

***Big data* : une mise en perspective empirique des impacts du Big Data sur la mutation des processus décisionnels**

Résumé:

Le *big data* est aujourd'hui en passe de faire muter les systèmes décisionnels. En effet, la décision ne s'appuie à présent plus uniquement sur les informations structurées qui étaient jusque-là collectées et stockées par l'organisation, mais aussi sur toutes les données non structurées se trouvant à l'extérieur du carcan des entreprises. Le *cloud* et les informations qu'il contient impactent dorénavant les décisions et nous assistons à l'émergence de la *business intelligence* 3.0. Avec la croissance d'internet, des réseaux sociaux et des objets connectés et communicants, les informations sont aujourd'hui plus abondantes que jamais et la croissance de leur production est chaque jour plus rapide. En 2012, 2.5 exaoctets de données (1 exaoctet représentant un millions de gigaoctets de données) venaient chaque jour grossir les rangs des *big data* (McAfee, 2012), qui devraient peser dès 2020 plus de 40 zettaoctets (Valduriez, 2014) pour 30 milliards d'appareils connectés (The economist, 2014) et 50 milliards de capteurs (Davenport, Soulard, 2014). L'un des aspects les plus critiques de tous ces flux d'information est l'impact qu'auront ceux-ci sur la manière dont les décisions sont prises. En effet, dans le cadre d'un environnement au sein duquel les données étaient rares et difficiles à obtenir, il était logique de laisser la prise de décision se conditionner à l'intuition du décideur d'expérience (Klein, 2007). Or, l'information et la connaissance étant dorénavant à la disposition de tous, le rôle des experts et décideurs change peu à peu. Les *big data* permettent en particulier aux systèmes analytiques et décisionnels de baser leurs prises de décision sur des modèles globaux. Or, la prise en considération de toutes les dimensions des situations rencontrées n'était jusque-là pas à la portée de l'homme, rationnellement limité (Simon, 1971). Les *big data* et le traitement des données non structurées imposent cependant de modifier l'architecture des systèmes d'aide à la décision (SAD) des organisations. Le *Time aware decision support system* permet par exemple de prendre en compte la temporalité des données du big data, il présente un fort intérêt dans le cadre du management des équipes opérationnelles de terrain (Police, Pompiers, etc.). Son intégration impose toutefois une redéfinition complète de l'architecture décisionnelle. Ce papier est un état des lieux des évolutions que subissent les systèmes d'aide la décision sous la pression des *big data*. Il présente l'impact que le *TADSS* aurait eu sur l'architecture décisionnelle de la Police néerlandaise dans le cadre de la gestion de l'afflux non programmé de 500 000 personnes dans la ville d'Enschede. Enfin, il ouvre le débat sur les questionnements éthiques soulevés par ces nouvelles technologies, l'analyse des données personnelles des individus fait en effet plus que jamais débat.

Mots clés : Intuition – Décision – Systèmes d'aide à la décision – Rationalité – *Business Intelligence*.

INTRODUCTION :

La *business intelligence* fait partie des thèmes incontournables du domaine de la gestion des systèmes d'information et de communication (Lebraty, 2006). Plusieurs chercheurs des sciences de gestion tels que Simon, Kahneman ou Klein ont tenté de définir les mécanismes menant à la prise de décision individuelle et collective. Les résultats des différentes études tendent vers la conclusion suivante : dans le cadre de situations extrêmes, il est préférable pour l'expert de s'en remettre à son intuition. Il est en effet parfois très compliqué et risqué d'anticiper les réactions d'une machine programmée (qui suit des procédures figées) lors de la survenue d'une situation nouvelle : les fonctions d'autopilotage des avions de ligne se désactivent par exemple face à des situations non routinières (Klein, 1999). Le modèle *RPD* (*Recognition Primed Decision*) de Klein permet de donner un cadre et une direction à la prise de décision en situation. Il apparaît plus adapté que les systèmes décisionnels qui se révèlent incompatibles voir dangereux lorsque utilisés hors du champ de problèmes bien caractérisés et dont l'évolution est connue. Or, le système d'aide à la décision (SAD) est défini de la manière suivante : il s'agit « *d'un système informatisé interactif aidant le décideur à manipuler des données et des modèles pour résoudre des problèmes mal structurés* » (Gorry, Morton, 1971). Nous nous trouvons de ce fait face à une incohérence : la notion de « problèmes mal structurés » induit en effet inévitablement la résolution de difficultés fréquemment complexes.

Les SAD permettent de partager des idées et ont pour principal objectif d'améliorer la qualité des décisions prises. Plusieurs limites en lien avec l'adoption de ces systèmes ont cependant pu être observées : il est en effet parfois compliqué pour les décideurs de ne pas s'en remettre à leurs expériences ou de gérer de grandes quantités d'informations. L'incompréhension des théories intégrées au SAD peut en outre mener au rejet du système par les utilisateurs qui préféreront dans ce cas s'en remettre aux avis de personnes physiques (Lebraty, 2006).

Nous assistons actuellement à une révolution dans le domaine de la *business intelligence* dont les systèmes se dotent peu à peu de *clusters Hadoop* ouvrant la voie vers l'analyse et l'exploitation des grandes données (ou *big data*). Jim Gray (Balazinska et al., 2007) propose la définition suivante du *big data* : Il s'agit d'« *un important volume d'informations non structurées et générées par une grande pluralité de nouvelles sources* ». De nombreuses analyses sont engagées autour de l'intérêt que représente le *big data* pour la *business intelligence* et les SAD (George et al., 2014). Le paramètre principal à prendre en compte

n'est pas la dimension du *big data* mais la qualité et la vélocité des données (structurées, semi structurées, non structurées) qu'il contient, ainsi que les compétences nécessaires à leur exploitation : il est moins question de *big data* que de *smart data*. Le décideur doit trouver les données susceptibles de générer de la valeur pour l'entreprise et avoir les moyens de les intégrer dans la prise de décision : 30% des dirigeants estiment en ce sens que l'absence de structure des données du *big data* est l'élément le plus problématique (Davenport, Patil, 2012). Les compétences nécessaires à l'analyse et au traitement des données du *big data* freinent encore grandement les entreprises. En ce sens, le *big data* ne concerne pour le moment qu'une minorité d'entreprises : 9 sur 10 d'entre elles estimaient en 2014 manquer de compétences technologiques ou humaines pour se lancer dans une stratégie big data (Cointot, Eychenne, 2014).

Cette communication est un état des lieux des évolutions des SIAD sous la pression des big data et traite donc la question de recherche suivante : **Les SAD actuels sont-ils adaptés aux environnements *big data* ?** Afin d'offrir des éléments de réponse à cette question, la première section s'intéresse aux modèles principaux de la décision, définit les SAD et leurs limites tout en expliquant l'intérêt que présente les *big data* pour ceux-ci. La deuxième section présente l'intégration des SAD actuels en environnement *Big Data*. Enfin la troisième section fait office de conclusion et ouvre le sujet traité.

1. PRISE DE DECISION ET SAD

1.1. Les trois modèles principaux de décision

Dans le domaine des sciences de gestion, plusieurs modèles ont été définis pour caractériser la prise de décision :

- **Le modèle analytique** : Il permet de traiter des problèmes simples ou compliqués (du moment qu'ils restent décomposables) dont l'ensemble des déterminants est connu. L'analyse rationnelle a pour objectif de définir tous les éléments d'un contexte, de lister toutes les options possibles et d'opérer un choix par rapport à une pondération des éléments ayant le plus d'importance aux yeux du décideur. De nombreux outils (sous forme de tableaux ou de cartographies) permettent de réaliser les étapes précédant la prise de décision dans le cadre de cette approche. L'objectif est ici

d'optimiser la décision pour identifier la meilleure solution. La décision doit de plus être celle qui répondra le mieux aux attentes du décideur mais peut toutefois être le résultat d'un compromis entre les éléments à prendre en compte (du moment qu'il n'existe aucune autre meilleure possibilité). Ce modèle est le plus adapté pour les processus de prise de décision en environnements certains (Kahneman, 2012).

- **Le modèle de la rationalité limitée (Simon, 1955):** Il concerne un individu novice dans son domaine et confronté à un problème complexe. La notion de rationalité fut la pierre angulaire de nombre de travaux de Simon qui met en avant les limites des décideurs en termes de capacités cognitives : limites physiques (ex : vitesse que celui-ci peut atteindre) et computationnelles (relatives à la capacité de l'individu à pouvoir programmer efficacement des événements). Simon argue du fait que la rationalité humaine est en réalité approximative, au contraire de celle qui devrait être exigée par certaines situations (ex : anticipation du cours des actions). *L'intelligent design choice model* (Simon, Newell, 1971) définit quatre phases de la prise de décision. Dans le cadre de celui-ci, l'individu se trouve face à plusieurs solutions potentielles et sélectionne celle qui lui convient le mieux au regard du contexte, son choix est donc relativement rationnel.
- **Le modèle *Recognition Primed Decision* (Klein, 1999):** Gary Klein a construit le modèle *RPD* par rapport à des observations réalisées dans le cadre de missions sur le terrain. Ce modèle ne s'intéresse pas à ce que les individus (experts) doivent faire face à un problème mais à ce qu'ils font effectivement lorsque confrontés à celui-ci. Klein a pu remarquer que les experts tels que les commandants de pompiers, s'en remettent intuitivement à leurs expériences pour prendre des décisions face à des situations extrêmes ou nouvelles. Par la projection mentale, les experts se révèlent capables de tester des hypothèses et de définir un cours de l'action afin d'anticiper les conséquences de la décision qu'ils prendront. Ils comparent la problématique face à laquelle ils se trouvent à d'autres, rencontrées par le passé, et usent d'analogies pour adapter ce qu'ils savent. Ce mode décisionnel permet aux experts de ces domaines de ne pas avoir à effectuer de comparaison entre les différentes possibilités s'offrant à eux. La réponse à une problématique est en effet très souvent présente dans la « mémoire » de l'expert qui ne dispose bien souvent pas du temps nécessaire à la réalisation d'un *mapping* complet des éléments de la situation rencontrée. La décision est ici directement fonction de l'expérience de l'expert.

Chaque modèle décisionnel est donc plus ou moins bien adapté à certaines problématiques (simples, compliquées, complexes). Le choix du modèle orientera par ailleurs vers un certain type de SAD en fonction de la décision à prendre. Un commandement de pompier ne peut pas (dans le cadre de missions opérationnelles) s'en remettre au même système d'aide à la décision que celui utilisé par un statisticien pour réaliser des projections à moyens ou long terme. Les *big data* se révèlent précieuses pour la globalité des modèles présentés mais les informations ne seront cependant pas valorisées de la même manière en fonction du SAD et du modèle décisionnel choisi.

1.2 Quels SAD pour quels types de décision ?

Les SAD peuvent être orientés données ou modèles (Alter et al, 1977). Ils permettent de partager des idées et ont pour finalité d'améliorer les décisions des décideurs.

Les SAD orientés données sont adaptés aux décisions « intuitives » dans la mesure où ils n'obligent pas à réaliser un choix entre différentes options. Les réseaux de neurones et les systèmes usant de l'intelligence artificielle appartiennent à ce type de SAD. L'article « *Human problem solving* » (Simon, Newell, 1971) a été précurseur dans le domaine de l'intelligence artificielle. Cet article décrit la mise au point de programmes capables d'émuler l'humain pour la réalisation de tâches telles que les échecs ou les puzzles.

Les réseaux de neurones sont construits, à l'égard de cette recherche, sur la base de l'émulation (Lebraty, 2006). Ils permettent de lier la conception du SAD et la cognition du décideur. L'objectif est de savoir comment un individu lambda traite et exploite l'information qu'il a à disposition. Le concept de style cognitif était jusque-là sujet de discorde entre les groupes de chercheurs, les uns dénonçant l'immaturité du concept pour être utilisé comme la base d'un système SAD, les autres affirmant qu'un SAD adapté au décideur ne pouvait qu'être plus efficace. L'approche cognitive a cependant mis en avant l'importance de l'interface Homme-machine (IHM) et la nécessité de créer des outils s'adaptant au modèle mental du décideur. L'avènement des *big data* est très prometteur pour les SAD puisque l'analyse et l'exploitation de ces données permettront de prendre en compte le style cognitif du décideur, ainsi que le contexte de la situation rencontrée, et ce en accord avec le modèle *RPD* (Klein, 1999).

Les SAD orientés modèles (c'est le cas des systèmes experts) proposent des choix entre plusieurs options aux décideurs. La principale limite de ces systèmes est liée à leur rationalité limitée à quelques points précis (en fonction des modèles qu'ils contiennent), ce qui ne leur permet pas de développer de vision globale (Klein, 1999). Les temps de traitement des requêtes sont par ailleurs particulièrement importants et bloquent de facto la génération de résultats en continu (Brasseur, 2013). Nous nous trouvons donc face à des machines équipées d'algorithmes capables de fournir des résultats rationnels bien plus pertinents que ceux issus des raisonnements de l'homme (Meehl, 1959), mais cependant incapables de réaliser des associations entre tous les différents éléments d'une situation (notamment la temporalité) et d'agir en fonction des contextes.

Nous constatons que chacun de ces SAD a ses propres faiblesses. Les *big data* pourraient néanmoins bientôt palier celles-ci et faire de l'analytique la pierre angulaire des futures prises de décision. Elles permettront aux machines de s'émanciper de l'obligation de se focaliser sur un ou quelques éléments des situations, nous faisant ainsi entrer dans une ère où l'intelligence artificielle s'appuiera sur des modèles prenant en compte tous les attributs du contexte. Les capacités computationnelles de la machine ne sont en outre pas limitées (au contraire de celles de l'homme) ou tout du moins évolutives en fonction des avancées technologiques. Le système expert *Watson* (conçu par IBM), combine les intelligences artificielles algorithmiques à un comportement calqué sur les capacités de déduction de l'homme (Cointot, Eychenne, 2014). Ce système ne se base pas uniquement sur des calculs (au contraire du système analytique ayant battu Garry Kasparov aux échecs en 1996) mais est capable d'aller « lire » par exemple des contenus textes divers sur Wikipédia ou les réseaux sociaux, etc. Ces analyses l'amènent à réaliser des déductions sans se borner à un simple raisonnement algorithmique. *Watson* est dorénavant capable de générer des diagnostics de cancer équivalents à ceux d'un spécialiste expérimenté. Il se tient par ailleurs informé des dernières nouveautés et découvertes en temps réel, ce qu'un homme n'est pas capable de faire. Au contraire de *Watson* et des SAD développés autour des *big data*, la plupart des SAD actuels (qu'ils soient orientés données ou modèles) n'ont pas été construits dans le but d'aller chercher de l'information en dehors de leurs bases de données respectives et il convient de définir quelle infrastructure est nécessaire pour prendre en compte ces données.

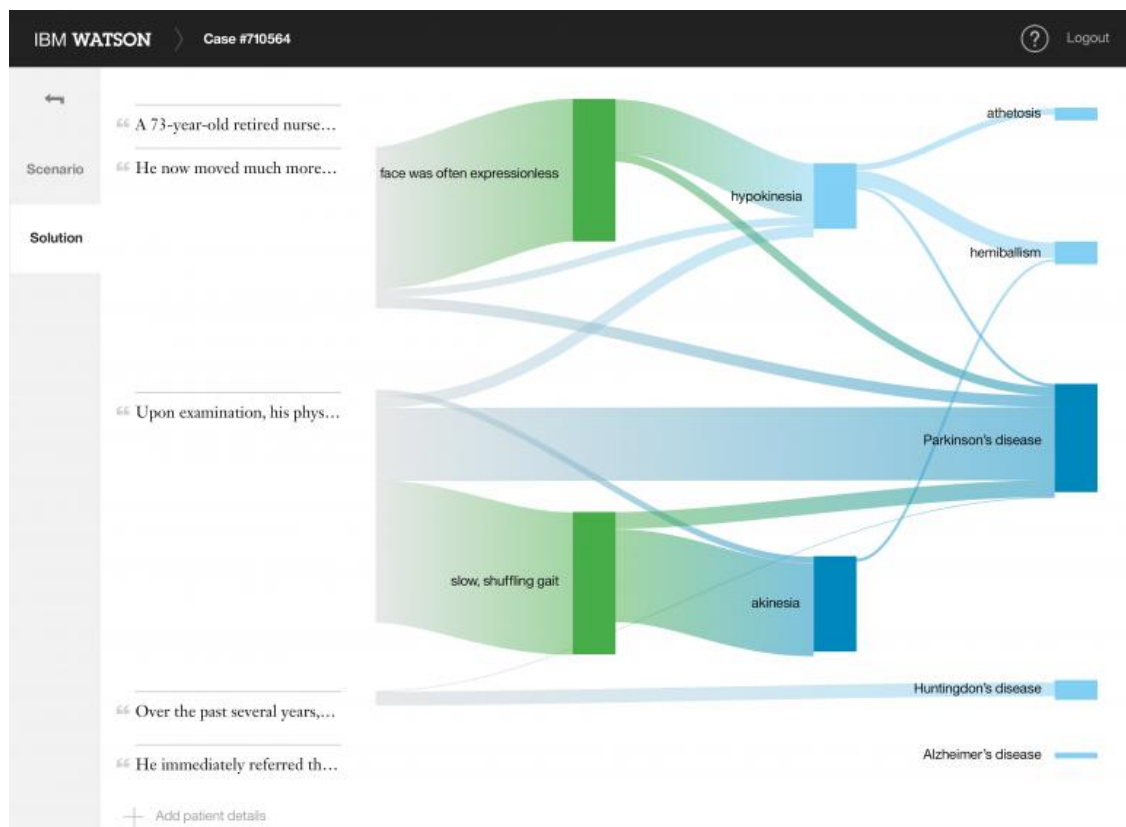


Figure 1: Interface Watson - Diagnostics médicaux

2. LES SAD ACTUELS EN ENVIRONNEMENT *BIG DATA* ?

2.1. L'infrastructure nécessaire à la prise en compte du *big data*

Plusieurs modèles relatifs à l'implantation du *big data* peuvent être choisis par une organisation (Cointot, Eychenne, 2014):

Le modèle disruptif : Le *big data* est dans ce cas l'élément central de la décision et tout est construit autour de son architecture. Les sociétés telles que *Yahoo* ou *Google* se basent sur ce modèle.

Le modèle évolutif : Le *big data* est intégré pour enrichir le modèle informatique déjà présent dans l'entreprise. Il est ici simplement une source potentielle de données.

Le modèle hybride : L'entrepôt de données est dans ce cas lié au système *big data*. Les deux systèmes sont bien intégrés et chacun traite des données particulières. Ce modèle permet de ne

pas impacter trop fortement le modèle décisionnel des sociétés. Les solutions proposées par *Oracle* ou *Terradata* sont de ce type.

La majorité des entreprises françaises souhaitant intégrer le *big data* à leurs processus décisionnels dispose déjà d'une architecture dédiée à la *business intelligence*. Presque aucune ne fait donc le choix de mettre en place un modèle disruptif par crainte de devoir redéfinir en totalité son « ossature » décisionnelle. La plupart des compagnies n'ont toutefois pas besoin de solution totale *big data* mais de solutions spécialisées pour traiter un problème spécifique (Van Rijmenam, 2014), le choix d'un modèle évolutif ou hybride se révèle donc valable.

Les organisations, font cependant souvent face à des difficultés lorsqu'elles cherchent à traiter des grandes quantités de données en conservant leur existant (Schmarzo, Baland, 2014):

- Aucun de leurs outils de base n'a été conçu pour traiter des informations transitant sur les réseaux sociaux ou du contenu non structuré. Ils sont dédiés à l'analyse de bases de données classiques et structurées (en lignes et en colonnes).
- Des entrepôts de données (*datawarehouses* reposant sur le système *online analytical processing* : *OLAP*) ont été mis en place par le passé en vue d'aider les décideurs des organisations. L'objectif de ces entrepôts est de rendre des données très hétérogènes accessibles à tous les utilisateurs. Cependant, les systèmes *OLAP* ne prennent pas en compte certains formats (par exemple les formats vidéo) et ne peuvent ainsi analyser et traiter certaines données du *big data* (Brasseur, 2013).
- Les anciens systèmes ne permettent pas de réaliser des recommandations prédictives (Lebraty, 2006). Les outils dédiés à l'intelligence artificielle nécessitent par ailleurs de recourir à une personne physique pour réaliser les requêtes. Les nouveaux outils de prédiction créent des modèles qui seront transformés en continu par l'analyse de données historiques, ils gagnent ainsi en performance au fil de leur utilisation. Le *Knowledge Graph of Google* fonctionne selon ce modèle et permettra in fine de passer du moteur de recherches au moteur de connaissances. Il sera ainsi possible de questionner directement *Google* qui réalisera des associations par rapport à l'historique complet des recherches opérées mondialement, mais aussi par rapport aux données personnelles des utilisateurs (Van Rijmenam, 2014).
- Il est difficile pour les systèmes d'aides à la décision d'interroger plusieurs bases de données (Lebraty, 2006). Or une stratégie *big data* impose de prendre en compte les bases de données du *cloud computing* (qui est le seul élément offrant des capacités de

stockage compatibles avec les données massives) en parallèle de celles de l'entreprise. Plusieurs offres *big data cloud* existent (*Cloudburst, Oracle Cloud service, etc.*).

- La prise en compte de la dimension temporelle est problématique pour les SAD classiques (Cointot, Eychenne, 2014). Elle nécessite le recours à une plateforme de traitement analysant et mettant en corrélation les données et informations issues de milliers voire de millions de sources en continu (ex : MapReduce). Plusieurs millions de messages peuvent devoir être analysés par seconde. Cette analyse permet de ne plus définir des tendances *a posteriori* (systèmes classiques) mais en temps réel : c'est une révolution. Cette analyse fait appel aux données GPS, informations techniques diverses, etc. et permet de donner des conseils aux utilisateurs en fonction des événements (exemple : modification de la conduite suite à la survenue d'une panne moteur). Les données sont issues des capteurs, des smartphones ou des réseaux sociaux. La dimension temporelle sera plus amplement étudiée dans la partie suivante de cette communication et un outil permettant de prendre en compte celle-ci sera présenté.

En 2013, les entreprises ont dépensés près de 31 milliards pour intégrer des technologies *big data* à leurs structures (Van Rijmenam, 2014). Lorsqu'une entreprise intègre le *big data* à ses stratégies décisionnelles, elle passe d'une vision tournée vers le passé à une autre orientée vers l'avenir (exploitation des données en temps réel pour définir des tendances, etc...). Nous avons pu constater plusieurs limites relatives à la prise en compte de l'environnement *big data* par les SAD des entreprises. Il est à noter que l'environnement *big data* impacte aussi fortement les outils de visualisations. Le traitement d'une quantité importante de données nécessite de définir des représentations capables d'amener à la compréhension des résultats. Les relations doivent aussi être mise en avant (courbes, tableaux, etc...) et peuvent dorénavant faire appel à des modèles multidimensionnels très complexes (cubes, étoiles, etc...). Cette complexité impose de mettre en avant des continuums (que l'œil humain, au contraire de la machine repère facilement). Il devient très aisé de générer des histoires issues des données brutes. Le système peut ainsi réaliser des projections mentales par rapport aux informations qu'il contient et les retranscrire automatiquement sous forme de récits. L'objectif est de rendre accessible au décideur le cours de l'action qu'il propose (Davenport, Soulard, 2014), *Narrative Sciences* a par exemple mis en place une fonction de ce type afin d'aider les utilisateurs dans le cadre de la rédaction d'articles.

L'interface doit en outre être particulièrement adaptée, au vu de l'influence que celle-ci peut avoir sur le décideur (Schmarzo, Baland, 2014). La complexité est par ailleurs à éviter afin de permettre à l'utilisateur de s'appropriier le système. Il veut simplement savoir ce qu'il se passe et ce qu'il doit faire (en se basant sur les bonnes pratiques ayant fonctionnées par le passé). Bill Schmarzo propose une interface axée sur 3 sections (perspectives, recommandations, indicateurs) :

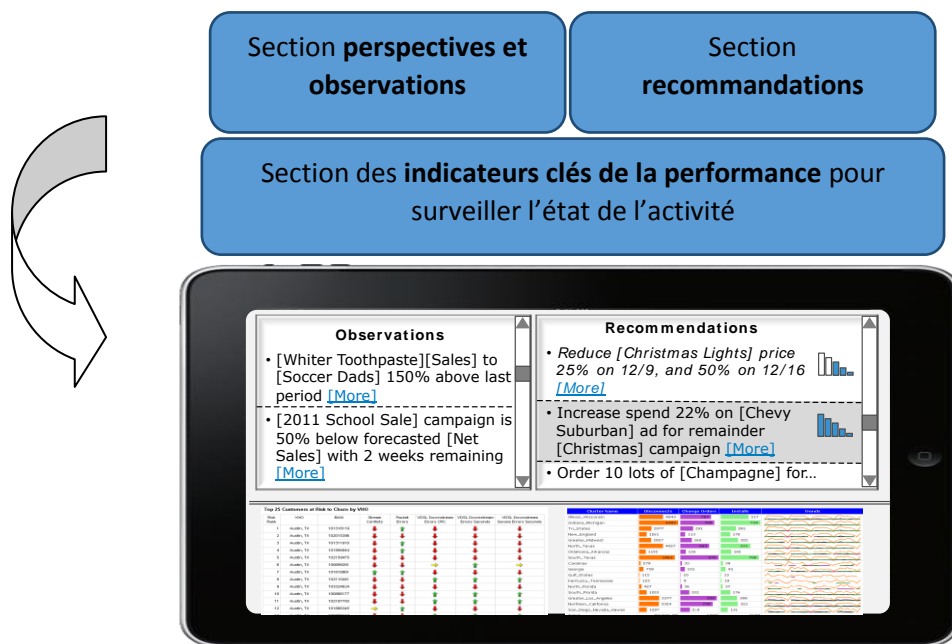


Figure 2: Interface perspectives-recommandations (Bill Schmarzo, 2014)

2.2. La prise en compte du big data par les SAD : le TADSS

Le *big data* est en passe de faire muter les processus décisionnels actuels et concerne tous les systèmes d'informations opérationnels (stocks, CRM, etc.) (Van Rijmenam, 2014). L'intégration et l'exploitation des données non structurées du *big data* dans les systèmes d'aide à la décision se fait par le recours à différents outils. Dans cet environnement, *Hadoop* fait office de logiciel *open source* de référence pour exploiter les grandes données. En parallèle, des systèmes tels que *Bigtable*, *Hbase*, etc. sont grandement déployés dans les infrastructures *cloud* (privées et publiques) (Liebowitz, 2015). Ces systèmes permettent l'analyse et le traitement de petaoctets de données et de milliers de requêtes en continu. La notion de péremption de la donnée prend toute son importance dans le cadre de ces analyses en temps réel et il convient de définir les intervalles de temps au sein desquels la donnée analysée est valable. Le management de la donnée évolutive apparaît comme une priorité des SAD à venir. L'apparition de systèmes de management des bases de données évolutives

(*Database management system*) traduit cette volonté de faire muter tant les applications permettant d'exploiter des données que les SAD orientant les prises de décision. Au niveau du processus décisionnel du SAD, l'impact des *big data* se ressentira de la phase d'identification du problème jusqu'à la proposition de différents cours de l'action par le système. Le décideur conservera le choix de mettre en œuvre les propositions ou non.

Le TADSS (Time aware decision support system)

Afin d'illustrer l'impact des *big data* sur la prise de décision en situation et l'intérêt de la prise en compte de la temporalité, nous pouvons prendre en exemple le *TADSS (time aware decision support system)*. Cet outil a été déployé dans un domaine très incertain au sein duquel la prise de décision intuitive prévaut sur la prise de décision analytique : les places financières (Milea et al., 2013). Nous pourrions imaginer intégrer ce système au processus décisionnel des équipes opérationnelles de pompiers ou de police dans le cadre de la prise de décision en situation.

Les différentes équipes opérationnelles peuvent en effet agir différemment face à une même situation puisqu'elles se basent sur des données différentes pour fonder leurs raisonnements. La mise en place du *TADSS* permettrait de prendre en considération des données en temps réel, l'historique des situations mais aussi la dimension temporelle des environnements et des changements ayant lieu. Cette approche évalue les caractéristiques des systèmes experts tout en prenant en compte les applications temporelles séparément. Elle construit ensuite un cadre de travail liant ces deux démarches. Le langage ontologique *tOWL*, supporté par le système, permet de réaliser des liaisons temporelles entre les éléments (par rapport à des intervalles de temps). Ce modèle est une extension des réseaux Bayésien. Des nœuds temporels sont intégrés à la représentation, les événements se voient attribuer un état et des causalités temporelles sont mises en avant entre les nœuds (Milea et al., 2013). Les recommandations ne sont plus limitées dans le temps et ne se périment pas au fil des changements de situations.

Le *TADSS* pourrait, comme nous l'avons vu, avoir un impact très important sur les prises de décisions des équipes opérationnelles de police, qui sont d'ailleurs de plus en plus en contact avec les technologies *big data* (Van Rijmenam, 2014). Nous allons à présent tenter d'observer les bénéfices qu'aurait pu apporter cet outil si intégré au processus décisionnel appliqué dans le cadre d'une mission de police réelle :

En 2012, un événement organisé par une radio des Pays Bas a pu être encadrée par la police néerlandaise : l'afflux non programmé de 500 000 personnes en 6 jours dans la ville d'Enschede. Trois outils ont été nécessaires pour garantir la sécurité de manière anticipée :

<i>Twitcident</i>	<i>Urbanshield</i>	<i>Blue Mark</i>
Analyse en temps réel des tweets locaux : 1,1 milliards de tweets furent scannés.	Analyse des données GPS, cartographies du positionnement des équipes terrain	Géolocalisation des smartphones de la foule et analyse des mouvements
Coordination en temps réel des actions par la <i>Crowd control room</i>		

Cette combinaison de systèmes illustre assez bien les quantités de données qu'ont été amenés à analyser en temps réel les décideurs (ici la *Crowd control room*) dans le cadre de cette mission qui a fonctionné selon le modèle suivant :

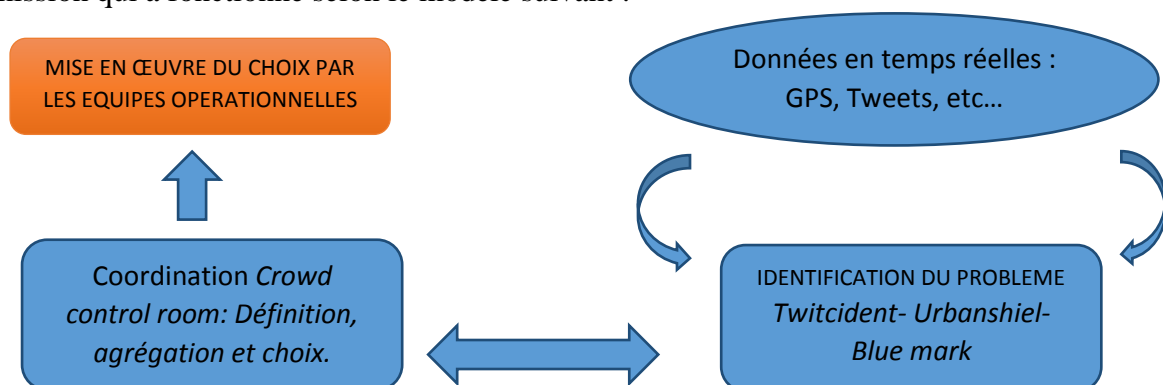


Figure 3: Processus décisionnel-Police Néerlandaise

Nous pouvons cependant imaginer aller encore plus loin et intégrer en sus des outils précédemment cités les éléments composant le système *TADSS* afin de permettre la prise de décision en reliant des valeurs à des intervalles de temps :

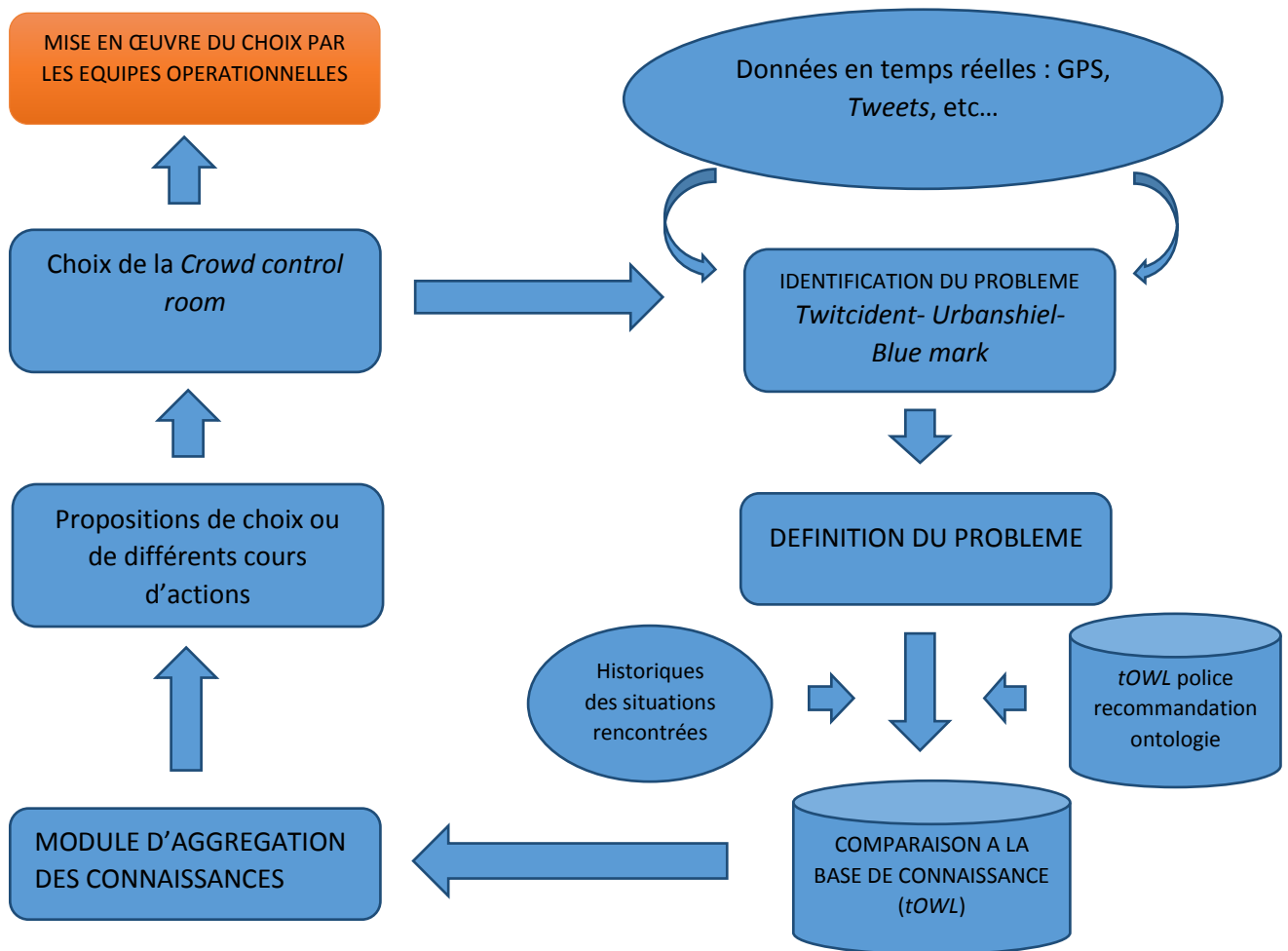


Figure 4: Le TADSS intégré au processus décisionnel de la police Néerlandaise

Le TADSS aurait dans le cadre de cette mission de police notamment permis de comparer les problèmes définis en temps réel grâce aux outils d'analyse (*Twitcident*, *Urbanshield*, *Blue Mark*) à des situations rencontrées par le passé en fonction de contextes similaires. Cet outil aurait permis d'alléger la charge portant sur l'organe de coordination de la mission (*Crowd control room*) et d'augmenter la rationalité des choix proposés. Le TADSS ne peut être intégré au processus décisionnel de la police Néerlandaise sans une refonte totale préalable de l'architecture de la prise de décision et l'intégration de nouveaux outils adaptés (*Twitcident*, *Blu mark*, etc.) pour traiter les données non structurés (provenant notamment des réseaux sociaux). L'action est ici conduite en fonction des informations circulant sur les réseaux sociaux et le *big data* est de fait l'élément central de la prise de décision.

2.3. La mesure de la performance du SAD

Elle peut être jugée en fonction du processus décisionnel ou du résultat (Lebraty, 2006). Lorsqu'en lien avec le processus décisionnel, la rationalité (du processus) est un critère à prendre en compte. A ce sujet nous avons expliqué précédemment que l'avantage des systèmes d'aide à la décision s'appuyant sur des outils *big data* résidait dans leur faculté à prendre en compte des modèles globaux et à orienter vers des décisions plus rationnelles que celles issues des raisonnements humains. La réalisation d'une argumentation formelle permet d'améliorer la rationalité d'une décision. De ce fait, l'interface des SAD doit rendre accessible l'information en évitant de mettre en avant des calculs mathématiques complexes : les cartes d'arguments peuvent être utilisées dans le cadre de raisonnements semi-automatisés. L'argumentation permet de sensibiliser au contexte, de donner une légitimité à la décision. Elle renforce en sus l'évaluation des individus vis-à-vis des informations à leur portée.

Lorsque la performance est évaluée en fonction du résultat, le retour sur investissement (*ROI*) apparaît comme un indicateur clé de cette mesure. Beaucoup de compagnies pensent que la mise en place d'une stratégie *big data* représente un coût important sans garantie de résultat : elles estiment que le *ROI* du projet envisagé sera faible voir nul. Ces projets se révèlent cependant payants pour la plupart des entreprises et seules 2% d'entre elles se retrouvent face à un échec total (Sweeney research, 2014). Le principal coût relatif à la mise en place de ces stratégies est lié aux ressources humaines : les *data scientist* sont en effet très coûteux (Davenport, Patil, 2012) et de plus en plus demandés (Wixom et al., 2011).

Les gains que devraient permettre le *big data* seront très importants pour les organisations dans un futur proche mais les secteurs ne seront pas impactés de la même manière. L'usage des technologies *big data* peut permettre une augmentation de 60% des marges dans le secteur du détail et une baisse de 50% des coûts d'assemblages et de développement dans la manufacture (Manyika et al., 2011). Les projets *big data* permettent de réduire le temps alloué à la réalisation d'une tâche ou d'un processus, ils sont de ce fait source de rentabilité pour les organisations (Brasseur, 2013).

Afin d'illustrer concrètement les gains induits par les projets *big data*, nous pouvons prendre en exemple la chaîne de magasins Macy's. Cette société a fait le choix de mettre en place une application afin d'optimiser le prix de ses marchandises (en fonction de la météo, de l'approvisionnement, etc...). Les prix des 73 millions d'articles peuvent être dorénavant

optimisés en 1 heure contre 27 heures avant l'implantation de l'application. Macy's économise ainsi 70% sur le coût du matériel. Le temps libéré peut enfin être utilisé pour répondre plus rapidement aux événements imprévus (Davenport, Soulard, 2014).

Globalement, le *ROI* des projets *big data* est très élevé. Le département policier d'une grande ville d'Amérique du Nord a par exemple lancé un projet d'analyse prédictive (fusion de la base de données locale avec une base nationale recensant les crimes). Les résultats du projet ont été une baisse nette de la criminalité et un *ROI* de 863% (Nucleus Research, 2012). *Nucleus Research* conclut son étude par le fait que le *ROI* d'un projet *big data* est en moyenne supérieur de 241% à un projet de *business intelligence* classique. Un euro investi dans un projet *big data* rapporte environ 9 euros. Le délai de remboursement est par ailleurs beaucoup plus faible que pour les projets classiques d'informatique décisionnelle (6 mois contre 27 mois).

3. Conclusion et futur des SAD

Cette communication a pour but de répondre à la problématique suivante : Les SAD actuels sont-ils adaptés aux environnements *big data* ? Plusieurs éléments tels que la nécessité de pouvoir traiter des contenus non structurés et la prise en compte de la temporalité de la donnée indiquent que les systèmes d'aide à la décision classiques ne sont que peu adaptés aux environnements du *big data*. Les architectures décisionnelles au sein des entreprises doivent donc être redéfinies pour tirer le meilleur profit des stratégies *big data*. Le *TADSS* est un exemple d'outil pouvant être intégré aux SAD classiques afin de considérer les *big data* et ce, sans imposer un remplacement complet des systèmes existants. Les outils de ce type devraient séduire sans problème les entreprises françaises qui restent frileuses à faire du *big data* l'élément central de la prise de décision (modèle disruptif). Cette méfiance tient son fondement dans les coûts imposés par le renouvellement de leurs systèmes décisionnels. Nous avons cependant pu constater que le retour sur investissement des projets *big data* est très élevé et que les délais de remboursement sont bas.

L'association de l'homme et des *big data* dans le cadre des prises de décisions pourrait former un magnifique partenariat (Brasseur, 2013). Il faut pour cela inventer de nouvelles structures organisationnelles permettant de tirer profit des points forts de la machine pour pallier les

points faibles de l'Homme (Andrew McAfee, 2012). Le crime et la corruption coûtent mondialement 2 trilliards de dollars par an tandis que les embouteillages pèsent 7 à 8 milliards de dollars dans le seul budget du Royaume Unis (Helbing, 2014b). Face à toutes ces pertes, un algorithme capable d'apporter une amélioration de tout juste 1% des décisions amènerait un bénéfice immense pour les sociétés. Or, la promesse du *big data* va bien au-delà de cet unique point de pourcentage (Helbing, 2014a). Le développement des objets connectés devrait lui-même influencer fortement les SAD, renforçant la capacité de ceux-ci à proposer des cours d'actions temporels : 30 milliards d'appareils connectés (The economist., 2014) et 50 milliards de capteurs (New Vantage Partners, Davenport, 2014) seront éparpillés à travers le monde en 2020.

Le *big data* et notamment l'utilisation des données qu'il contient soulève cependant de nombreux questionnements éthiques. A l'heure où Me-trics, une start-up établie au Michigan, tente de développer un outil permettant de collecter et de superviser les données personnelles d'utilisateurs dans presque la totalité des aspects de leurs vies (santé, humeur, budget, forme physique, activité en ligne, etc...), il est de plus en plus question de définir un cadre d'utilisation du *big data* (Davenport, Soulard, 2014). Les gouvernements ne possédaient jusque-là pas les données nécessaires pour contrôler tous les pans des sociétés mais cela est en train de changer. Chaque nouveau capteur intégré à nos appareils est un pas de plus vers la connaissance ultime du comportement des individus. Ils sont les relais de nos intérêts, de nos passions, etc... Les ordinateurs exécutent dorénavant 70 % des transactions financières tandis les premières *hotlines* intégralement informatisées sont opérationnelles (ex : *Hotline IBM Watson*). Les ordinateurs sont plus performants que les hommes dans beaucoup de domaines : faut-il craindre de voir ceux-ci faire un jour de meilleurs policiers, avocats ou politiciens que nous ? Ce n'est pas de la science-fiction : un système informatique capable de simuler toutes les actions et interactions entre humains est envisageable à moyen terme. Il pourrait équiper ces milliards d'agents qui posséderaient dès lors des capacités cognitives et une forme d'intelligence (Helbing, 2014a).

Il ne faut toutefois pas avoir peur des grands bouleversements (les grandes innovations ont toujours été vectrices de peurs, à l'instar par exemple de la télévision) mais les accompagner et s'adapter à eux. L'informatique et les *big data* sont la troisième révolution industrielle et celle-ci va s'amplifier pendant encore plusieurs décennies (Andrew McAfee, 2012). Les ordinateurs font actuellement des percées dans des univers nouveaux et nous assistons à une véritable partie d'échecs, avec d'un côté l'homme et de l'autre la technologie (Levy, Murnane,

2013). Nous devons rester en alerte constante vis-à-vis des interférences que peut provoquer la machine vis à vis de nos intuitions (Klein, 2004). Il faut enfin imposer aux programmeurs de développer des technologies de l'information à des fins de support et non de domination.

BIBLIOGRAPHIE

ALTER, Steven et al, 1977. A taxonomy of decision support systems. In : *Sloan Management Review*. 1977. Vol. 19, n° 1, p. 39–56.

ANDREW MCAFEE, Erik Brynjolfsson, 2012. Big data: the management revolution. In : *Harvard business review*. 2012. Vol. 90, n° 10, p. 60-6, 68, 128.

ANON, 2014. The internet of nothings. In : *The Economist* [en ligne]. 10 mai 2014.
[Consulté le 3 février 2015]. Disponible à l'adresse :
<http://www.economist.com/blogs/babbage/2014/05/difference-engine-1>.

BALAZINSKA, Magdalena, DESHPANDE, Amol, FRANKLIN, Michael J., GIBBONS, Phillip B., GRAY, Jim, HANSEN, Mark, LIEBHOLD, Michael, NATH, Suman, SZALAY, Alexander et TAO, Vincent, 2007. Data management in the worldwide sensor web. In : *IEEE Pervasive Computing*. 2007. Vol. 6, n° 2, p. 30–40.

BRASSEUR, Christophe, 2013. *Enjeux et usages du big data: technologies, méthodes et mise en oeuvre*. Paris : Hermès science publications-Lavoisier. ISBN 9782746245204 2746245205.

COINTOT, Jean-Charles et EYCHENNE, Yves, 2014. *La révolution big data: les données au cœur de la transformation de l'entreprise*. Paris : Dunod. ISBN 9782100711420 2100711423.

DAVENPORT, Thomas H. et PATIL, D. J., 2012. Data scientist. In : *Harvard business review*. 2012. Vol. 90, p. 70–76.

DAVENPORT, Thomas H et SOULARD, Hervé, 2014. *Stratégie Big Data*. Paris : Pearson. ISBN 9782744066177 2744066176.

GEORGE, Gerard, HAAS, Martine R. et PENTLAND, Alex, 2014. Big data and management. In : *Academy of Management Journal*. 2014. Vol. 57, n° 2, p. 321–326.

GORRY, George Anthony et MORTON, Michael S. Scott, 1971. *A framework for management information systems* [en ligne]. S.l. : Massachusetts Institute of Technology.

[Consulté le 3 février 2015]. Disponible à l'adresse :
http://cpe.njit.edu/dlnotes/MIS645/Frame_Management_Info.pdf.

HELBING, Dirk, 2014a. Crystal Ball and Magic Wand-the Dangerous Promise of Big Data (Chapter 3 of Digital Society). In : *Available at SSRN* [en ligne]. 2014. [Consulté le 3 février 2015]. Disponible à l'adresse : http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2502561.

HELBING, Dirk, 2014b. ID 2501356 : *Big Data Society: Age of Reputation or Age of Discrimination?* [en ligne]. SSRN Scholarly Paper. Rochester, NY. Social Science Research Network.
[Consulté le 3 février 2015]. Disponible à l'adresse : <http://papers.ssrn.com/abstract=2501356>.

KAHNEMAN, Daniel, 2012. *Thinking, fast and slow*. London : Penguin Books.
ISBN 9780141033570 0141033576.

KLEIN, Gary A, 1999. *Sources of power how people make decisions* [en ligne]. Cambridge, Mass. : MIT Press. [Consulté le 3 février 2015]. ISBN 0585078173 9780585078175. Disponible à l'adresse : <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=9353>.

KLEIN, Gary A, 2004. *The power of intuition: how to use your gut feelings to make better decisions at work*. New York : Currency/Doubleday. ISBN 0385502893 9780385502894.

LEBRATY, Jean-Fabrice, 2006. Les systèmes décisionnels. In : *Encyclopédie de l'informatique et des systèmes d'information*. 2006. p. 1338–1349.

LEVY, Frank et MURNANE, Richard J., 2013. Dancing with robots: Human skills for computerized work. In : *Washington, DC: Third Way NEXT*. 2013.

LIEBOWITZ, Jay, 2015. *Bursting the big data bubble the case for intuition-based decision making* [en ligne]. S.l. : s.n. [Consulté le 3 février 2015]. ISBN 9781482228878 1482228874. Disponible à l'adresse : <http://www.crcnetbase.com/isbn/9781482228878>.

MANYIKA, James, CHUI, Michael, BROWN, Brad, BUGHIN, Jacques, DOBBS, Richard, ROXBURGH, Charles, BYERS, Angela Hung et INSTITUTE, McKinsey Global, 2011. Big data: The next frontier for innovation, competition, and productivity. In : [en ligne]. 2011.
[Consulté le 3 février 2015]. Disponible à l'adresse : <http://kedeandig.talk4fun.net/y/big-data-the-next-frontier-for-innovation-competition-and-productivity-by-james-manyika-michael-chui-brad-brown-jacques-ughin-richard-dobbs-charles-roxburgh-angela-hun-jacques-ugin-and-angela-hung-byers.pdf>.

- MEEHL, Paul E., 1959. Some ruminations on the validation of clinical procedures. In : *Canadian Journal of Psychology/Revue canadienne de psychologie*. 1959. Vol. 13, n° 2, p. 102.
- MILEA, Viorel, FRASINCAR, Flavius et KAYMAK, Uzay, 2013. A general framework for time-aware decision support systems. In : *Expert Systems with Applications*. 2013. Vol. 40, n° 2, p. 399–407.
- NEW VANTAGE PARTNERS et DAVENPORT, Thomas, 2014. *Big data executive survey*. S.l.
- NUCLEUS RESEARCH, 2012. *Market focus report the value for social and CRM*. S.l.
- SCHMARZO, Bill et BALAND, Marie-Cécile, 2014. *Big data tirer parti des données massives pour développer l'entreprise*. Paris : First Interactive. ISBN 9782754059787 2754059784.
- SIMON, Herbert A., 1955. A behavioral model of rational choice. In : *The quarterly journal of economics*. 1955. p. 99–118.
- SIMON, Herbert A. et NEWELL, Allen, 1971. Human problem solving: The state of the theory in 1970. In : *American Psychologist*. 1971. Vol. 26, n° 2, p. 145.
- SWEENEY RESEARCH, 2014. *The big and small of big data in Australia*. S.l.
- VALDURIEZ, Patrick, 2014. Benefits and risks of using cloud and big data are analyzed at CMM|CMM. In : [en ligne]. 2014. [Consulté le 3 février 2015]. Disponible à l'adresse : <http://www.cmm.uchile.cl/?p=22543>.
- VAN RIJMENAM, Mark, 2014. *Think bigger: developing a successful big data strategy for your business* [en ligne]. S.l. : s.n. [Consulté le 3 février 2015]. ISBN 9780814434161 0814434169. Disponible à l'adresse : <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=686831>.
- WIXOM, Barbara, ARIYACHANDRA, Thilini, GOUL, Michael, GRAY, Paul, KULKARNI, Uday et PHILLIPS-WREN, Gloria, 2011. The current state of business intelligence in academia. In : *Communications of the Association for Information System*. 2011. Vol. 29, n° 16, p. 299–312.

Table des matières : *big data* : Un catalyseur des systèmes d'aide à la décision ?

Résumé:	1
INTRODUCTION :	2
1. PRISE DE DECISION ET SAD	3
1.1. Les trois modèles principaux de décision	3
1.2 Quels SAD pour quels types de décision ?	5
2. LES SAD ACTUELS EN ENVIRONNEMENT <i>BIG DATA</i> ?	7
2.1. L'infrastructure nécessaire à la prise en compte du <i>big data</i>	7
2.2. La prise en compte du <i>big data</i> par les SAD : le TADSS	10
Le TADSS (Time aware decision support system)	11
2.3. La mesure de la performance du SAD	14
3. Conclusion et futur des SAD	15
BIBLIOGRAPHIE	17

Table des illustrations : *big data* : Un catalyseur des systèmes d'aide à la décision ?

Figure 1: Interface Watson - Diagnostics médicaux	7
Figure 2: Interface perspectives-recommandations (Bill Shmarzo, 2014)	10
Figure 3: Processus décisionnel-Police Néerlandaise	12
Figure 4: Le TADSS intégré au processus décisionnel de la police Néerlandaise	13