

Gérald Duffing

Professeur Associé

ICN Business School, France

Laboratoire LORIA, équipe SITE, France

[Gerald.duffing@icn-groupe.fr](mailto:Gerald.duffing@icn-groupe.fr)



Odile Thiery

Professeur de l'Université Nancy 2

Laboratoire LORIA, équipe SITE France

[Odile.thiery@loria.fr](mailto:Odile.thiery@loria.fr)



Gestion et qualité de l'information stratégique :  
une approche par les risques des systèmes  
décisionnels.

**Résumé :** *L'aide à la prise de décision est aujourd'hui en partie réalisée grâce aux entrepôts de données, qui peuvent intégrer une grande quantité de données provenant aussi bien de l'entreprise que de son environnement. Cependant, la diversité et la complexité des données, des processus et des acteurs impliqués dans ce dispositif technique nous invitent à prendre en compte des risques nouveaux (comme, par exemple, la non qualité des données, la non détection de changements significatifs, etc.). Face à ces risques, il est souhaitable de concevoir un système plus proche de ses utilisateurs, capable d'assister leurs recherches ou leurs manipulations d'informations, de les conseiller, ou encore de les alerter à bon escient. Dans cette optique, nous proposons une approche par les risques qui permet, selon nous, une prise en compte complète et très spécifique du problème de l'assistance à l'utilisateur dans le cadre d'un Système Décisionnel. Nous en déduisons l'architecture générale d'un système collaboratif d'aide à la décision centré sur les besoins spécifiques de tous les acteurs du système décisionnel, lequel devient une sorte d'atelier de travail personnalisable.*

Gérald Duffing, Odile Thiery

Strategic Information quality and management:

A risk-based approach.

*Abstract: Decision making is often based on data warehouses, that store a large amount of data, coming from either inside or outside the organisation. However, the complexity and diversity of data, processes and actors involved require management of new kinds of risks (eg data quality, meaningful signal detection failures, etc.). To face these risks, a system that takes into account more of users' needs is to be devised: query and manipulation assistance, advices or alert generation are some of the intended features we foresee. We propose a risk-based approach that can, according to us, provide a comprehensive and specific answer to the users' assistance problem. We provide the architecture of such a collaborative, user-centric, decision support system, as a customisable data workbench.*

Mots clés et Key-words:

Qualité des données, Analyse des Risques, Systèmes d'Aide à la décision, Entrepôts de données, Intelligence Economique.

Data quality, Risk Analysis, Decision support systems, Data warehouse, Economic intelligence.

## 1. Introduction

Le décideur dispose aujourd'hui de nombreux outils pour faciliter son travail : souvent désignés sous l'appellation d'outils de *business intelligence*, ces systèmes décisionnels mettent à la disposition des utilisateurs une grande quantité d'informations, exploitable grâce à des outils d'interrogation, de manipulation, ou de représentation graphique. Puisque les décisions se construisent en partie à la suite d'analyses réalisées à l'aide de ces outils, plusieurs problèmes liés à la fois aux données et aux manipulations peuvent se poser : la qualité des données initiales est-elle avérée ? La sémantique des données manipulées est-elle claire ? Les manipulations opérées sur ces données ont-elles pu altérer leur qualité ? A ces questions, il n'existe pas de réponse simple et unique car, d'une part, le processus qui a permis de mettre ces données à la disposition des utilisateurs est complexe et, d'autre part, l'expertise et les besoins propres de chaque utilisateur vont avoir une influence directe sur la qualité de son processus de traitement des données et, donc, sur le résultat final, la décision.

Le problème de la qualité des données et des décisions est encore un problème ouvert, notamment parce qu'il est impossible d'apporter une réponse universelle, compte tenu de l'extrême diversité des domaines d'applications, des processus et des acteurs entrant en jeu. Pour mieux prendre en compte cette réalité complexe et apporter des réponses ciblées, nous proposons dans cet article une approche par les risques qui prend en compte à la fois les risques techniques (liés au fonctionnement même du système décisionnel) et les risques métiers (liés au domaine d'application).

Dans un premier temps nous définirons l'information comme matière première de la décision ainsi que ce que nous entendons par donnée de qualité. Dans un deuxième temps nous introduirons la gestion du processus d'Intelligence Economique dans l'équipe SITE et définirons comment se place le Système d'Information Décisionnel dans ce processus. Ensuite, nous expliciterons ce que nous entendons par approche par les risques et analyserons les risques encourus par la prise de décisions reposant sur des données de mauvaise qualité. Nous terminerons par la proposition d'une architecture pour le traitement des risques.

## 2. L'information, matière première de la décision

L'information peut être considérée comme l'une des matières premières utiles à la prise la décision. Nous nous intéressons dans cet article aux décisions qui nécessitent la mise à disposition des décideurs un ensemble de données et d'informations à la fois internes et externes. Si, en effet, les données à prendre en compte pour une décision de niveau opérationnel sont le plus souvent issues du système d'information de l'organisation, les décisions tactiques ou stratégiques réclament des informations hétérogènes qu'il peut être nécessaire de collecter hors de l'entreprise. C'est notamment indispensable dans le cadre de l'activité de veille stratégique, mais ce peut être aussi utile pour analyser et interpréter une situation donnée, en corrélant les deux types de données. Notons également que les besoins en informations des décideurs sont variés, souvent propres à chacun d'eux, et peuvent évoluer dans le temps.

D'un point de vue technique, ces informations seront souvent mémorisées dans un entrepôt de données (Inmon, 1994), qui se révèle être un outil clé d'aide à la prise de décision (Yong-Tae Park, 2006). Au travers de quatre phases successives (Gouarné, 1998), il a vocation à identifier et extraire toutes les données utiles à la prise de décision (phase de collecte), à les intégrer dans un schéma général unique (phase d'intégration), à les

transformer selon un modèle multi-dimensionnel et/ou à les segmenter en magasins de données – les *datamarts* – pour faciliter leur manipulation (phase de diffusion), et enfin à les mettre à la disposition des utilisateurs finals, via des outils appropriés (phase de présentation).

Durant toutes étapes du cycle de vie d'un entrepôt, on doit veiller à garantir une qualité optimale de ces données. Il est évident que la qualité d'une décision est en partie liée à la qualité des données utilisées pour la construire (Chengalur-Smith et al, 1999 ; Ge et Helfert, 2006). De nombreuses recherches sur la qualité de données ont été menées et continuent d'occuper la communauté scientifique aujourd'hui (Ramos Lima et al, 2006). La notion de qualité peut cependant revêtir de nombreuses formes et il n'existe pas de définition universelle, car elle ne peut être appréciée que par rapport à un utilisateur donné, dans un contexte précis. On retient cependant souvent les critères de précision, de pertinence, de crédibilité, d'accessibilité, d'interprétabilité, etc. Ainsi, la personnalisation des entrepôts de données est impérative : l'outil doit pouvoir reconnaître et gérer de manière adéquate divers profils d'utilisateurs (Thiery et David, 2002).

Le problème de la qualité des données est tout à fait central à plus d'un titre (Huh et al, 1990). En amont, les sources de données doivent être, prises individuellement, fiables. L'évaluation d'un degré de qualité peut être difficile, surtout s'il s'agit de données externes à l'organisation. La phase de collecte est suivie d'une phase d'intégration des données de différentes sources dans l'entrepôt : il convient de s'assurer que ceci ne mettra pas en péril la qualité des données, étant entendu que l'intégration peut donner lieu à des transformations, des agrégations, etc. En aval, lors de la phase de diffusion, les données ou les indicateurs calculés que l'on peut en déduire doivent eux aussi être de qualité ; ils doivent être par ailleurs suffisamment documentés pour minimiser les risques de mauvaise utilisation ou de mauvaise interprétation lors de la phase de présentation.

Plusieurs travaux se sont intéressés à la problématique de la qualité dans les entrepôts. Il s'agit par exemple d'instrumenter l'entrepôt avec des calculs d'indicateurs de qualité des données, qui seront périodiquement réévalués (Vassiliadis et al, 1999). Les entrepôts sont des environnements techniques complexes qui font l'objet de fréquentes modifications, notamment pour répondre aux besoins d'information des décideurs : il faut donc pouvoir ajouter de nouvelles données, calculer de nouveaux indicateurs etc. Dans ce contexte, une méthode de maintien de qualité tout au long de la vie l'entrepôt a été proposée (Quix, 2003). Parallèlement, plusieurs techniques de nettoyage de données ont été mises au point (Rahm et Hai Do, 2000), relayées par de nombreuses solutions commerciales.

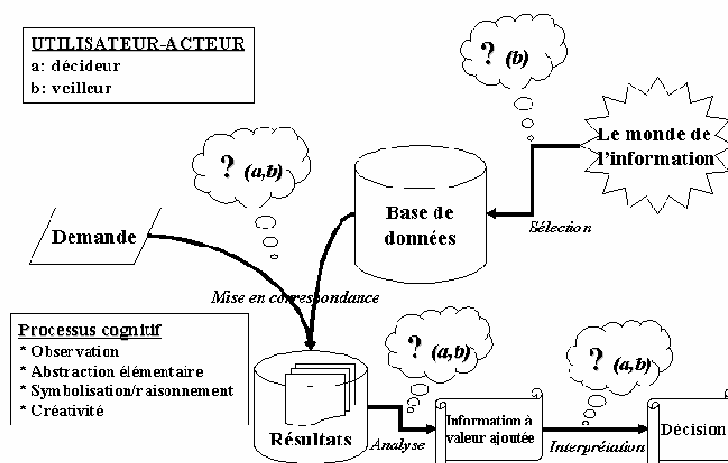
L'entrepôt est un système complexe qui s'efforce de mettre à la disposition de tous les utilisateurs toutes les informations utiles. Face à cette promesse ambitieuse, il est essentiel de pouvoir s'appuyer sur des méthodes permettant de gérer la mise en place et le fonctionnement de l'entrepôt. Les travaux menés dans le domaine de l'Intelligence Economique (IE) apportent une réponse adaptée.

### **3. La gestion des informations selon le modèle de l'Intelligence Economique**

#### **3.1. Le processus d'Intelligence Economique (IE)**

L'Intelligence économique est le processus de collecte, de traitement et de diffusion de l'information qui a pour objet la réduction de la part d'incertitude dans la prise de toute décision stratégique (Revelli, 1998). Les étapes principales de ce processus (Figure 1) reposent toutes sur la captation et l'utilisation de données provenant « du monde de

l'information », composé de l'organisation elle-même – c'est-à-dire des données qu'elle produit – et de son environnement. Nous appellerons « veille » le processus informationnel par lequel l'organisation recherche et collecte des informations utiles à une meilleure compréhension de son fonctionnement et de son environnement.



(Figure 1 : Le processus d'Intelligence Economique. Source : Thiery et David, 2002)

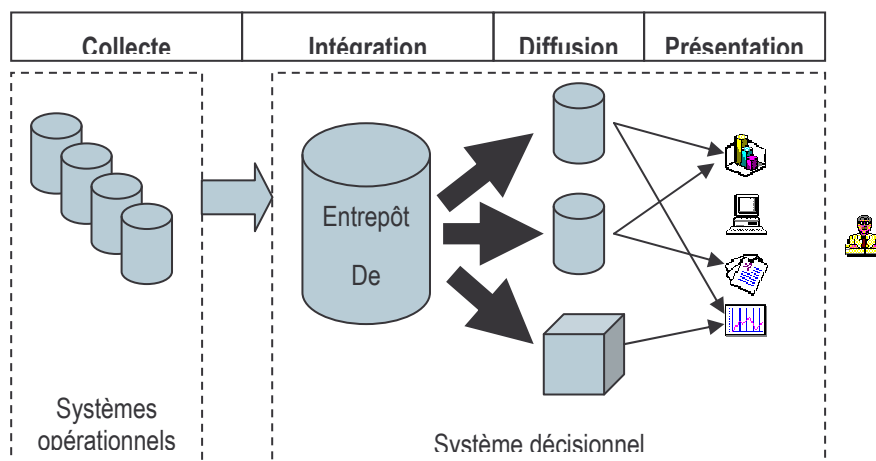
La démarche d'IE telle que nous la concevons dans l'équipe SITE se décompose en dix activités principales dérivées de ce processus :

1. Identification des besoins sous forme de problèmes à résoudre (menace, risque, danger),
2. Identification des types de résultat,
3. Identification des types d'informations nécessaires pour obtenir le résultat,
4. Identification des sources pertinentes,
5. Validation des sources,
6. Collecte d'information,
7. Validation des informations,
8. Traitement des informations collectées pour le calcul des indicateurs,
9. Interprétation des indicateurs,
10. Prise de décision pour la résolution du problème.

### 3.2. Le système décisionnel au cœur de l'activité d'IE

L'entrepôt de données représente l'élément fondamental du système décisionnel (Figure 2) et on peut mettre en relation les étapes du processus d'IE présentées ci-dessus avec les processus qui régissent le fonctionnement de l'entrepôt.

Pour décrire les activités de sa gestion, nous nous appuyons sur trois niveaux d'intervention et sur les modèles associés. Le niveau *conception* assure la mise en place du système décisionnel et des processus qu'il supportera ; il est fondé sur une démarche adaptée de développement de systèmes d'information (Afolabi, 2007). Le niveau *exploitation* décrit le fonctionnement effectif du système décisionnel, avec les activités correspondant au chargement des données en provenance des systèmes de production ou de l'environnement (on rencontre souvent le terme ETL, pour Extraction – Transformation – Loading). Le niveau *utilisation* regroupe les activités des acteurs de type « veilleur » et « décideur », qui effectuent des recherches de données, des manipulations, etc. Les étapes clés d'une recherche typique sont : la compréhension et la définition des besoins – supporté par le modèle MEPD (Bouaka, 2006), l'instanciation de ce besoin sous forme de requête – à l'aide notamment du modèle WISP (Kislin, 2007), le traitement de la requête, l'analyse et l'interprétation par le décideur, et enfin la prise de décision.



(Figure 2 : le Système décisionnel)

Les utilisateurs d'un tel système se divisent en deux grandes catégories d'acteurs : le *décideur* est l'utilisateur confronté à un problème de décision ; il doit fonder sa réponse sur des indicateurs pertinents par rapport à son problème. Il a la responsabilité de l'identification du problème décisionnel, de l'analyse qui lui succède et de la décision finale. Puisque ce n'est souvent pas un spécialiste en système d'information, il est assisté du *veilleur*, qui a la responsabilité de l'alimentation en informations des différentes bases de l'organisation. D'un point de vue technique, il fait appel à un autre acteur, le *technicien*, dont nous parlerons peu dans cet article, mais dont le rôle est important, puisqu'il adapte le système selon les besoins exprimés par les décideurs et « rationalisés » par le veilleur. Le rôle du veilleur est d'aider le décideur à transformer un problème décisionnel en un problème de recherche d'information. Au sein de ces deux classes d'utilisateurs principales, des profils variés peuvent exister, amenant la nécessité de disposer d'un véritable modèle de l'utilisateur : la connaissance des habitudes associées à chaque profil connu peut aider le système à affiner son comportement (David, 1999). En effet, les utilisateurs adoptent des attitudes différentes par rapport à une information (Choo, 2007) : il existe donc différents

niveaux d'acceptabilité quant à la qualité ou l'importance accordée à une donnée. Par exemple, un utilisateur A acceptera qu'une information soit incomplète, tandis que B jugera alors cette même information comme étant de faible qualité. C'est pour cette raison que certains travaux laissent à l'utilisateur la possibilité d'interroger le système en spécifiant eux-même des niveaux de qualité exigés (Burgess et al, 2007).

Nous pensons que le système d'IE peut être vu comme un atelier de travail collaboratif où différents experts collectent, consultent, manipulent, analysent des données, en vue de construire des tableaux de bords (surveillance), des modèles d'analyse (interprétation) ou tout autre artefact qui leur seront propres, ou qui pourront être partagés. Les objectifs de chaque utilisateur contribuent à la raison d'être de l'entrepôt : aboutir à une vision la plus claire possible de l'organisation et de son environnement et, le cas échéant, prendre les meilleures décisions possibles.

Or, comme l'entrepôt est un système en constante évolution, ces manipulations ne sont pas sans risque sur la qualité des données, dont le maintien est un véritable défi. Nous avons déjà noté que la notion de qualité est propre à chaque utilisateur (Strong et al, 1997) : il ne suffit donc pas de mémoriser des indices de qualités globaux sur chaque donnée. Comment assurer la qualité d'une information produite (en terme, par exemple, d'intégrité, de crédibilité, d'actualité) ? Comment garantir sa bonne interprétation par d'autres utilisateurs ? Ces deux exemples simples suggèrent que les évolutions de l'entrepôt soient contrôlées, ainsi que les actions des utilisateurs.

A ce stade, nous cherchons à mettre en place un système décisionnel adapté à l'activité d'IE, où différents profils d'utilisateurs peuvent interagir en toute sécurité. Ses principales caractéristiques sont les suivantes : il prend en compte les trois niveaux *conception*, *exploitation* et *utilisation* ; il est *personnalisable* et apporte une assistance ciblée à tous les types d'acteurs ; il est une plate-forme *d'observation* et *d'expérimentation*, qui supporte une évolution constante et différenciée selon les acteurs ; il fournit des mécanismes de surveillance de la qualité des données. Pour satisfaire à ces conditions, la démarche que nous proposons réunit les aspects techniques et métier à l'aide d'un objet d'étude unificateur : le risque.

#### **4. L'approche par les risques**

Les notions de qualité et de risque sont intimement liées : pour une information précise, exiger un critère de qualité donné va nous aider à identifier les risques de non qualité. De façon complémentaire, un risque mis sous surveillance doit nous aider à garantir le critère de qualité attendu.

##### **4.1. Méthode générale de gestion des risques**

L'approche par les risques a surtout été appliquée à la gestion de projet (Bruckner et al, 2001) et à la sécurité des systèmes d'information (CORAS, 2000). L'approche CORAS est intéressante car elle permet de définir les « actifs » à mettre sous surveillance, d'énumérer les « menaces » qui pèsent sur eux et de spécifier les « traitements » permettant de traiter ces menaces. Dans les organisations, la gestion des risques prend des formes variées et nous avons choisi de nous intéresser exclusivement dans cet article aux risques liés à la gestion et à l'utilisation des systèmes d'information. Dans ce sous-domaine, la gamme des travaux existants s'étend de normes très spécifiques, comme la gestion de la sécurité des informations, jusqu'à des processus généraux de gestion du risque, comme la norme

australienne AS/NZS 4360 (1999), sur laquelle nous nous appuyons pour établir notre démarche ; elle se décompose en cinq étapes principales :

1. La définition du contexte, où l'on précise les objectifs, les intervenants et leurs enjeux, les critères d'appréciation des risques.
2. L'identification des risques, qui est une première approche où on indique les scénarios de risques principaux, en expliquant leur déroulement.
3. L'analyse des risques est l'étape qui poursuit l'exploration en indiquant les contrôles possibles, les probabilités d'occurrence, ainsi que les conséquences attendues.
4. L'évaluation des risques va préciser l'étape précédente et aboutit à un classement des risques.
5. Le traitement des risques, enfin, énumère les options possibles de gestion des risques identifiés précédemment, et explique comment les parades peuvent être mises en place.

Des étapes présentées ici, nous dégageons deux niveaux de modélisation utiles dans le cadre d'un entrepôt de données. Les aspects statiques (étapes 1 à 4) nous conduisent à élaborer une liste de méta données propres à analyser les risques possibles, tandis que les aspects dynamiques (étape 5) nous amènent à proposer des scénarios d'actions possibles pour pallier aux risques identifiés.

#### **4.2. Représentation du risque par la notion d'enjeu**

Notre proposition poursuit les travaux de (Bouaka, 2004), qui introduit la notion d'enjeu dans son Modèle d'Explicitation du Problème Décisionnel (MEPD). Pour nous, le risque peut être la conséquence d'une suite d'événements détectables dans le système décisionnel, et qui y sont représentés sous forme d'enjeux métiers ou techniques. Nous reprenons et complétons la proposition de Bouaka qui décompose l'enjeu en trois éléments fondamentaux : l'objet, le signal et l'hypothèse.

- L'objet représente un élément concret du monde du décideur, auquel il s'intéresse au travers d'une ou plusieurs données particulières. Par exemple, le volume des ventes ou le prix de vente d'un produit donné. On suppose que cet objet et ses données associées existent dans le système décisionnel. Cette modélisation permet de s'assurer de la définition unique, sémantiquement claire et partagée par tous d'un objet d'intérêt.
- Le signal est un constat réalisé sur l'objet. Il est associé à une formule de calcul et à un algorithme particulier d'analyse qui opère sur les données de l'entrepôt. Par exemple, il peut s'agir d'une hausse modérée mais continue du prix de vente sur une période de temps donnée. Un grand nombre de signaux peut être associé à un même objet, en fonction de l'acteur qui s'y intéresse, de la méthode de détection de ce signal, etc.
- L'hypothèse est la conséquence que la détection du signal a sur un ou plusieurs objets donnés, ce qui permet de construire un réseau. Il existe deux types d'hypothèse, *l'assertion* et le *risque* : le risque est surveillé systématiquement et périodiquement.

Chacun de ces éléments est décrit par une liste de méta données élémentaires (Tableau 1): un nom, un numéro de version, un propriétaire, une ACL (*Access Control List*, utile pour assurer la confidentialité), une liste d'observateurs (acteurs intéressés par cet élément, son évolution et les actions qu'il déclenche), un système d'horodatage (date de création, de dernière modification, de dernière mise-à-jour, de dernier déclenchement, etc.). L'hypothèse inclut des éléments de probabilité permettant d'évaluer ses chances d'occurrence, sa vraisemblance (la théorie de Dempster-Shafer, par exemple, permet d'introduire des notions intéressantes de crédibilité et de plausibilité). Elle permet également de définir des actions à exécuter suite à sa propre activation.

<b>Objet</b>	
Donnée	Elément d'information issu du SD (assorti de méta-données indiquant s'il s'agit d'une donnée élémentaire, agrégée, calculée, etc.)
Source(s)	Adresse précise de la donnée dans le système (par exemple, pour une table relationnelle : BD, table, colonne)
Type	Type de donnée : temps, volume, prix, etc.
Fréquence	Fréquence de mise-à-jour.
Granularité	Niveau de détail (exemple : jour, semaine, ...)
<b>Signal</b>	
Analyseur	Elément logiciel chargé de surveiller l'évolution de l'objet sous-jacent (nom, version, ACL, propriétaire)
Evénement(s)	Analyseur sollicité, plage ou algorithme de détection du signal sur les données produites par l'analyseur.
<b>Hypothèse</b>	
Type	Assertion ou Risque.
Objet(s) impacté(s)	Désignation de l'objet et de la donnée impactés.
Probabilité(s)	Eléments utiles à l'évaluation de l'hypothèse.
Action(s)	Action à déclencher dès lors que l'hypothèse est générée.

(Tableau 1 : Données spécifiques de l'enjeu)

### 4.3. Identification des risques

Nous présentons successivement les deux catégories de risques (métiers et techniques) prises en compte dans notre proposition, qui peuvent être représentés par un réseau appelé « vue ». Une vue est *spécifique à un acteur*, même si chacun des éléments qui la constituent peut être associé à un ou plusieurs acteurs (lesquels peuvent ainsi modéliser leur propre vue d'un même problème). Cette personnalisation est l'un des objectifs majeurs de notre proposition.

D'un point de vue « métier », l'exemple présenté (Figure 3) est un réseau d'enjeux très simples concernant deux objets : la boisson « Mangolight », le produit fini, et l'une de ses matières premières, la mangue. Considérons les quatre enjeux suivants, surveillés automatiquement par le système décisionnel, permettant de mettre sous contrôle un risque métier : l'effritement de la marge (nous présentons uniquement et succinctement l'Objet concerné *O*, le signal surveillé *S* et l'hypothèse générée *H* :

(j) O = Marge ; S = Baisse ; H = Diminution du coût de production

(c) O = Coût de Production ; S = Hausse ; H = Hausse coût du stock

(d) O = Volume stocké ; S = Faible ; H = Hausse du coût du stock

(e) O = Volume disponible marché ; S = Baisse ; H = Hausse du coût du stock

Le système évalue périodiquement tous les enjeux ayant des hypothèses de type risque. A chaque évaluation, le système effectue un contrôle qualité de l'objet concerné (ce qui permet notamment d'associer des risques métiers à des risques techniques). L'enchaînement suivant pourrait être obtenu.

L'enjeu (j), dont l'hypothèse définit un risque, est évalué. Suite à la détection du signal S, l'hypothèse H est générée : une alerte « Diminution de marge » est déclenchée et l'enjeu (c) est examiné.

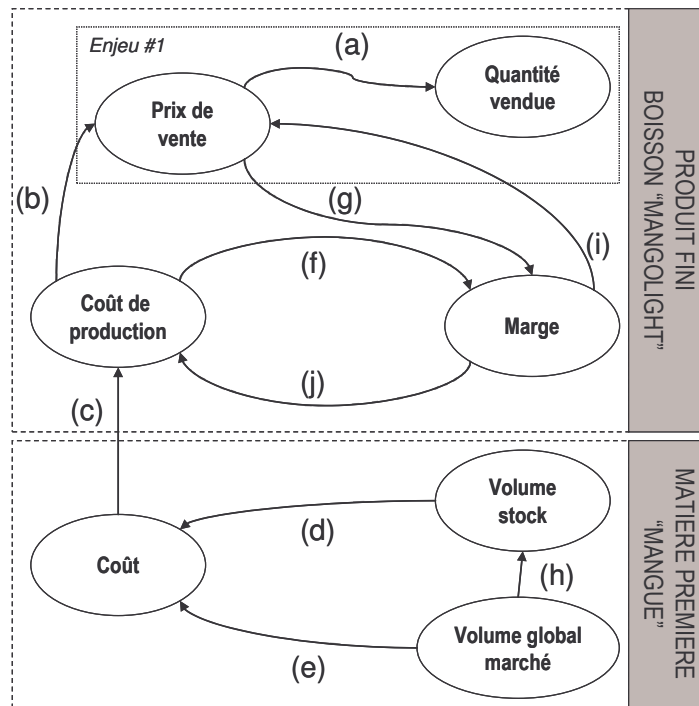
Pour l'enjeu (c), un contrôle qualité repère que l'objet (le coût de production) est une donnée agrégée et propose à l'utilisateur un accès immédiat aux constituants de ce coût. Si le signal S de (c) est avéré, l'examen des enjeux (d) et (e) est demandé.

Pour l'enjeu (d), le contrôle qualité détermine si les valeurs du volume stocké sont disponibles et à jour pour la période considérée et vérifie que cette période est suffisamment longue pour autoriser une évaluation correcte de la tendance analysée pour la détection du signal S. Si non, une extraction est automatiquement demandée. Si la disponibilité et la qualité des données utiles à l'évaluation de S est avérée, les données sont proposées à l'utilisateur. Dans le cas contraire, une alerte qualité est générée.

Pour l'enjeu (e), le contrôle qualité détermine si les données sont disponibles (il s'agit ici d'une donnée externe à l'entreprise). Si non, une alerte est générée à l'attention du veilleur, suivie d'une notification à l'utilisateur. On peut aussi imaginer que l'entreprise soit abonnée à une source de données externes qui pourrait alors être mobilisée automatiquement.

On peut voir au travers de ce court exemple que le système peut réagir de manière à :

- Identifier, au sein de l'énorme volume de données que représente l'entrepôt, les informations utiles au traitement d'un phénomène modélisé comme un « objet d'intérêt », ici l'évolution de la marge d'un produit ;
- Extraire automatiquement les données pertinentes à l'interprétation (accès aux données de détail d'un indicateur agrégé, extraction d'une série chronologique, etc.) ;
- Surveiller la qualité des données et alerter l'utilisateur en cas de doute, voire limiter l'utilisation de données inadaptées (vérification des données utiles à la détermination d'une tendance, ici en s'assurant de la longueur suffisante d'une série) ;
- Rechercher, de manière automatique ou non, des données externes utiles à l'interprétation d'un phénomène (ici, les données relatives à un marché de matières premières) ;



(Figure 3 : Exemple d'enjeux)

D'un point de vue technique cette fois, chacune des étapes du processus général d'IE présente des difficultés : aux étapes 1 à 5, il faut accéder à des informations pertinentes (c'est-à-dire possédant une sémantique claire, relatives ou contributives au sujet, disponibles facilement, manipulable et d'une précision adaptée). Puis il faut s'assurer que toutes les données utiles ont pu être collectées (étapes 6 et 7). Ensuite, il convient de s'assurer que les données auront été correctement manipulées et interprétées (étapes 8 et 9). Alors seulement, l'ultime étape de prise de décision pourra être réalisée avec les meilleures chances de succès. Pour identifier les risques potentiels génériques, nous effectuons une analyse croisée entre les éléments Acteurs-Information-Processus, pris deux à deux. Ce travail doit être réalisé aux trois niveaux : conception, exploitation et utilisation du système décisionnel :

- *Acteur-Information* : les risques correspondent à des problèmes de qualité de données, selon l'interprétation de chaque acteur. Citons la pertinence des données, la confidentialité, la pertinence de leur présentation, la précision, l'exhaustivité, etc.
- *Acteur-Processus* : les risques sont liés à la bonne connaissance du processus par l'acteur, à l'adaptation de son niveau d'expertise par rapport au processus qu'il manipule.
- *Information-Processus* : les risques sont relatifs à l'adaptation du processus à l'information considérée, à la pertinence de la présentation, au type de données et au niveau de qualité de donnée exigé par le processus, etc.

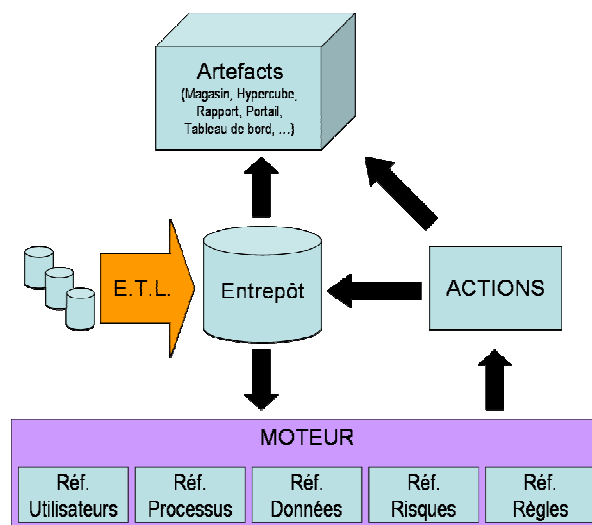
Dans le cadre de cet article, nous donnons un aperçu des risques possibles au seul niveau utilisation, les autres niveaux conception et exploitation relevant plus des aspects strictement informatiques (Tableau 2), sachant que tous ne sont pas détectables automatiquement par le système décisionnel : cette liste non exhaustive peut cependant être utilisée comme une *check-list* qui vise à mettre en évidence les problèmes devant attirer l'attention des acteurs du système, ou à conduire l'audit d'un système existant.

Acteur - Information	Violation de confidentialité. Non détection d'un signal important. Mauvaise interprétation (niveau d'expertise insuffisant). Mauvaise qualité de données (obsolète, erronée, manquante, ...).
Acteur - Processus	Expertise insuffisante avec le processus, l'outil. Manipulation inadéquate.
Information - Processus	Représentation incorrecte (incompatibilité information – processus). Volumétrie. Informations non accessible, non calculable. Incompatibilité de version entre l'information et le processus.

(Tableau 2 : Risques au niveau de l'utilisation)

#### 4.4. Proposition d'une architecture pour le traitement des risques

A la suite de l'analyse et de l'évaluation des risques, un ensemble de scénarios peut être mis en place pour les risques qui se prêtent à un traitement automatisé. Cet ensemble est nécessairement très spécifique à chaque cas ; dans le cadre de nos recherches, notre objectif est de proposer l'architecture générale d'un système décisionnel apte à rendre opérationnels ces scénarios (Figure 4).



(Figure 4 : Proposition d'architecture pour un SD collaboratif)

Sur ce schéma, un ensemble de référentiels permet de décrire le système décisionnel. Le référentiel des utilisateurs reconnaît chaque personne pouvant accéder au système ; il en stocke le profil (caractéristiques, habitudes, etc.) et il aide ainsi le système à personnaliser ses réponses. Le référentiel des processus reprend tous les processus techniques d'un système décisionnel. Le référentiel des données contient les méta données qui décrivent toutes les données stockées et notamment leur qualité. Le référentiel des risques permet de mémoriser tout ce qui est représentatif d'une menace pour le système (risque technique) ou pour l'utilisateur (risque métier). Enfin, le référentiel de règles indique ce qu'il y a lieu de détecter et de faire, le cas échéant. Ces deux derniers référentiels mémorisent les enjeux.

Tous ces référentiels sont utiles pour piloter les actions nécessaires à la gestion des risques de l'entrepôt. Parmi les actions possibles, citons :

- Alerter un utilisateur sur un risque potentiel, ou bien mettre à sa disposition des informations complémentaires (action « assistance ») ;
- Interdire une opération, c'est-à-dire bloquer la commande qui a conduit à déclencher l'action, après détection d'un risque (action « sécurité ») ;
- Déclencher des opérations spécifiques sur le système décisionnel, par exemple collecter des données supplémentaires ou les retraiter (action « exploitation », transparente pour l'utilisateur).

Nous avons donné dans ce paragraphe une grille d'identification des principaux risques à considérer et proposé l'architecture d'un système décisionnel capable de les prendre en compte, fondé sur une représentation du risque sous forme d'enjeu personnalisable par chaque acteur.

## 5. Discussion

Notre approche se veut intégratrice : elle couvre l'ensemble du cycle de vie de l'entrepôt (conception, exploitation, utilisation), puisque les risques pouvant mettre en péril la qualité des données se rencontrent à chaque étape du processus général d'IE. Par exemple, on peut gérer à la fois un problème d'information manquante (conception), un problème de fréquence de rafraîchissement (exploitation) ou encore un problème de représentation graphique inadéquate (utilisation). L'approche par les risques permet également d'unifier les problèmes techniques et métiers, tout en s'adaptant aux spécificités de tous les acteurs (types d'acteur, niveaux d'expertise, attitudes, exigences hétérogènes, etc...). Le système décisionnel tel que nous le concevons prend en compte les préoccupations spécifiques de chacun des acteurs, grâce à la modélisation sous forme d'enjeu faisant apparaître, d'une part, les relations causales existant entre les objets métiers et, d'autre part, les risques qui traduisent la vulnérabilité de l'organisation.

La collaboration entre les acteurs est favorisée par le fait que, par exemple, un enjeu d'un acteur A peut avoir des conséquences sur d'autres enjeux d'un acteur B. Ces deux caractéristiques complètent utilement l'entrepôt de données classique sans se substituer aux outils finals auxquels les acteurs sont habitués, mais aussi et surtout sans cloisonner les différentes fonctions de l'organisation ; il faut souligner ici le rôle important du veilleur, qui matérialise dans l'entrepôt cette vue consensuelle de l'organisation (Knauf, 2007, Kislin 2007). Le système est bien adapté à la veille collaborative, puisqu'il offre une assistance à la surveillance des données internes et externes, tout en permettant d'établir les relations existant entre ces deux types d'information, au travers d'une analyse générale de risque. Il offre de plus un accès « multi-facettes » à un même objet, puisque l'expertise de tous les acteurs concernés est traduite dans les risques et enjeux associés, rendus accessibles à tous (dans la limite des règles de confidentialité définies). Nous pensons que la rencontre de plusieurs profils d'acteurs est propice à la mise en évidence de relations causales nouvelles, dont l'observation permettra d'évaluer la pertinence. Ceci souligne le caractère d'« atelier » de l'activité de veille que notre architecture souhaite promouvoir, dans un contexte d'incertitude où l'expérimentation joue un rôle important. La modélisation et la représentation de liens causaux dans cette architecture peuvent contribuer à une activité plus générale de gestion des connaissances au sein de l'organisation, en fournissant un outil de capitalisation de connaissances opérationnel.

Notre approche doit être complémentaire aux outils que l'on trouve sur le marché : en amont avec les logiciels de gestion de la qualité des données et en aval avec les outils de présentation de données ou de data mining. L'apport principal est la mise en correspondance entre l'aspect technique de la qualité d'une donnée, et l'utilisation de cette donnée dans un réseau de risques métier : ceci permet, à notre sens, de mieux mesurer la « valeur métier » d'une donnée et de mieux apprécier sa criticité, pour arriver ainsi à une meilleure évaluation de la qualité contextuelle, c'est-à-dire par rapport à l'expression d'une « valeur métier » (Even et Shankaranarayanan, 2005).

## **6. Conclusion**

Les systèmes décisionnels prennent une place croissante dans les organisations, si bien que de plus en plus de décisions sont prises à partir d'analyses réalisées sur des données et informations extraites de ces systèmes. Leur facilité d'accès et d'utilisation peut masquer des risques. Dans ces conditions, la fiabilité du système décisionnel doit être maximale. La difficulté réside, notamment, dans l'importance du volume à traiter et dans le caractère fortement évolutif d'un tel système. Nous avons présenté une approche des systèmes décisionnels visant améliorer ce niveau de fiabilité, qui s'attache à limiter leur utilisation inadéquate en proposant de prendre en considération les risques à la fois techniques et métiers qui peuvent conduire à une mauvaise décision.

Notre proposition met en avant deux éléments fondamentaux : l'utilisateur, en tant qu'acteur ayant une expertise et une expérience du système connues de ce dernier, et le risque, objet unificateur autour duquel s'organise le fonctionnement du système. En nous appuyant sur une méthode de gestion des risques classique, nous aboutissons à un modèle collaboratif de système décisionnel qui peut être adapté à tout domaine d'application et qui permet de couvrir les trois grandes périodes du cycle de vie : la conception, l'exploitation et l'utilisation.

A l'aide de cette architecture, il est possible de mettre sous contrôle à la fois les données et les processus qui agissent sur elles en modélisant les risques qu'ils encourent, dans le but de garantir le maintien de la qualité du système décisionnel, en déclenchant différents types d'actions de sécurité ou d'assistance. Nous pensons que cette démarche peut amener les acteurs d'un projet décisionnel à prendre en compte un maximum de sources potentielles de risque et aboutir ainsi à un système plus fiable.

## 7. Références bibliographiques

Afolabi B. S. (2007), La conception et l'adaptation de la structure d'un système d'intelligence économique par l'observation des comportements de l'utilisateur. Thèse de doctorat de l'Université Nancy 2.

Australian/New Zealand Standard (1999), AS/NZS 4360 Risk Management.

Bouaka N. (2004), Développement d'un modèle pour l'explicitation d'un problème décisionnel : un outil d'aide à la décision dans un environnement d'intelligence économique. Thèse de doctorat. Université Nancy 2. Nancy.

Bruckner R., List B., Schiefer J. (2001), Risk-Management for Data Warehouse Systems. Lectures Notes in Computer Science, vol. 2114. pp. 219-229. Springer-Verlag, Berlin.

Bueno D., David A. (2001), METIORE: A Personalized Information Retrieval System. Lecture Notes in Computer Science, vol. 2109, pp. 168.

Burgess M., Gray W. A., Fiddian N. J. (2007), Using quality criteria to assist in information searching. International Journal of Information Quality (1)1, pp. 83-99.

Chengalur-Smith I. N., Ballou D. P., Pazer H. L. (1999), The Impact of Data Quality Information on Decision Making: An Exploratory Analysis. IEEE Transactions on Knowledge and Data Engineering (11)6, pp. 853-864.

Choo C. W. (2007), Information seeking in organizations: epistemic contexts and contests. Information Research (12)2, (Paper 298. Available at <http://InformationR.net/ir/12-2/paper298.html>).

CORAS (2000), A platform for risk analysis of security critical systems. IST-2000-25031. <http://coras.sourceforge.net>

David A. (1999), Modélisation de l'utilisateur et recherche coopérative d'information dans les SRI multi média en vue de la personnalisation des réponses. HDR en Sciences de l'Information et de la Communication. Université Nancy 2.

David A., Thiery O. (2001), Prise en compte de l'utilisateur dans un Système d'Information Stratégique. Actes de la conférence « Veille Stratégique Scientifique et Technologique, VSST01 ». Barcelona.

David, A. and Thiery, O. (2003), L'Architecture EQUA2te et son Application à l'intelligence économique. In David A. (Ed.) Actes de la Conférence « Intelligence Economique: Recherches et Applications ». INRIA Lorraine, Nancy.

Even A., Shankaranarayanan G. (2005), Value-Driven Data Quality Assessment. International Conference on Information Quality.

Gackowski Z. J. (2005), Operations Quality of Data and Information. International Conference on Information Quality.

Gackowski Z. J. (2006), Redefining Information Quality and its Measuring: The Operations Management Approach. International Conference on Information Quality.

Ge M., Helfert M. (2006), A Framework to Assess Decision Quality Using Information Quality Dimensions. International Conference on Information Quality.

Gouarné J.-M. (1998), Le projet décisionnel. Enjeux, modèles, architectures du Data Warehouse. Eyrolles.

Huh Y. U., Keller F. R., Redman T. C., Watkins A.R. (1990), Data quality. Information and Software Technology (32)8, pp. 559-565.

Inmon W.H. (1994), Building the Data Warehouse, John Wiley&Sons.

Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P. (2003), Metadata and Data Warehouse Quality, in « Fundamentals of Data Warehouses ». 2<sup>nd</sup> edition. Springer.

Kislin Ph. (2007), Modélisation du problème informationnel du veilleur dans la démarche d'intelligence économique. Thèse de doctorat Université Nancy 2.

Knauf A. (2007): Caractérisation des rôles du coordinateur-animateur: émergence d'un acteur nécessaire à la mise en pratique d'un dispositif régional d'intelligence économique. Thèse de doctorat Université Nancy 2.

Peguiron F. (2006), Application de l'Intelligence Economique dans un Système d'Information Stratégique universitaire : les apports de la modélisation des acteurs. Thèse de doctorat Université Nancy 2.

Quix C. (2003), Repository Support for Data Warehouse Evolution. ICFAI Journal of Systems Management (1)2, pp. 57-72.

Rahm E., Hai Do D. (2000), Data cleaning: Problems and current approaches. IEEE Data Engineering Bulletin (23)4, pp. 3-13.

Ramos Lima L. F., Gastaud Maçada A. C., Vargas L. M. (2006), Research into Information Quality: A Study of the State-of-the Art in IQ and its Consolidation. International Conference on Information Quality.

Revelli C. (1998), Intelligence stratégique sur Internet. Dunod.

Salber D., Coutaz J., Decouchant D., Riveill M. (1995), De l'observabilité et le de l'honneteté : le cas du contrôle d'accès dans la communication Homme-Homme Médiatisée. Actes du congrès IHM'95, pp. 27-34. CEPAD Ed.

Strong D. M., Lee Y. W., Wang R. Y. (1997), Data Quality in Context. Communication of the ACM (40)5, pp. 103-110.

Thiery, O. David, A. (2002). Modélisation de l'Utilisateur, Système d'Informations stratégiques et Intelligence Economique. Revue Association pour le Développement du Logiciel (ADELI). n° 47.

Vassiliadis P., Bouzeghoub M., Quix C. (1999), Towards Quality-Oriented Data Warehouse Usage and Evolution. Lecture Notes in Computer Science, vol. 1626.

Wang R., Strong D. (1996), Beyond accuracy : What data quality means to data consumers. Journal of Management Information Systems (12)4.

Yong-Tae Park Y. Y. (2006), An empirical investigation of the effects of data warehousing on decision performance. Information Management (43)1, pp. 51-61.