. Il va donc falloir comparer en faisant preuve de jugement, les divers besoins des différents groupes d'utilisateurs. Il faut cerner également les possibilités d'obtentions et d'extensions des ressources initiales qui permettraient de répondre pleinement à l'ensemble de la demande client.

d. Suivi de la performance

Nous pouvons citer deux principaux types de suivi de la performance en matière de pertinence.
Premièrement, un suivi par des descriptions des mécanismes particuliers utilisés qui sont appuyés aux mieux sur des mesures, à défaut sur des exemples de leur incidence qui montrent que les processus préalablement décrit sont en place.
Deuxièmement, on peut apporter la preuve de la pertinence des données par l'usage qui en est fait selon les résultats d'évaluations remplis par les clients eux-mêmes.
La seconde dimension que nous exposerons ici est l'exactitude, à savoir dans quelle mesure l'information décrit bien le phénomène qu'elle doit mesurer.

3.2 Exactitude

Habituellement, elle se définie par rapport aux estimations statistiques erronées et est traditionnellement décomposée en composante de biais[vii] et de variance[viii]. On peut également la définir par rapport aux sources d'erreurs principales susceptibles de mener à des données inexactes[ix].
Il faudra tout particulièrement s'assurer de l'exactitude des données aux cours des trois principales étapes du projet : la conception, la mise en œuvre et l'évaluation.

a. Conception

Il est assez rare que les programmes de recherche définissent des cibles en ce qui concerne l'exactitude des données. Ils indiquent plus souvent des quantités à estimer, des zones géographiques à étudier, des niveaux de détails ou de précisions voulus, des

estimations, mais l'exactitude des données demeure au mieux que très vague tout comme la définition des niveaux d'erreurs non imputable à l'échantillonnage. Lors de la définition des programmes de recherche, on essaie souvent de trouver le juste milieu entre exactitude et rapidité de diffusion en fonction des contraintes de ressources. Il convient alors d'envisager dans le processus de conception les diverses options en matière de niveaux d'exactitude.

b. Mise en œuvre

Il faut tout d'abord des renseignements qui permettent de surveiller et de corriger en temps réel les problèmes qui peuvent survenir lors de l'acquisition des informations. L'utilisation d'un système d'information permettra d'obtenir au bon moment les renseignements dont il a besoin afin d'adapter son propos ou pour corriger les situations difficiles. Il faudra également pouvoir accéder à l'information après coup si le modèle a été exécuté tel que prévu et que certains aspects ont posés problèmes. Les leçons à tirer pour les programmes suivant permettront d'entrer dans le cercle vertueux de l'amélioration continue. Bien entendu, l'information se rapportant directement à l'exactitude ne constitue qu'une partie infime des données que requière la gestion opérationnelle. Les renseignements liés aux choix et aux coûts du moment des opérations doivent tout autant être pris en considération pour les modèles ultérieurs.

c. Evaluation de l'exactitude

Cette troisième phase, probablement la plus importante du processus, est l'évaluation de l'exactitude de l'information.

Nous pourrons décomposer cette notion d'exactitude en trois composantes.

La première composante réside dans l'évaluation du **biais**. D'une façon générale, le terme *« biais »* désigne un écart systématique (non aléatoire) entre une grandeur et la prédiction de cette grandeur. C'est donc une valeur **d'erreur systématique**.

La seconde composante est la composante de « *variance* ». Habituellement, la **variance** d'erreur est définie en statistique comme la part de la variance observée (ou variance totale) qui est imputable aux fluctuations aléatoires de l'échantillonnage. Ces fluctuations concernent notamment le choix (aléatoire précisément) des composantes du dispositif de mesure utilisé, ainsi que des conditions dans lesquelles la mesure a lieu, choix des items, des classes, des moments d'observation, des correcteurs, soit une valeur d'**erreur aléatoire**.

La troisième composante est constituée **d'erreurs principales** susceptibles de mener à des données inexactes. On regardera alors les erreurs issues des incertitudes de couverture, d'échantillonnage, de réponses, de non réponses etc.

Les efforts à investir dans la mesure de l'exactitude sont avant tout une décision de gestion qui doit être prise compte tenu des compromis habituels lors de la conception de l'échantillon et des méthodes mises en œuvre

3.3 Rapidité de diffusion de l'information

On entend par rapidité de diffusion de l'information, le délai entre la fin du recueil de l'information et la date à laquelle les données sont disponibles.

Comme nous l'avons vu, il existe un lien fort entre la rapidité à laquelle l'information doit être diffusée et la pertinence de l'information. En effet, cette dernière influe fortement sur la célérité avec laquelle l'information va être diffusée.

On peut également se poser la question de la durée pendant laquelle l'information est utile. Nous pouvons dire que la réponse à cette question dépend fortement du phénomène observé.

Reste à trouver le bon compromis. En effet, est-il préférable d'avoir des données plus exactes plus tard ou moins exactes plus tôt ? On admettra donc que la rapidité de diffusion n'est pas un objectif intrinsèque.

Les principales dates de rendu de production doivent être annoncées longtemps à l'avance, laissant ainsi aux chercheurs le temps de planifier leurs recherches ainsi que la publication de leurs résultats. Les mêmes précautions sont nécessaires pour les reporting et les tableaux de bord fonctionnels réalisés pour le pilotage des établissements.

3.4 Accessibilité

Répondre à la question d'accessibilité revient à travailler sur quatre points essentiels.

Tout d'abord, avec quelle facilité peut- on se procurer les données ? Avec quelle facilité constate-t-on que l'information existe ? Quelle est le caractère approprié de la présentation de l'information ? Ou encore, quel est le coût d'accès à l'information.

Il faut veiller qu'aucune erreur ne se glisse en transférant l'information d'un programme de recherche vers les utilisateurs. A cette étape du processus, il conviendra de vérifier le transfert des bonnes informations dans les bases de données, de publier des tableaux dans la bonne version. Comme de telles erreurs sont susceptibles d'avoir lieu à l'étape de la livraison, les risques existent et les systèmes qualité réduisent la possibilité de telles erreurs.

Dans le même esprit, le retour des utilisateurs est essentiel. La rétroaction des utilisateurs peut provenir de statistiques automatisées sur l'utilisation des diverses composantes des systèmes, d'enquêtes sur le degré de satisfaction des utilisateurs quand à des produits, services ou systèmes de livraison ou il peut s'agir de commentaires de suggestions, de plaintes ou de témoignages d'approbations librement exprimés par les utilisateurs.

3.5 Interprétation

On entend par possibilité d'interprétation, la disponibilité de **renseignements supplémentaires** ou de **métadonnées** nécessaires à l'exploitation des données.

Une fois acquise cette notion de métadonnées, on s'intéressera également aux méthodes de collectes de données ainsi qu'aux indicateurs de l'exactitude des données.

L'information nécessaire à la compréhension des données peut se grouper en trois catégories. Tout d'abord les **concepts et les classifications** à la base des données c'est-à-dire quels sont les éléments mesurés afin que l'utilisateur puisse juger de la pertinence par rapport à son besoin, ensuite les **méthodes de collectes et de rassemblement** des données c'est-à-dire quelles méthodes j'utilise pour pouvoir juger de la pertinence des outils utilisés, enfin, les mesures de l'exactitude des données, quel est le résultat obtenu pour garantir la confiance dans les résultats. De ce fait, la description technique devient un véritable indicateur de l'exactitude des informations.

Lorsque l'on parle interprétation, on se doit également d'aborder la forme des rapports. Prétendre que l'information éclairant les données doit être compréhensible relève d'une évidence. Néanmoins, le producteur d'information doit s'efforcer de communiquer dans la langue de l'utilisateur et non pas dans un jargon interne, moyennant quoi l'utilisateur aura beaucoup de mal à interpréter les données.

3.6 Cohérence

Dernière branche pouvant qualifier la qualité de l'information, la cohérence des données concerne la liaison avec d'autre renseignement statistiques dans un cadre analytique au fil du temps.

Cette propriété se rapporte à la cohérence entre les mêmes éléments de données se rapportant à des unités de temps différentes, à la cohérence entre divers éléments de données se rapportant à la même unité de temps ou à la cohérence internationale.

Concrètement, une terminologie est la même pour tous les programmes. Par exemple, le « niveau d'instruction » aura la même signification dans une enquête sur le recensement de la population que dans une enquête sur la scolarité. En conséquence, les quantités exprimées ont des liens connus entre elles.

On arrive à cette notion de cohérence par l'adoption de cadres nationaux voir internationaux comme par exemple les normes IFRS.[x] qui rendent cohérent les systèmes de comptabilité mondiaux. Tous les systèmes de classifications types s'appliquant à tous types de variables permettent de garantir cette cohérence des données. La comparaison sur le plan international se fait en vérifiant si les normes adoptées sont conformes aux normes internationales quand il y en a.

Il faut également s'assurer que le processus de mesure n'entraine pas d'incohérence entre les sources de données même quand les quantités à mesurer sont garanties de façons uniformes.

Pour évaluer dans quelle mesure les données sont cohérentes, on peut définir trois grands ensembles de mesures. Il s'agit en premier lieu de valider l'existence et de mesurer le niveau d'utilisation des cadres de références de variables et des systèmes de classification types. Dans une deuxième phase, de valider l'existence et de mesurer le niveau d'utilisation des méthodes et des outils pour la conception et la mise en œuvre des programmes. Et enfin de mettre en place des indicateurs pour mesurer la fréquence et l'incohérence des données publiées.

4. INDICATEURS ET MESURES

Comme dans toute démarche qualité, la mise en place d'indicateurs est une étape essentielle pour s'engager dans le processus d'amélioration continu.

A partir de ces définitions théoriques, chaque organisme devrait s'approprier les concepts et créer sa propre définition opérationnelle en fonction des objectifs et des priorités des programmes et/ou des projets afin de définir des indicateurs pour chacune des dimensions et vérifier par des mesures régulières leurs évolutions dans le temps. Chacune des dimensions que nous avons exposées peuvent être mesurées, soit de manière objective au travers de suivis automatiques d'indicateurs spécifiques, soit de manière subjective en recueillant la perception des utilisateurs.

Nous pouvons citer quelques indicateurs à titre d'exemple.

L'âge des données est-il conforme au besoin métier ? Se poser cette question permet de s'interroger sur le critère d'**opportunité**. Les indicateurs que nous pouvons alors mettre en œuvre pourraient être *la date de la collecte des données, la date du dernier traitement et/ou encore le contrôle de la version*.

Est-ce que toutes les données nécessaires sont disponibles ? Répondre à cette question permet de mesurer si la donnée est **complète**. Mesurons alors si l'intégralité des données optionnelles est disponible, le nombre de valeurs non renseignés, le nombre de valeurs par défaut par rapport à la moyenne.

Quelles sont les données sources des informations contradictoires ? Permet de mesurer la **cohérence** des données. Nous pouvons alors mesurer par exemple la valeur de la déviation standard.[xi] ou effectuer des vérifications de plausibilité.

Afin de vérifier **l'exactitude** des données, interrogeons-nous afin de savoir en quoi les valeurs représentent la réalité ? On pourra alors mesurer la fréquence des changements de valeurs ou encore le feed-back des bénéficiaire.

Les données sont-elles compréhensible par les utilisateurs ? Cette question est un exemple d'interrogation permettant de répondre à la notion d'**interprétabilité des données**. On pourra alors mesurer la valorisation des données utilisateurs

Quels sont les données saisies, stockées, ou affichées dans un format non standard ? On s'intéressera dans ce cas à la mesure de **standardisation,** de **conformité** d'une donnée. On pourra alors mettre en œuvre un *certificat de conformité* par exemple.

Quelles sont les données répétées ? Permettra de s'interroger sur le caractère de **duplication** d'une

donnée. Nous pouvons mesurer alors le *nombre d'enregistrement dupliqués.*
Une fois les indicateurs définis, il faut mettre en place un système de mesure qui permet de surveiller leur évolution dans le temps. L'utilisation de ces indicateurs permettra de définir les plans d'actions nécessaires à l'amélioration des différents processus.

5. QUI EST CONCERNE ?

La qualité des données concerne tous les personnels de l'organisation. Nous l'avons déjà vu, le chercheur, l'ingénieur, le spécialiste métier, comme l'utilisateur de base ont un intérêt différent, mais un intérêt tout de même à participer à la démarche de qualité des données.
Les *dirigeants,* doivent être les premiers promoteurs de la qualité de l'information. Comme dans toute les démarches qualité, la direction de l'organisme à un rôle prépondérant [xii] dans la réussite d'une démarche qualité.
La direction doit avant tout **définir la stratégie** de l'établissement en matière de qualité des données. Elle doit **prendre les décisions** nécessaires au bon fonctionnement du système qualité de l'organisme ou de l'établissement. Enfin, elle doit assurer le **pilotage** du projet « qualité des données ».
Les *responsables opérationnels* sont avant tout les garants de la qualité de l'information à sa création.
Le responsable opérationnel a pour mission de *gérer au quotidien d'important volume* de données. *Travailler sur les interfaces* entre la direction et les métiers. Pour lui, le besoin d'informations se traduit par le besoin d'outils de reporting puissants et d'éléments pertinents et synthétiques.
Les *collaborateurs opérationnels* sont créateurs et utilisateurs des données correctes, ils ont quant-à eux *besoin d'informations fiables* et *besoins d'un accès simple et rapide* à la donnée. Nombres de problèmes de qualité des données sont dues à des erreurs de saisies de l'information. Il faudra veiller à sensibiliser les collaborateurs opérationnels à l'importance de la qualité de l'information, et ce afin qu'à la source cette information soit la plus exacte possible.
Enfin le *responsable du système d'information*. Il est garant de l'intégrité et de la disponibilité de l'information. Son rôle est de former, conseiller, aider et accompagner les utilisateurs à maintenir les données aux meilleurs niveaux de qualité. Il est également garant de l'infrastructure technique qui va permettre d'acquérir, stocker, modifier, restituer et sécuriser l'information. Il travaille en étroite collaboration avec les équipes chargées de l'aide au pilotage et de la qualité notamment dans le cadre des universités

6. GOUVERNANCE

Dans le cadre de la démarche qualité des données, l'établissement doit définir son modèle de gouvernance, c'est-à-dire la formalisation de son modèle de pilotage des technologies, des personnes et des processus.
On pourra distinguer deux instances dans le pilotage de ce projet : la direction de l'unité appuyé par les directions métiers et le comité « qualité des données »

6.1 Direction de l'unité, de l'établissement et directions métiers

Son rôle comme on a déjà pu le voir réside dans la sponsorisation du projet. Comme dans toute démarche qualité cette sponsorisation et indispensable. Il faut convaincre la direction générale et les directions métiers de l'impact de la non-qualité des données et surtout démontrer que la qualité est source d'excellence pour le laboratoire ou l'établissement. La conduite du changement et toutes ses méthodes seront des outils précieux pour convaincre les directions métiers de l'intérêt de la démarche. Elle s'assurera que la démarche est lancée.
Autre rôle de la direction générale, mettre en place un comité qualité des données, dont l'objectif sera de s'assurer que l'ensemble des projets intègrent la gestion de la qualité des données dans l'ensemble de leurs processus

6.2 Comité qualité des données

Ce comité est responsable de la qualité des données de l'entité. Ce modèle de pilotage doit comprendre une structure organisationnelle en charge de l'amélioration de la qualité de l'établissement, sous la responsabilité d'un sponsor qui doit avoir une influence sur les directions métiers.
Il définit les objectifs, les priorités. Il s'assure que tous les projets intègrent la qualité des données dans leurs processus. Il s'assure également de la disponibilité des financements nécessaires au projet.
Ce comité se réunit régulièrement pour assurer le suivi du projet qualité des données et assurer le suivi des actions d'améliorations. Enfin, il décide des nouvelles priorités.
Ce comité est composé d'experts issus de l'ensemble des directions de l'établissement. Ces derniers sont responsables de la définition, de la surveillance des mesures et indicateurs. L'analyste Qualité des Données doit avoir un lien fort avec la direction des systèmes d'informations. Son rôle est de mettre en place les outils de contrôle de la qualité des données, définir les principaux indicateurs et mesures de la qualité des données, justifier des programmes d'amélioration à mettre en œuvre et de mesurer de façon régulière les progrès effectués.

7. ENJEUX POUR LA RECHERCHE OU LES ETABLISSEMENTS D'ENSEIGNEMENT SUPERIEUR

Nous pouvons distinguer quatre impacts majeurs de la qualité des données. Ces quatre impacts sont à croiser en fonction du métier exercé
En premier lieu, il conviendra de réfléchir aux **enjeux stratégiques liés** à la qualité des données. Le décideur a avant tout besoin d'avoir confiance dans la

qualité et la pertinence des données qu'il a à analyser pour pouvoir effectuer ses recherches dans des conditions optimum. Nous verrons également que des réflexions visant le contrôle du processus de l'information sont d'ores et déjà engagées.

Dans un second temps, nous pouvons réfléchir aux **enjeux économiques**. Cette démarche qualité permet également d'élargir le champ de reconnaissance, notamment par des organismes certificateurs et/ou accréditeurs. Cette reconnaissance permet d'assurer une certaine confiance aux organismes de financement, aux commanditaires, aux donneurs d'ordres, aux bailleurs de fonds, à la communauté scientifique. La qualité de l'information peut devenir un critère important de sélection pour bénéficier d'un financement national, européen ou international. Si aujourd'hui l'évocation d'une norme dans un appel d'offre n'est qu'évoqué, il y a fort à parier que demain ce sera un critère déterminant dans le choix du laboratoire partenaire par exemple.

Nous pouvons également évoquer les **enjeux sociétaux et environnementaux**. Dans une société de plus en plus sceptique, la responsabilité du chercheur et des laboratoires par rapport à leurs recherches vis-à-vis de la société, apparait aujourd'hui normale. La qualité de l'information vise à garantir un niveau de confiance vis-à-vis des générations actuelles et futures.

Ces enjeux sont à croiser en fonction de l'impact que peut avoir la mise en œuvre d'un référentiel sur la qualité des données et sur les personnels de recherche. Si pour l'opérateur, la mise en œuvre d'un référentiel « qualité de l'information » n'aura pour effet qu'une meilleure gestion des données internes, pour l'ingénieur, producteur et praticiens des données, l'impact sera plus important notamment par la mise à disposition de méthodes pour mesurer et décrire la qualité des données. Enfin, pour le chercheur, la confiance dans ses corpus de donnée aura un impact très fort sur le déroulement de sa recherche ainsi que sur les résultats qui vont en découler.

8. CONCLUSION

La qualité des données peut être aujourd'hui un des enjeux majeurs des laboratoires de recherche ou des établissements d'enseignement supérieur. Même si actuellement les appels d'offres de recherche ne mentionne les démarches qualité que comme un plus, et non comme une obligation, il y a fort à parier que dans les années à venir, ces démarches deviennent indispensables pour obtenir des contrats de recherche voir pour faire face à la concurrence des laboratoires étrangers. L'attribution de moyens financier est directement liée à la performance et celle-ci se mesure notamment à partir des données.

La non-qualité des données a un coût important, tant humain que financier. Les enjeux à la fois scientifiques, économiques, sociétaux et environnementaux font que les établissements doivent être en mesure aujourd'hui de contrôler la qualité de leurs données.

Pour répondre à ces questions, la mise en place d'un comité de pilotage de la qualité des données veillera à traiter la pertinence, l'exactitude, la rapidité de diffusion ou encore l'accessibilité, l'interprétation ou alors la cohérence des données en liaison forte avec la direction du système d'information, direction garante du stockage, de la sécurisation ou encore de la diffusion de l'information.

L'idée de démarche et de pérennité est essentielle et caractéristique de l'approche qualité. Elle va à l'encontre d'une opération unique et isolée qui ne permet de nettoyer et d'améliorer les données que ponctuellement. Cela signifie que les objectifs, mesures et indicateurs associés doivent être portés par l'ensemble des acteurs concernés et en particulier se traduire par une implication forte de la direction.

---

[i] *Livre blanc JEMM research – Des données de qualité - Informatica*

[ii] *Electronic Data Interchange ou, en français, Echange de Données Informatisées.*
L'EDI peut être défini comme l'échange, d'ordinateur à ordinateur, de données concernant des transactions en utilisant des *réseaux et des formats normalisés.*
Les informations issues du système informatique de l'émetteur transitent par l'intermédiaire de réseaux vers le système informatique du partenaire pour y être intégrées automatiquement.

[iii] *Métadonnée signifie concrètement « des données sur des données ».*

Une métadonnée est une donnée qui a pour but de décrire une autre donnée. Les métadonnées peuvent être des informations complémentaires, nécessaires à la compréhension d'une autre information ou dans le but de permettre une utilisation pertinente. Par exemple, les métadonnées d'une page web peuvent informer sur le créateur de la page, de la date de création, de la date de publication, son contenu, etc.
Des métadonnées de qualité facilitent la consultation d'informations. Elles permettent d'améliorer la pertinence des résultats affichés lorsqu'un utilisateur effectue une requête.
http://www.graphic-evolution.fr/definition/metadonnees-64.html
Le terme de métadonnées est utilisé pour définir l'ensemble des informations techniques et descriptives ajoutées aux documents pour mieux les qualifier.
http://www2.cndp.fr/standards/metadonnees/general.htm

[iv] *Partie d'un enregistrement composée de valeurs multiples issues d'un même domaine.*

[v] *http://www.iso.org/iso/fr/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=7238*

[vi] *Le cloud computing est un concept de déportation sur des serveurs distants des traitements informatiques traditionnellement localisés sur le poste utilisateur*

Le concept d'informatique dans le nuage est comparable à celui de la distribution de l'énergie électrique. La puissance de calcul et de stockage de l'information est proposée à la consommation par des compagnies spécialisées. De ce fait, les entreprises n'ont plus besoin de serveurs propres, mais confient cette ressource à une entreprise qui leur garantit une puissance de calcul et de stockage à la demande.

[vii] *Erreurs systématiques*

[viii] *Erreurs aléatoire*

[ix] *Erreur de couverture, d'échantillonnage, de non-réponse et de réponses*

[x] *Normes internationales d'information financière, plus connues sous leur nom anglais d'International Financial Reporting Standards*

Normes comptables internationales, élaborées par le Bureau des standards comptables internationaux

[xi] *La déviation standard, permet d'évaluer la dispersion des mesures autour de la valeur moyenne*

[xii] *Cf Paragraphe 5 de la norme Iso9001*