Revue COSSI

N° 4 | 2018
Méthodes et stratégies de gestion de l'information par les organisations : des big data aux thick data

Faire parler ses données : de la masse à la substance, le sens en question

Francis BEAU *Chercheur*
DeVisu
University of Valenciennes and Hainaut-Cambresis

Édition électronique :
URL :
https://revue-cossi.numerev.com/articles/revue-4/72-faire-parler-ses-donnees-de-la-masse-a-la-substance-le-sens-en-question
DOI : numerev_1606
Date de publication : 04/09/2018

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : BEAU, F. (2018) Faire parler ses données : de la masse à la substance, le sens en question. *Revue COSSI*, (4). https://doi.org/10.34745/numerev_1606

Les progrès des outils informatiques et des réseaux permettent l'accès à des masses considérables de données (big data) dont le traitement semble devoir se prolonger pour apporter de la substance (thick data) à des données au préalable affinées (smart data). Afin de légitimer une stratégie de gestion de l'information originale dont nous présentons le principe appliqué à une méthode opérationnelle d'organisation des connaissances dans une mémoire collective, nous nous appuyons sur l'observation des processus de construction de sens, pour étudier la nature des traitements associés à ces différentes approches de la notion de data. Puisant nos arguments dans les humanités plus que dans le génie numérique, nous montrons que, du cumul des données à l'agrégation des savoirs en passant par la représentation des connaissances, le sens mû par une volonté et déterminé par un besoin d'agir, joue en effet un rôle essentiel dans le fonctionnement d'un système d'information.

The progress of computer tools and networks enable access to massive amounts of data (big data) whose processing seems likely to be extended in order to bring substance (thick data) to data previously refined (smart data). In order to legitimate an original information management strategy as the one we present the principle applied to an operational method of knowledge management in a collective memory, we rely on observation of building sense processes, to study the nature of the processes associated with those different approaches to this notion of data. Drawing our arguments from humanities better than from digital engineering, we show as a matter of fact that, from accumulation of data to aggregation of knowledge passing by representation of cognition, sense moved by will and determined by need for action, play an essential role in the operation of an information system.

Mots-clés :
Data, Numérique, Mémoire, Connaissance, Savoir, Sens, Analogique, Facettes, Data, Digital, Knowledge, Cognition, Sense, Memory, Analogical, Facets

---

INTRODUCTION

Comment concevoir un système d'information permettant la mise en œuvre d'une véritable stratégie de gestion de l'information qui fasse sens des données recueillies ? Pour poser le problème, il faut passer par des considérations théoriques, à commencer par le vocabulaire utilisé que l'on doit bien préciser : la mémoire et tous ses ingrédients qui participent à la construction du sens, puis les notions d'épistémè chez Foucault ou d'épistêmê aristotélicienne et de doxa qui en éclairent le fonctionnement. Mais cette étude très théorique n'aurait à son tour aucun sens si elle n'obéissait pas à une volonté, qui s'exprime dans des applications, celles-là bien pratiques, dont les métiers d'exploitation de l'information éprouvent un besoin croissant.

Nos développements théoriques ont donc pour première ambition de légitimer l'originalité d'une démarche essentiellement méthodologique qui va ainsi à contre-courant d'une tendance générale à la suprématie incontestée de la technique. Pour mettre en place une stratégie de gestion de l'information efficace, il nous a semblé en effet important de redonner toute sa place aux sciences humaines dont les sciences de l'information et de la communication relèvent, dans un paysage assurément trop dominé par des technologies de l'information et de la communication en pleine effervescence, qui relèvent quant à elles des sciences dites exactes, sciences de l'ingénieur ou « sciences du calcul ». Résolument théoriques, mais fondés sur l'expérience, nos travaux s'inscrivent ainsi dans une démarche foncièrement pragmatique d'adaptation des pratiques aux progrès technologiques, qui revendique toutefois la volonté de ne rien abandonner à une technique de plus en plus hégémonique.

PROBLÉMATIQUE : INFORMATION ET DONNÉE, UN MÊME OBJET, DEUX POINTS DE VUE DISTINCTS

L'information n'existe pas en tant que telle si elle n'est pas effectivement reçue. Pour l'esprit qui la reçoit, elle est connaissance, et vient modifier son savoir implicite ou explicite. (Meyriat, 1985)

Comment concevoir une stratégie de gestion de l'information efficace faisant sens de données recueillies massivement ? Autrement dit comment passer de la donnée reçue à l'information utile pour la décision dans l'action ? Ou encore « comment transformer les torrents de données en fleuves de connaissances ? » (Lévy, 2015). Avant toute tentative de réponse à cette question, il s'avère nécessaire de bien distinguer la notion d'information de celle de donnée.

Il faut, entend-on souvent, « donner un sens aux données » (Blais & Martineau, 2006). Un trait caractéristique du concept de data, qui s'impose comme un leitmotiv à la lecture, tant de la littérature scientifique que de la presse généraliste ou technique, tient dans l'expression « faire parler ses données » (Allard, 2015). Quelle que soit la forme qu'elle peut prendre au fil des pages (faire sens de ses données, donner du sens

à ses données, tirer du sens de ses données…)[1], la question du sens semble donc bien centrale en matière de traitement des données, au point de transformer la perception que l'on a de leur silhouette, qui passerait de « grosse » à « épaisse »[2] en donnant de la consistance ou de la substance à la masse disponible.
La notion de sens s'inscrit dans un triptyque :

- les 5 sens, auxquels on ajoutera l'intuition qui enclenche le passage du sensible à l'intelligible (les capteurs ou instruments de collecte des données, en réception) ;
- le sens de la flèche (le processus de traitement des données, de conception ou de conceptualisation, passage du sensible à l'intelligible ou du sentiment à l'idée, puis du concept au symbole et de l'opinion au discours) ;
- et la signification (le produit de ce traitement, c'est-à-dire le discours ou un ensemble de signaux porteurs de sens, en émission).

La donnée passerait donc de « grosse » (big data) à « épaisse » (thick data), tout au long d'un processus de construction de sens qui passe par un stade intermédiaire de conceptualisation, indiquant paradoxalement plus de finesse ou « d'intelligence » (smart data).

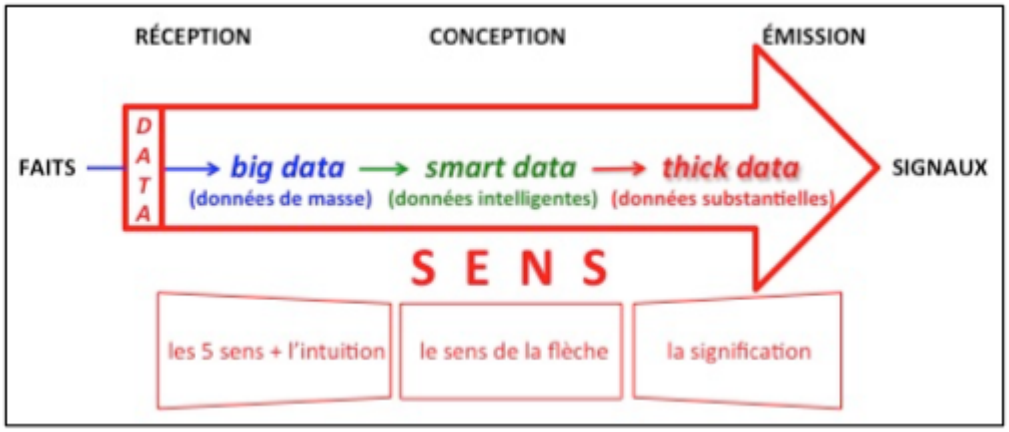


Figure 1. Le processus de construction de sens.

De tels adjectifs utilisés pour qualifier les données traduisent bien, l'énormité des gisements de données auxquels nous sommes désormais confrontés, ainsi que les immenses capacités de calcul dont nous disposons pour les traiter. Mais le recours même à ces épithètes, qui répond à un besoin de distinguer bien concret, conduit à s'interroger sur la terminologie en usage pour désigner d'un même nom des réalités sensiblement différentes. La donnée change en effet de statut selon la nature du traitement qui lui est appliqué. Allant de l'observation sélective et du calcul algorithmique portant sur de simples traces numériques en quantité massive (big), celui-ci fait place à un traitement plus fin (smart) semblable à celui de la pensée émanant de l'intuition et de l'interprétation analogique via la visualisation d'une idée, pour évoluer jusqu'à l'énonciation linguistique d'un jugement de nature dialectique par

agrégation de données (thick) plus substantielles (voir plus loin figure 3).
Le terme « donnée » n'est en réalité, dans l'usage courant, qu'une formulation particulière de ce concept d'information si difficile à définir : dès lors qu'elle a été massivement recueillie (big data), la donnée se transforme méthodiquement, pour générer une information plus substantielle (thick data) susceptible d'être communiquée. Le lieu de cette transformation est un système d'information et de communication ou une mémoire collective, dont nous allons examiner le fonctionnement que l'on peut rapprocher de celui de notre mémoire individuelle. On y retrouve en effet à une échelle macroscopique, sous le terme générique d'information, la donnée en entrée du système, puis la connaissance en perpétuelle transformation dans le système, et le savoir en sortie, produit de la rencontre d'une connaissance avec un sujet qui donne sens à son action en l'éclairant.
Dans une logique de communication appliquée à une mémoire humaine impliquant un émetteur (donateur) et un récepteur (donataire), quelle que soit sa position dans le processus de construction de sens, et quel que soit son mode de traitement, automatique (calcul), intellectuel (idée) ou logique (jugement), on peut comprendre que l'information soit considérée par le donataire/récepteur à qui celle-ci est « donnée », comme une « donnée » du problème qui se pose à lui. Mais ce donataire/récepteur devient progressivement concepteur en donnant sens à la donnée pour en faire une connaissance intelligible puis un savoir, en lui donnant de la consistance ou de « l'épaisseur » dans sa mémoire. L'information qui lui est donnée change ainsi de statut au fur et à mesure de son traitement : elle prend sens en devenant connaissance, puis savoir, acquérant ainsi de la consistance ou de « l'épaisseur » dans sa mémoire.
Les spécialistes des transmissions ou des « télécommunications », terme révélateur du caractère technicien de l'approche applicative souvent choisie pour aborder les problèmes de traitement de l'information et de communication, s'intéressent au signal et à son conditionnement physique pour une transmission optimale par des machines. Les informaticiens s'intéressent quant-à-eux aux données et à leur traitement automatique par des machines destinées à suppléer l'homme dans toutes les tâches accessibles à des opérations purement calculatoires. Ce sont tous des ingénieurs dont les préoccupations théoriques répondent à des motivations éminemment techniques : leurs applications pratiques sont des machines destinées à augmenter artificiellement les performances individuelles. Les recherches en intelligence artificielle connaissant un essor nouveau avec le phénomène des big data, tout comme la cybernétique, placée sur le devant de la scène médiatique depuis que l'interconnexion des ordinateurs en réseau a fait la fortune du préfixe « cyber » en réalisant un espace d'information et de communication mondial que l'usage a consacré sous le nom de cyberespace, relèvent d'une même approche très technicienne des systèmes d'information et de communication.
Tandis que l'ingénieur se préoccupe des données et de la manière d'en traiter (informatique) ou d'en transmettre (télécoms), le plus possible, le plus vite possible, le plus loin possible et au moindre coût, le chercheur en sciences de l'information et de la communication s'intéresse quant-à-lui à leur transmission, c'est-à-dire à leur transformation, qui répond à un besoin de sens attaché à l'élaboration des connaissances (information) pour aboutir à la restitution d'un savoir (communication).

Son attention se concentre sur le facteur humain, qui joue un rôle majeur dans le fonctionnement d'une mémoire collective. Il s'intéresse en particulier à l'effet produit par le signal chez un sujet (le sens) et à sa transmission, dans le cadre de la construction d'un jeu collectif mettant en scène les différents acteurs du système d'information et de communication (ou mémoire collective). Ses développements applicatifs s'attacheront à l'amélioration des passes entre les joueurs, de la réception des données à la restitution d'un savoir commun en passant par l'élaboration d'une intuition collégiale donnant sens au jeu collectif, plutôt qu'aux progrès des performances techniques demandées aux outils de traitement et de transmission des données.

BIG DATA ET THICK DATA : LA LUNETTE ASTRONOMIQUE ET LA BIBLIOTHÈQUE

Les données d'aujourd'hui correspondent à ce que l'épistémologie des siècles passés appelait les phénomènes. Pour continuer de filer cette métaphore, les algorithmes d'analyse de flux massifs de données d'aujourd'hui correspondent respectivement aux instruments d'observation de la science classique. Ces algorithmes nous montrent des patterns, c'est-à-dire en fin de compte des images. (Lévy, 2015, 83)

En entrée de la mémoire collective qui les traite, certaines données peuvent s'identifier à des ensembles discrets de valeurs exactes accessibles au calcul (data). Ce sont des faits tangibles, fruits de l'observation empirique, dûment répertoriés et catégorisés, c'est-à-dire que leur trace s'inscrit à la fois physiquement et conceptuellement dans une mémoire numérique. Il suffit alors de les accumuler, en confiant leur traitement à des algorithmes conçus pour en tirer de nouvelles informations par le calcul et apporter ainsi des solutions aux problèmes posés par l'incertitude inhérente à la décision dans l'action. Au départ, il y a une multitude de données éparses qui peuvent être des traces d'évènements[3], d'actions ou de pensée, c'est-à-dire l'expression numérique de faits observés, dont le sens global nous échappe du fait de leur éparpillement dans le temps ou dans l'espace, mais dont l'accumulation puis la quantification fait sens. Des algorithmes intelligemment conçus pour ordonner dans des ordinateurs toutes ces données selon une logique correspondant à un besoin de savoir dûment identifié, peuvent alors proposer des solutions statistiques ou calculées qui viennent enrichir nos savoirs pour résoudre nos problèmes de décision dans l'action. Ce que les Anglo-Saxons nomment big data, que nous traduirons en français par « cumul de données », « données cumulées » ou « données de masse », ou encore depuis peu par celui de « mégadonnées » (Brasseur, 2016) c'est ce travail sur la donnée qui va être recueillie, puis soumise au calcul pour proposer de nouvelles informations.

Sans préjuger des progrès des outils algorithmiques dans les années à venir, particulièrement en matière de visualisation des données (Reymond, 2016), d'intelligence artificielle et « d'apprentissage profond » (LeCun, 2016) ou de codage de la signification linguistique (Lévy, 2015), observons qu'il ne s'agit là que de progrès techniques appliqués aux instruments de la connaissance. Il en a été de même par le passé avec la lunette astronomique qui a permis un bond significatif en matière de

recueil des données, ou avec la bibliothèque, en matière de diffusion des savoirs. Dans un monde dominé par la technique, il semble important de ne pas focaliser notre attention sur les seuls instruments que la technologie nous propose (voire parfois nous impose), en s'intéressant d'abord au besoin opérationnel qui donne sens à l'information dans l'action, et à la volonté qui l'anime.

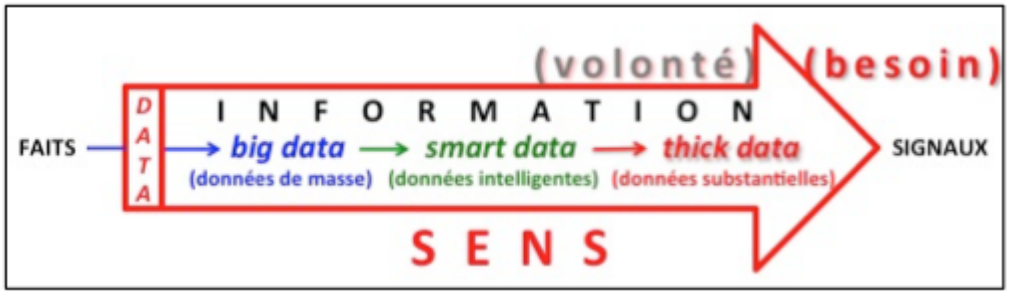


Figure 2. Le besoin et la volonté au service de l'information.

Notre travail repose ainsi sur un principe fondamental imposant d'admettre que le processus de décision dans l'action collective soit toujours mû par une volonté humaine qui fait sens et intervient tout au long des opérations de gestion de l'information, même fortement automatisées. Ce principe incontournable aux allures de postulat, certes peu susceptible d'être établi scientifiquement mais humainement rassurant, nous apparait en effet pour l'heure, comme le plus apte à fonder la conception d'une mémoire collective garantissant une pertinence maximale de l'information en sortie du système. C'est la raison pour laquelle nous avons jugé nécessaire de penser le système d'information autrement que sous l'angle des algorithmes et des seuls outils informatiques, en le considérant sous l'angle du sens et de sa construction dans la mémoire.

Pour fixer le vocabulaire, on peut ainsi procéder par analogie en observant que le Big data est au thick data ce que la lunette astronomique est à la bibliothèque scientifique en astronomie : un instrument de recueil des faits au service de la diffusion du savoir. Le Big data ou l'analyse des données massives constitue le télescope moderne qui permet de voir plus loin, plus large ou tout simplement mieux. Il est ainsi clairement cantonné à l'observation et au recueil des faits (données de masse), tandis que le thick data ou la composition de données substantielles se consacre à la diffusion du savoir qui procède de leur traitement.

Tout comme les livres dans une bibliothèque, les thick data obtenues à l'issue du processus de construction de sens, mû par une volonté elle même animée par un besoin, sont des instruments œuvrant à la diffusion d'un savoir, terme que l'on préfèrera à celui de data popularisé par une informatique omniprésente. À l'heure des technologies numériques désormais incontournables, le document qui demeure plus que jamais cet instrument élémentaire de diffusion du savoir assurant une fonction sociale de transmission de sens entre individus, est amené à jouer un rôle central dans le fonctionnement de notre mémoire collective, comme dans celui de toute bibliothèque

dont il est le composant élémentaire sous forme de livre.

LA CONSTRUCTION DU SENS DANS UNE MÉMOIRE : DE LA DONNÉE À L'INFORMATION

Il paraît que les nombres parlent d'eux-mêmes. Mais c'est évidemment oublier qu'il faut, préalablement à tout calcul, déterminer les données pertinentes, savoir exactement ce que l'on compte, et nommer – c'est-à-dire catégoriser – les patterns émergents. (Lévy, 2015)

Nous nous intéressons donc à la construction du sens dans le système d'information et de communication que nous assimilons à une mémoire. En entrée de notre mémoire collective, d'autres données ne peuvent pas s'identifier à des valeurs exactes, comme peuvent l'être les data accessibles au calcul. C'est en particulier le cas des informations portées par des signaux analogiques, comme de toutes celles directement perçues par nos sens. C'est également le cas de la plupart des informations issues de la pensée (connaissances ou savoirs en forme de jugements), sauf à n'en considérer que les simples traces numérisées (métadonnées documentaires) y-compris celles liées à l'ingénierie linguistique pour la traduction automatique, seules porteuses de valeurs exactes accessibles au calcul numérique.

Le traitement algorithmique ne peut plus alors s'appliquer à ce type de données et doit céder la place à un traitement intellectuel, substituant aux faits une « représentation abstraite » (analogie) de leur « manifestation concrète », qui « fait sens », soit une « information » (Beau, 2017), en s'attachant à nommer – c'est-à-dire catégoriser – les patterns émergents. Dans notre mémoire individuelle, c'est la pensée qui réalise ce traitement en donnant naissance à de nouvelles informations qui ne sont plus seulement le produit d'un calcul numérique, mais celui d'un jugement émanant d'une idée à caractère analogique par essence. Pour distinguer ces informations qui sont des objets traités par la pensée, des données qui sont des objets traités par des algorithmes, nous leur réservons le terme de « connaissance » qui illustre bien, par analogie, cette genèse réalisée par la pensée (cf. figure 3).

De la fonction statistique de traitement des « données de masse » (big data) en entrée, à la fonction documentaire d'agrégation des « données substantielles » (thick data) en sortie, en passant par la fonction cognitive de traitement des « données intelligentes » (smart data), ce travail est celui de la mémoire dont nous avons décrit quelques grandes lignes dans une communication récente portant sur la « construction de sens » dans un « système d'information et de communication » envisagé comme « une mémoire collective » (Beau, 2016). Celles-ci que nous avons introduites aux paragraphes précédents peuvent être résumées dans le schéma de la figure 3 ci-dessous.

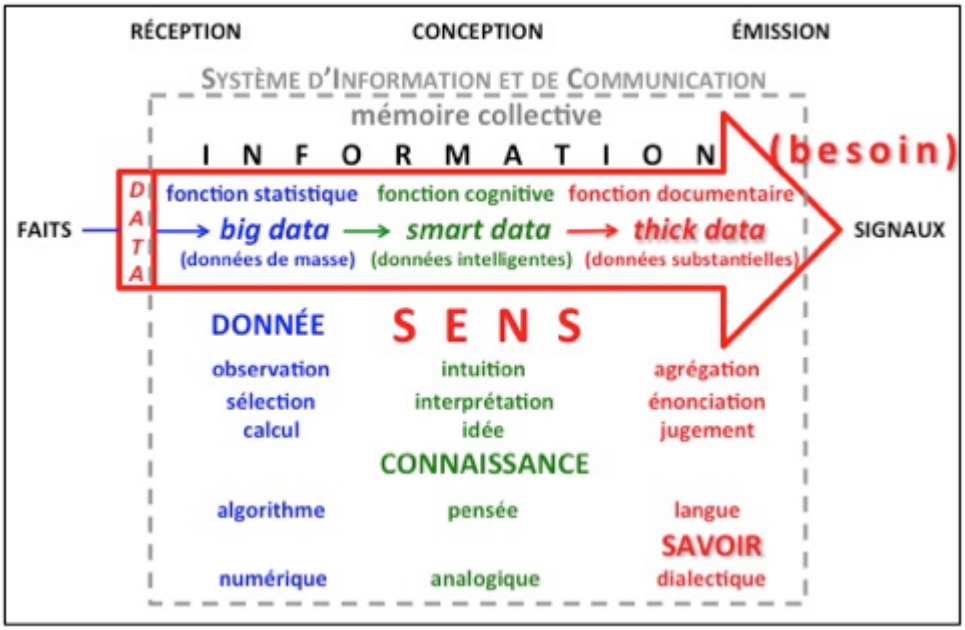


Figure 3. Le processus de construction de sens dans la mémoire.

À partir de données multiples issues du calcul (big data ou données de masse) ou de tout autre type de ressource (observation directe, veille, enquête…), délivrant des données que l'on dira « substantielles » pour traduire le thick data des Anglo-Saxons, recherchées puis recueillies pour rencontrer un besoin de sens dans l'action, une information de synthèse est élaborée. Cette information que l'on dira actionnable pour exprimer le fait qu'elle est utile à l'action passe par l'état de connaissance transformée en savoir pour agir. Pour compléter l'idée que nous nous faisons des notions de connaissance et de savoir dans le système d'information, nous limiterons le sens attribué à la donnée à celui que lui donne Bergson (1888) dans le titre de son « Essai sur les données immédiates de la conscience », qui désigne ce qui est connu immédiatement par observation directe, indépendamment de toute élaboration de l'esprit.

Envisager une mémoire collective uniquement sous l'angle de l'échange de données, qu'elles soient « massives », « intelligentes » ou « substantielles », ce serait ainsi la réduire à sa dimension tacite, c'est-à-dire se priver de toute sa dimension consciente ou explicite. Sans conscience, la mémoire ne serait que technique sans théorie, pratique sans sagesse ou savoir-faire sans savoir. Elle ne serait alors que « ruine de l'âme » pour plagier la formule que Rabelais appliquait à la science. À partir d'une observation et d'un traitement numérique en entrée, dont les performances s'améliorent avec les progrès de la technique, on souhaite donc aboutir en sortie à une information en forme de produit dialectique (un discours) accessible à tous ceux qui pratiquent la langue adoptée par la collectivité.

L'ÉPISTÊMÊ ARISTOTÉLICIENNE ET LES ÉPISTÉMÈS

FOUCALDIENNES

Pour préciser encore ces différentes notions qui interviennent dans la construction du sens, on peut s'appuyer sur nos racines grecques et en particulier avec Aristote sur la notion d'épistêmê, reprise beaucoup plus tard par Michel Foucault (1966) dans « Les mots et les choses » pour désigner la pensée ou le discours scientifique d'une époque.
Tantôt traduite par le mot « science », tantôt par le mot « savoir », la notion d'épistêmê s'applique chez Aristote à la science théorique (theôría), l'observation exercée avec discernement ou sagesse (sophia), prolongée par la science pratique (praxis) exercée avec méthode, sagacité ou prudence (phronesis), puis par la science productive (poïésis) exercée avec habileté ou technique (technè). Dans notre système d'information et de communication (figure 4), on retrouve en entrée, le récepteur qui observe (theôría) avec discernement (sophia) une information qui est une donnée de son problème, mise en œuvre (praxis) par un concepteur qui élabore avec méthode (phronesis) une nouvelle information intelligible, puis en sortie, un émetteur qui produit (poïésis) avec habileté (technè) une information substantielle.

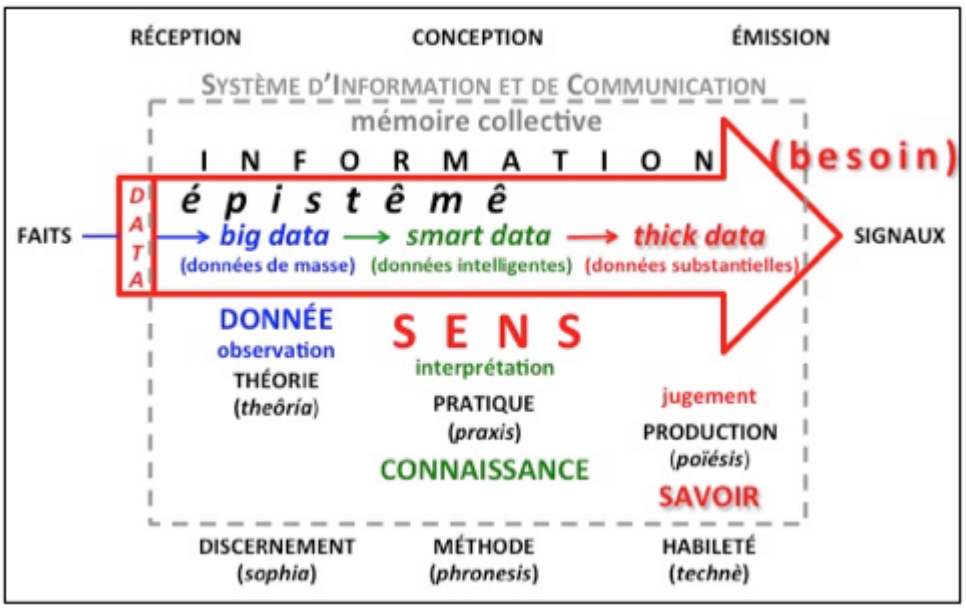


Figure 4. L'épistêmê dans la mémoire collective.

C'est donc tout naturellement sur la méthode que nous avons fait porter notre effort pour améliorer les pratiques. En effet, l'épistémè, entendue au sens que Michel Foucault lui donne dans Les mots et les choses, est illustré par trois grands moments de la culture occidentale : la Renaissance avec l'invention de l'imprimerie, la période Classique avec ses académies, puis l'époque Moderne qu'il craint de voir évoluer vers une disparition de l'homme qui « s'effacerait, comme à la limite de la mer un visage de sable » (Foucault, 1966). Pour éviter que ces craintes ne se concrétisent, nous pensons

qu'il faut dépasser la seule dimension numérique et sa vertu exclusivement technique, qui domine le paysage épistémique actuel, en limitant la technique à son rôle instrumental et en donnant la priorité à la méthode et à son rôle essentiel pour la science.

DOXA ET INTUITION, LANGUE ET PENSÉE, LA QUESTION DU SENS

C'est ainsi, dans l'esprit de cette épistémè foucaldienne, que nous pensons utile d'appeler de nos vœux une période post-moderne redonnant à la méthodologie toute sa place aux côtés de la technologie numérique qui domine la période moderne. Nous nous appuyons pour cela sur une notion complémentaire de l'épistêmê chez Aristote, la doxa, mot que certains auteurs traduisent par « endoxe » (Pelletier, 2007). Pour imager notre propos, nous pouvons dire que l'endoxe est à l'intuition ce que la langue est à la pensée : une réponse analogique à la question du sens.
Chez Aristote en effet, la doxa est une idée partagée par tous (ou la plupart) parce qu'elle répond à une attente (un besoin) des sages, pour faire progresser la théorie avec discernement, la pratique avec méthode et la production avec habileté (Beau, 2016). On peut ainsi dire qu'elle est complémentaire de l'épistêmê (figure 5), c'est-à-dire de la théorie (theôría) et de la sagesse (sophia) associée, qu'elle fait progresser en lui donnant un sens déterminé par un besoin (l'attente des sages), pour mettre en œuvre une pratique (praxis) avec méthode (phronesis), puis réaliser un produit (poïésis) dont une des habiletés (technè) peut être une nouvelle doxa dans une sorte de boucle de rétroaction revenant à l'observation des données disponibles pour améliorer les connaissances pratiques et produire de nouveaux savoirs techniques.

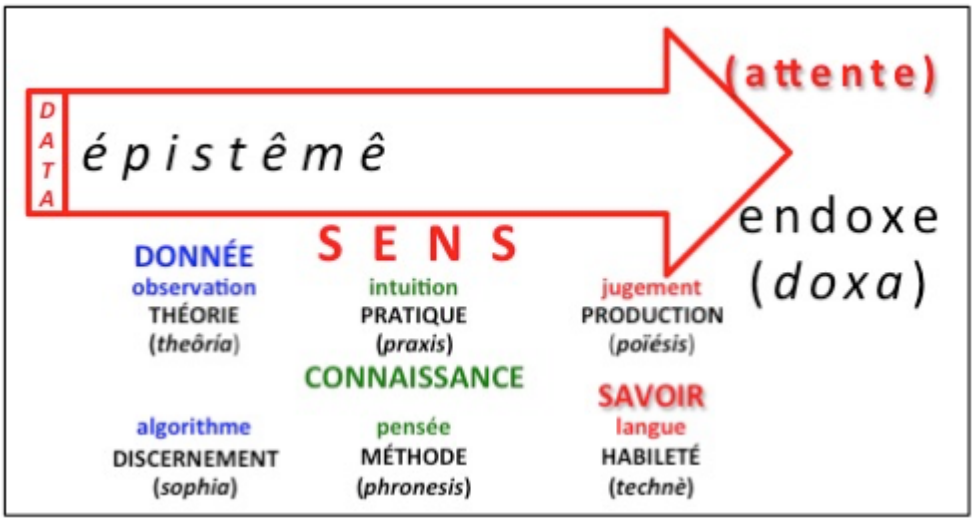


Figure 5. Épistêmê et doxa.

L'endoxe est l'élément central de la construction de sens dans notre mémoire collective. Elle façonne l'intuition collective et le jugement commun, comme la langue conditionne la pensée et le jugement individuels : en réponse à une attente. Face à une masse de données à exploiter en constante augmentation, elle doit faire évoluer l'épistémè actuelle pour permettre de concevoir une connaissance intelligible et émettre un savoir substantiel.

Cette épistémè « moderne » demeure en effet fondée sur une technè numérique omniprésente, dont on observe qu'elle bouscule profondément la fonction documentaire dans nos systèmes d'information au point de placer cette dernière au cœur de ces grandes « révolutions culturelles et cognitives » décrites par Michel Serres (2007), dont celles de l'écriture puis de l'imprimerie et maintenant du numérique. Celles-ci affectent en effet le « couplage entre un support et un message », dont la « quadruple caractéristique (stocker, traiter, émettre et recevoir de l'information) » est « commune aux sciences humaines et aux sciences dures ». Entre une fonction statistique en pleine effervescence et une fonction documentaire à vocation collective de transmission de sens dans un espace-temps de plus en plus contracté par les nouvelles technologies de l'information et de la communication, c'est la fonction cognitive qui est désormais sollicitée en abordant l'exploitation des data comme une « formidable machine à produire du sens nouveau » (Leleu-Merviel, 2004).

On touche là au noyau dur des problèmes d'organisation des connaissances en environnement électronique qui est précisément celui de la transformation des « big data » en « thick data ». La nouvelle épistémè doit ainsi pouvoir reposer sur un corpus méthodologique consistant, dont un des objectifs pourra être d'organiser le fonctionnement d'une mémoire collective. Le document « électronique », adjectif que nous préférons à « numérique » afin de nous démarquer de l'emprise informatique, est amené à y jouer tout son rôle, qui est essentiel dans le processus de construction de sens. Il nous faut donc donner à cette épistémè « post-moderne » toute la consistance de la dimension analogique d'un discours plus dialectique, mais pas moins scientifique pour autant, en revenant avec Aristote à la source de la Grèce antique et à un mode de pensée plus géométrique qu'algébrique. Si les traitements numériques semblent donc parfaitement adaptés au passage de la fonction statistique à la fonction cognitive, et encore loin d'avoir épuisé toutes leurs ressources en la matière, le passage de la fonction cognitive à la fonction documentaire semble quant à lui plus adapté à un traitement de nature analogique.

Pour s'en convaincre, on peut s'appuyer sur ce que l'on observe du fonctionnement de la langue dans la mémoire individuelle. Contrairement au calcul qui fait de la mémoire des ordinateurs un outil numérique dédié au traitement des données, la langue est en effet un outil que l'on peut dire analogique pour indiquer le fait qu'elle procède par analogies, c'est-à-dire en entretenant un rapport de ressemblance avec les objets qu'elle manipule. C'est elle qui conditionne la pensée et fait de notre mémoire spécifiquement humaine un système analogique dédié à l'exploitation d'une information qui s'agrège dans un discours. C'est donc aussi une langue (ou un langage documentaire) qui va faire de notre mémoire collective un système analogique dédié à l'exploitation d'une information qui s'agrège dans des documents assurant une fonction sociale de transmission de sens entre individus, à la fois dans l'espace et dans le temps.

CONCLUSION : ENJEUX ET PERSPECTIVES

Nous devons maintenant nous appuyer sur la puissance de calcul de l'Internet pour « théoriser » (catégoriser, modéliser, expliquer, partager, discuter) nos observations, sans oublier de remettre cette théorisation entre les mains d'une intelligence collective foisonnante. (Lévy, 2015)

Sans aller jusqu'à espérer la « société datacentrique » que Pierre Lévy nous annonce, fondée sur le « médium algorithmique » et l'adoption généralisée d'un métalangage tel que « le langage IEML (ou tout autre système universel de codage computationnel du sens) », « la maîtrise des données, à commencer par leur maîtrise intellectuelle » à l'aide d'une intelligence collective foisonnante, devient sans aucun doute « un enjeu scientifique et social majeur ». Nous pensons néanmoins que les sciences de l'information et de la communication doivent contribuer à remettre de l'humain dans les technologies de l'information et de la communication en proposant une « théorie scientifique » de l'information différente de toutes celles issues de la « théorie mathématique de la communication » (Shannon, 1948), qui permette de concevoir des systèmes d'information documentaires capables « d'optimiser la distribution de l'information pour qu'elle soit appropriée aux besoins de chacun », ce que « nous ne savons pas encore » faire (Babinet, 2016).

Cette théorie doit selon nous se fonder plus sur l'observation des pratiques des hommes que sur celle du développement à marche forcée de techniques numériques de plus en plus performantes au point de laisser croire que l'on pourrait un jour s'affranchir du travail de la pensée. L'intuition collective et le besoin de sens qui l'anime, soutenu par une volonté commune elle-même initiée par des sensations ou des émotions partagées, doivent être selon nous au cœur de toute réflexion théorique en matière de systèmes d'information. Une telle « théorie scientifique » de l'information pourrait marquer l'avènement d'une épistémè post-moderne ou d'une « ère post-numérique » (Beau, 2015), alliant l'ordre et l'harmonie d'une culture analogique classique au génie numérique moderne en apportant à la méthode le soutien indispensable susceptible d'assurer son ascendant sur la technique.

Afin de légitimer une telle approche des systèmes d'information empruntant donc plus aux humanités qu'aux sciences de l'ingénieur, nous travaillons à la formulation des principes d'une méthode de recherche et de partage de l'information, dont l'instrument principal n'est pas le support numérique gestionnaire de données massives, ni l'intelligence artificielle pourvoyeuse de connaissances nouvelles, mais le document électronique recueil de savoirs humains. Son exploitation y est envisagée comme un sport d'équipe, dans un système d'information documentaire à facettes impliquant le lieu et le temps, à l'image des cinq sens complétés par l'intuition qui fédère l'ensemble. L'enjeu est d'améliorer la pratique du système par une communauté organisée autour d'une fonction commune qui donne sens à son jeu collectif. Notre expérience de ce jeu dont le document est le principal instrument, nous a en effet conduit à la réalisation d'un système opérationnel de planification de l'activité documentaire d'une communauté réunie autour d'un besoin de savoir pour exercer sa fonction. Celui-ci est associé à un système « d'indexation analogique » fondé sur un langage documentaire et

une « grammaire » qui l'organise, calquée sur une hiérarchie des sens, dont nous avons pensé utile d'approfondir les fondements scientifiques en puisant à la source des sciences humaines. L'homme étant au cœur du système, c'est en effet, selon nous, bien à ces dernières qu'il revient de théoriser cette expérience de jeu collectif.
À l'heure des technologies numériques et du web sémantique, nous avons donc fait le choix délibéré de faire appel à une théorie de l'information, ancrée dans les humanités de la pensée, de la langue et de la grammaire qui l'organise, plutôt que dans la technique du calcul et des algorithmes qui le programment. Nous voulons affirmer ainsi la nécessité d'une distinction claire entre logique et algorithmique, entre sciences humaines et sciences de l'ingénieur ou entre l'humain et l'automate. Une telle démarcation n'est possible qu'à condition de se donner les moyens d'une interdisciplinarité forte, dont nous ambitionnons de montrer, que les méthodes de partage dynamique de l'information documentaire auxquelles nous travaillons peuvent être l'instrument.
« Malgré leurs progrès constants », nous dit le général Jean Rannou (2015), ancien chef d'état-major de l'armée de l'air française, « les capacités technologiques, n'apportent pas de solution aux analystes » du renseignement « quand les intentions des adversaires restent inaccessible ». Ces progrès, « qui étaient censés suppléer les capacités des hommes, voire les remplacer, leur donnent en réalité une place plus importante en termes de responsabilité et plus exigeante en termes de compétences ». C'est là tout l'enjeu de notre travail : donner aux hommes les moyens méthodologiques pour répondre à des exigences de plus en plus fortes en raison des formidables progrès technologiques auxquels ils ont désormais accès.
Mais si la machine, nous dit encore Cédric Villani (2016), « n'est pas capable de trouver et d'avoir l'intuition de la direction vers où aller, elle va se retrouver piégée par cet océan de possibles ». L'énorme intérêt suscité par le phénomène des big data, qui restent malgré tout en demande « d'épaisseur », nous donne un assez bon exemple de l'insuffisance de ce « tout-technologique » qui domine le paysage actuel. Il justifie en tout cas selon nous cette recherche « d'épaisseur » que le recours à la notion de « thick data » illustre bien et que l'effort méthodologique que nous avons entrepris a l'ambition de satisfaire en donnant du sens aux données grâce à l'intuition qui doit être aux gouvernes pour savoir « vers où aller ».

BIBLIOGRAPHIE

Allard, L. (2015). L'engagement du chercheur à l'heure de la fabrication numérique personnelle. *Hermès, La Revue,* 73,(3), 159-167. http://www.cairn.info/revue-hermes-la-revue-2015-3-page-159.htm.
Babinet, G. (2016). *L'ère numérique, un nouvel âge de l'humanité*. Le Passeur.
Bachimont, B. (2007). Nouvelles tendances applicatives. De l'indexation à l'éditorialisation. Dans Patrick Gros (dir.), *l'indexation multimédia : description et recherche automatiques*, Paris, hermès sciences.
Balslev, K., Saada-Robert, M. (2006). Les microgenèses situées. Unités et procédés d'analyse inductive - déductive. *Recherches qualitatives*, vol. 26, n° 2, 85-109.

Beau, F. (2017). Facettes et système d'information : Une approche de la classification focalisée sur un besoin de savoir pour agir. *Les Cahiers du numérique*, vol. 13,(1), 115-142.
Beau, F. (2016). Aristote et l'épistémè numérique, *Vers une épistémè numérique ? Actes du 19e colloque international sur le document électronique (CIDE19) Athènes*, europia, 1-13.
Beau, F. (2015). Le document numérique au coeur de l'action collective. *Documents et dispositifs à l'ère post-numérique. Actes du 18ème Colloque International sur le Document Électronique (CIDE.18) Montpellier*, europia.
Bergson, H. (1888). *Essai sur les données immédiates de la conscience*, Les Presses universitaires de France, 1970, 144 e édition.
Blais, M. et Martineau, S. (2006). L'analyse inductive générale : description d'une démarche visant à donner un sens à des données brutes. *Recherches Qualitatives – Vol.26(2)*, pp. 1-18.
Bollier, D. (2010). *The promise and peril of big data*. technical report. Washington, D.C. : the aspen institute.
Brasseur, Ch. (2016). *Enjeux et usages du big data : technologies, méthodes et mise en œuvre*. Lavoisier hermès (2ème édition).
Cardon, D. (2012). Regarder les données. *Multitudes 2012/2 (n° 49),* p. 138-142.
Foucault, M. (1966). *Les mots et les choses*. Gallimard, Paris.
Grolier (de), E. (1962). *Étude sur les catégories générales applicables aux classifications et codifications documentaires*, Unesco.
Lecun, Y. (2016). Qu'est-ce que l'intelligence artificielle ? Chaire informatique et sciences numériques (2015-2016), recherches sur l'intelligence artificielle, http://www.college-de-france.fr/site/yann-lecun/recherches-sur-l-intelligence-artificielle.htm.
Leleu-merviel, S. (2004). Effets de la numérisation et de la mise en réseau sur le concept de document, *Revue i3, volume (4) n°1*, 121-140.
Lévy, P. (2015). Le medium algorithmique. *Sociétés*, 129,(3), 79-96.
Lortie, F. (2010). Intuition et pensée discursive : sur la fonction de l'ἐπιβολή dans les ennéades de plotin. *Laval théologique et philosophique, vol. 66, n° 1*, 45-59.
Merzeau, L. (2009). Du signe à la trace : l'information sur mesure, *Hermès, no 53*, 23-31.
Meyriat, J. (1985). Information vs communication ? *in* Laulan, A.-M., *L'Espace social de la communication : concepts et théories*, Paris, Retz-CNRS, 63-89.
Rannou, J. (2015). *Brève histoire du renseignement en France*, CEIS Les notes stratégiques, http://universite-defense-2014.org/fr/system/files/note_strategique_renseignement.pdf.
Serres, M. (2007). *Les nouvelles technologies révolution culturelle et cognitive*, Conférence, Quarante ans de l'INRIA, forum « Informatique et Société », Lille, www.acgrenoble.fr/ien.bourgoinashnord/IMG/pdf_Texte_de_la_conference.pdf.
Shannon, C.E. (1948). A mathematical Theory of communication, *Bell System Technical Journal, Vol. 27* (July, October), 379–423, 623–656.
Paillé, P. (1994). L'analyse par théorisation ancrée. *Cahiers de recherche sociologique*, Numéro 23 (Critiques féministes et savoirs), 1994, 147-181.

Reymond, D. (2016). Visualisation de données. *Introduction, les cahiers du numérique, 4/2016 (vol. 12)*, 9-18.
Vierset, V. (2016). Vers un modèle d'apprentissage réflexif. Recueil de traces d'apprentissage formulées dans les log books des stagiaires en médecine. *Approches inductives*, 3(1), 157-188.
Villani, C. (2016). *Des robots et des hommes, qui vaincra ?* Entretien avec Lila Meghraoua, L'Atelier BNP PARIBAS, https://atelier.bnpparibas/life-work/article/c-villani-robots-hommes-vaincra

[1] (Bollier, 2010), (Paillé, 1994), (Balslev & Saada-Robert, 2006), (Vierset, 2016), (Cardon, 2012).
[2] Traduction littérale de l'anglais thick que l'on pourrait traduire aussi par « concret », au sens de « consistant » (emprunté du latin concretus, « épais, compact », adjectif tiré du participe passé de concrescere, « croître par agglomération, se solidifier ».
[3] Il s'agit de transformer des empreintes en indices, de faire signe avec des traces (Merzeau, 2009).