

**UNE EXPLOITATION DU BIG DATA AU SERVICE DE LA RELATION ELECTEUR-
POLITIQUE**

A BIG DABA EXPLOITATION SERVING THE ELECTOR-POLITICAL RELATIONSHIP

JALAL BOUSSAID

Doctorant à l'Ecole Nationale de Commerce et de Gestion, Université IBN
TOFAIL, Kénitra, Maroc

Jalal.boussaid@encgk.ma

HASSAN AZDIMOUSA

Enseignant chercheur à l'Ecole Nationale de Commerce et de Gestion,
Université IBN TOFAIL, Kénitra, Maroc

Hassan.azdimousa@encgk.ma

Date de soumission : 20/11/2018

Date d'acceptation : 25/03/2019

DOI : <https://doi.org/10.5281/zenodo.3679428>

RESUME

La société contemporaine connaît une évolution rapide par rapport aux technologies de collecte, de traitement et de stockage de l'information. Ces évolutions génèrent des données toujours plus nombreuses, des traitements de plus en plus automatisés, complexes et rapides. L'exploitation de ces masses de données crée de nouvelles connaissances, de nouveaux services et de nouveaux comportements de la société. La terminologie utilisée pour décrire ces phénomènes est souvent « Big data ».

Cet article expose de ce fait, une revue de littérature approfondie sur le système Big data, ainsi que son exploitation pour améliorer la relation électeur-politique. Le cas des élections présidentielles d'Obama est aussi mis en détail. Les processus d'analyse des données du Big data sont aussi présentés et une chaîne de valeur virtuelle pour l'exploitation du Big data pour créer de la valeur est proposée.

MOTS CLES : TIC, Big data, Obama, électeur-politique, chaîne de valeur

ABSTRACT

Contemporary society is evolving rapidly in relation to information gathering, processing and storage technologies. These evolutions generate more data, treatments more automated, complex and fast. The exploitation of these masses of data creates new knowledge, new services and new behaviors of society. The terminology used to describe these phenomena is often "Big data".

This article presents literature review on the Big Data system, as well as its exploitation to improve the elector-policy relationship. The case of Obama's presidential elections is also detailed. Big data data analysis processes are also presented and a virtual value chain for the exploitation of Big Data to create value is proposed.

KEY WORDS: ICT, Big data, Obama, Policy-elect, value chain

INTRODUCTION

De nos jours, le plus grand défi qui se présente pour les organisations n'est pas un manque d'information pour prendre des décisions mais plutôt le contraire, être submerger par une quantité phénoménale de données (Edmunds et Morris, 2000). Les avancées technologiques ont fourni la capacité de stocker efficacement une vaste quantité d'informations, de transformer des données brutes, hétérogènes, difficilement lisibles, en une information visible et compréhensible par les utilisateurs des systèmes d'informations. Lesdits avancés ont permis d'organiser, mettre de l'ordre et rendre accessible des données en leur donnant ainsi un sens (Flichy, 2013). En effet, les évolutions technologiques et informatiques en matière de puissance de calcul ont atteint des niveaux impressionnants tout en devenant de plus en plus abordables et accessibles pour plus d'organisations que tout autre moment de l'histoire humaine (Uma et Ashok, 2016).

Communément appelé « Big data », il se présente comme une catégorie nouvelle et spécifique de systèmes d'information qui profite des progrès de l'informatique et des logiciels analytiques sophistiqués pour aider les entreprises dans le secteur privé à atteindre un avantage concurrentiel sur le marché (McGuire, Manyika et Chui, 2012). Ceci dit, même dans la politique, l'exploitation du Big data procure un avantage compétitif pour les partis politiques (Bille, 2015 ; Chadwick et Stromer-Galley, 2016).

Bien que le concept de Big data soit devenu rapidement un vocable populaire et trop évoqué par les professionnels, les académiciens et le simple citoyen, le terme lui-même est apparu il y a près de quinze ans et a fait écho à la fin des années 2000, lorsqu'il a trouvé sa définition la plus répandue, celle des « 3V » puis celle des « 5V » : **variété, volume, vélocité, véracité et valeur** (Uma et Ashok, 2016). Ce système d'information gigantesque en réseau (Flichy, 2013) continue de subir des raffinages adaptés au même rythme qu'aux valeurs ajoutées qu'il apporte à l'organisation en général et les organisations partisans en particulier. De ce fait, cet article expose une revue de littérature approfondie sur le Big data et ses différents « V ». Le cas des élections présidentielles d'Obama est aussi mis en détail pour examiner en profondeur les éléments clés de réussite de cette campagne. Les processus d'analyse des données du Big data seront aussi décrits. Le processus de création de valeur sera aussi mis en exergue pour pouvoir proposer en dernier lieu une chaîne de valeur virtuelle permettant l'exploitation du Big data afin de créer de la valeur électeur-politique.

1. REVUE DE LA LITTERATURE

La définition de Big Data peut être abordée à partir de plusieurs dimensions (Camacho et al., 2014), (Agrawal et al., 2015), (Bille, 2015) et (Gandomi & Haider, 2015) :

- Les sources de données :

Dans la plupart des cas, les chercheurs voulant définir le Big Data, mettent en exergue sa capacité à collecter, structurer, analyser et traiter des données provenant d'une grande variété de sources, y compris les sites de réseaux sociaux, le Web, les historiques de transactions en ligne et hors ligne, les téléphones mobiles, les tablettes, les ordinateurs et d'autres sources de données. En outre, de nouveaux types de données peuvent être générés à l'avenir, qui doivent être intégrés ou incorporés en douceur avec les pipelines de données existantes. Les systèmes qui ont la capacité de gérer et d'assimiler diverses sources de données font partie intégrante de la définition de Big Data.

- La variété des données :

Le Big Data se caractérise par sa capacité à stocker et intégrer dans une ressource globale et facilement accessible, des données structurées (données des listes électorales, données des hôpitaux, données de gestion des caisses d'allocation familiale ou de l'administration fiscale...) et non structurées (données des réseaux sociaux, des forums ou des blogs...) collectées à partir de plusieurs sources. Le Big data collecte et stocke aussi des données semi-structurées comme le « Extensible Markup Language (XML) » qui est un langage textuel utilisé pour échanger des données sur le Web (Gandomi & Haider, 2015). Grâce au Big Data, la base de données a évolué d'un simple outil documentaire à une plate-forme d'« exploration générative » (Waterton, 2010).

- Le volume des données :

Le volume des données dans le Big data fait référence à l'ampleur de ces données. Les définitions du Big data selon le volume de données, sont relatives et varient selon les facteurs, comme le temps et le type des données. En effet, les données perçues comme étant de grands volumes aujourd'hui peuvent ne plus l'être demain parce que les capacités et les espaces de stockages s'améliorent et évoluent, ce qui permet de créer des ensembles de données encore plus considérables à conserver. Néanmoins, deux ensembles de données de même taille peuvent nécessiter différentes technologies de gestion de données en fonction de leur type, par exemple, des données tabulaires et des vidéos. Ainsi, les définitions du Big data dépendent également des technologies utilisées pour l'analyse. Ces considérations empêchent les chercheurs à établir un seuil spécifique pour définir des données comme volumineux. Toutefois, comme le soulignent Boyd et Crawford dans leurs travaux, "Big Data is less about data that is big than it is about the capacity to search, aggregate, and cross-reference large data sets" (Boyd et Crawford, 2012).

- La vélocité des données :

La vélocité des données se réfère à la vitesse à laquelle ces données sont générées, analysées puis traitées. La vulgarisation des technologies numériques tels que les Smartphones et les tablettes, a entraîné un taux sans précédent de création de données. Ces quantités énormes de données suscitent un besoin croissant de les analyser en temps réel et d'obtenir des avantages compétitifs prouvés. Les données émanant des appareils mobiles et des flux d'applications mobiles produisent des avalanches d'informations qui peuvent être utilisés pour générer des offres personnalisées en temps réel pour des clients ou électeurs potentiels. Ces volumes élevés de données fournissent plusieurs types d'informations, souvent « périssables » sur ses utilisateurs, telles que l'emplacement géographique, la démographie et les modèles d'achats ou de votes effectués. Ces ensembles de données peuvent être analysés instantanément pour créer une valeur tangible audit client ou électeur.

- La véracité des données :

La véracité des données a été évoquée comme quatrième V pour la première fois par IBM (Gandomi & Haider, 2015). Aborder la définition du Big data par cette dimension se veut nécessaire pour présenter le manque de fiabilité de certaines données vu leurs sources. Par exemple, les émotions des citoyens ressenties dans les médias sociaux à travers les mentions « j'aime », le partage ou les commentaires sont conditionnelles et imprécises. Pourtant, ils contiennent des informations précieuses. Ainsi, la nécessité de traiter des données approximatives et difficilement authentifiables, est une autre facette du Big data. Ladite facette est abordée à l'aide d'outils et d'analyses développés pour la gestion et l'exploitation de données incertaines.

- La valeur des données :

Le cinquième V, à savoir la valeur des données, a été introduit par l'entreprise américaine « Oracle » (Gandomi & Haider, 2015) comme un attribut contribuant à la définition du Big data qui se caractérise souvent par une surcharge informationnelle composée surtout de données à faible valeur. Autrement dit, les données collectées présentent habituellement une faible valeur par rapport aux volumes de données déjà stockés. Cependant, pour obtenir une grande valeur, il faut analyser de gros volumes de ces données tout étant capable de se focaliser sur celles qui sont réellement valorisables et actionnables.

- Le traitement des connaissances :

L'analyse dite prédictive se trouve au cœur des utilisations de Big Data (Bose, 2009 ; Earley, 2014). Cette analyse se base sur des modèles constitués par un ensemble de techniques statistiques

ainsi que d'autres outils tels que les algorithmes analytiques sophistiqués qui peuvent prédire les comportements, les scénarios et les paramètres qui intéressent l'utilisateur. En extrayant, traitant et croisant de grands ensembles de données, ces modèles sont révélés, puis contribuent à former la base de connaissances approfondies qui peuvent générer des avantages concurrentiels pour les organisations sollicitant lesdits Big data (Hair, 2007). Ces modèles d'analyse n'étaient pas faciles à déchiffrer ou à détecter avant l'ère du Big Data. Le matériel informatique ainsi que les logiciels nécessaires pour exploiter ces derniers, n'étaient tout simplement pas disponibles ou bien ils étaient trop coûteux et donc hors de portée de nombreux décideurs. Mais aujourd'hui, grâce au développement technologique (le cloud computing, les logiciels open source et Internet des objets), le Big Data peut fournir des informations, des données et des connaissances qui sont bien supérieures à ce qui étaient disponibles dans le passé (Jukić et al., 2015 ; Uma et Ashok, 2016).

En revanche, le Big Data peut-être définie selon Boyd et Crawford (2012), comme un phénomène culturel, technologique et scientifique qui repose sur l'interaction de trois éléments :

- 1) La technologie qui se développe pour maximiser le pouvoir de calcul et la précision algorithmique afin de rassembler, analyser, lier et comparer de gros ensembles de données ;
- 2) L'analyse qui s'appuie sur la collecte, le traitement et le croisement de grands ensembles de données pour identifier des modèles conceptuels dans le but de proposer des solutions économiques, sociales, techniques, juridiques (et aussi politiques).
- 3) La mythologie qui répand la conviction que les grands ensembles de données offrent une forme plus élevée d'intelligence et de connaissances pouvant ainsi générer des conceptions objectives, des visions véridiques et des représentations précises.

2. L'EXPLOITATION DU BIG DATA POUR DES FINS POLITIQUES : (LE CAS D'OBAMA)

L'avènement du digital et l'exploitation du Big data et de ses atouts dans les deux campagnes présidentielles d'Obama en 2008 et 2012 ont révolutionné les pratiques de marketing politique et ont contribué à un véritable changement dans les approches classiques usant des techniques habituelles. Avec les nouveaux outils de ciblage et de récupération des données personnelles que fournit le web, les stratégies politiques peuvent aller infiniment plus loin que le simple ciblage par couche sociale. C'est ce qu'a fait Obama avant même le lancement de la campagne de 2008, avec sa stratégie digitale qui ciblait en priorité, les électeurs encore indécis (LE GOFF, 2012). Les bases de données maîtrisées, la prédiction des comportements des électeurs, l'hyper-ciblage, la collecte de fonds et l'identification des influenceurs pour porter le message du candidat, sont tous des

techniques qui ont marqué lesdites campagnes et ont permis à Obama de gagner les élections à deux reprises.

2.1. LA NOUVELLE MAITRISE DES BASES DE DONNEES

Le point de départ du changement repose sur la maîtrise nouvelle des bases de données et de leur exploitation systématique. Jusque-là, les politiques et leurs stratégies travaillaient sur des données chiffrées, parfois complexes, qui dessinaient le paysage politique révélé par les études de marché sous toutes leurs formes (sondages, panels, etc.). Une connaissance fine de l'opinion, des enjeux et des comportements était ainsi permise et aussi bien les campagnes électorales que les études politiques étaient largement fondées sur ces éléments d'analyse. (Bille, 2015)

La collecte d'un nombre consistant de bases de données de différents types (politiques, commerciales...) puis l'analyse de ces données par approximation et croisement, constituent des nouveautés apportées par « le Digital Data Task force » d'Obama qui est constitué principalement de jeunes informaticiens venus de Google, Facebook et d'autre multinationales spécialisées du web 2.0. Le traitement de ces ensembles de données hétérogènes permet de dégager des informations précieuses et jusqu'ici indisponibles, sur la plupart des citoyens américains et sur leurs attitudes et comportements en permanence. Ainsi, un modèle de prise de décision qui prend en considération l'importance des données chiffrées et leur évaluation différenciée, se met en œuvre « *ledata-driven decision-making* ». Ce processus de prise de décision a procuré pour l'équipe d'Obama des avantages compétitifs comme la géolocalisation des électeurs indécis pour mener une campagne basée sur le porte à porte mais aussi connaître leurs préférences et besoins pour ainsi prédire leurs comportements.

Par ailleurs, aux Etats-Unis, plusieurs données politiques et électorales sont disponibles publiquement comme les agrégations de données sociodémographiques et nominatives sur une base géographique ou les corrélations prédictives des votes en fonction de l'historique de résultats électifs précédents par zone géographique. Même les rapports de sondages d'opinion et les études adéquates réalisées pour connaître les thèmes de « bascule » propres à influencer le vote des électeurs indéterminés en fonction des zones d'habitat, sont accessibles (Ruet, 2013). Autant plus que, les bases de données à spécificité économique, sociologique ou commerciale qui décrivent minutieusement la consommation des citoyens américains, sont mises en vente formellement. En conséquence, grâce au Big data l'équipe électorale d'Obama dispose à l'avance d'une base assez solide de connaissance de l'électorat qui lui permettra de pouvoir rapprocher un nombre considérable de données pour une même personne tout en donnant le moyen aux statisticiens des

campagnes électorales d'établir des « profils ». Ces profils vont, à leur tour, permettre d'élaborer des algorithmes de comportement individuel, notamment sur trois aspects essentiels : l'abstention ou la non-inscription sur les listes, la mobilisation précédente des électeurs, la conviction des indécis. (Bille, 2015)

2.2. LA PREDICTION DES ATTITUDES ET COMPORTEMENTS

Depuis 1998, Green et Gerber ont effectué des recherches dans plus d'une douzaine d'états américains, étudiant ainsi un large éventail d'élections fédérales, étatiques et municipales. Les recommandations prescrites dans leurs travaux (2008) découlent d'expériences approfondies menées dans des contextes électoraux réels, en examinant l'impact et l'efficacité de la prospection « porte-à-porte », les appels téléphoniques, le courrier direct et d'autres tactiques de campagne. Les stratégies élaborées pour réaliser les campagnes électorales d'Obama se sont basées sur les travaux de ces deux chercheurs. En sollicitant des bases de données de type commerciales ou sociologiques, les équipes d'Obama cherchaient à identifier et analyser les attitudes des électeurs non-inscrits ou indécis qui pouvaient voter démocrate. Dès qu'ils sont repérés, les militants se mobilisent pour les aborder directement avec des messages précis et adaptés à leurs attentes et besoins personnels conformément aux profils qu'ils leur ont dressés. Cela dit, cette politique de prospection est considérée par les observateurs et les chercheurs comme une innovation majeure dans le domaine du marketing politique.

En effet, c'est avec la campagne d'Obama en 2008 que les tendances de marketing digital ont largement investi le domaine politique. En deux ans de campagne, grâce à sa mobilisation sur Internet, l'ancien président des États-Unis a fait travailler près de deux millions de personnes pour sa cause et envoyé sur le terrain des millions de militants pour leur faire ouvrir quelques 70 millions de portes (LE GOFF, 2012). Issenberg, dans ses travaux, précise que l'équipe de campagne d'Obama a débuté l'année de l'élection en étant convaincue qu'elle connaissait le nom de chacun des 69 456 897 Américains qui, par leurs voix, l'avaient portée à la Maison-Blanche (Issenberg, 2012). Ainsi donc, en définissant les électeurs qui sont catégorisés comme indécis ou même républicains modérés, les équipes d'Obama pouvaient « cibler » ces individus et essayer de les convaincre avec des outils dialogiques personnalisés ou en d'autres termes, adapter une stratégie de *Nanotargeting* (micro-ciblage).

2.3. MICRO-CIBLAGE : VERS UN CIBLAGE PLUS FIN

Pour gagner les élections, les deux grands partis politiques les plus populaires aux États-Unis, à savoir le parti républicain et celui des démocrates, tentent de rallier le maximum de votants, chacun

de son côté, en optant pour des stratégies de marketing politique bien précises. Étant donné que le scrutin présidentiel américain passe par deux étapes. Les citoyens choisissent des représentants, qui, à leur tour, voteront pour un des deux candidats à la présidence. Partant du fait que lors des campagnes électorales américaines, il y a des États reconnus comme des états acquis à l'avance par les républicains « Red States » et d'autres comme traditionnellement acquis par les démocrates « Blue States ». Il existe une quinzaine d'États où le résultat des votes est méconnu par les analystes et les professionnels du jeu politique.

En effet, les électeurs de ces États «*Swing States* » constituent un réservoir de vote estimé autour de 6 à 10% des votants en début des campagnes électorales et souvent nommé comme le "marais" dans le vocabulaire des sondeurs car à eux-seuls, ils peuvent faire balancer les élections. C'est donc avec certitude que les équipes des candidats vont se focaliser sur ces États pour convaincre les indécis et non-inscrits d'aller voter pour leurs partis. Les efforts fournis, la mobilisation et l'argent dépensé dépendent des poids de ces États et ceci selon le nombre des représentants qui seront élus.

Vu le nouveau contexte dans lequel ce sont déroulées les élections, caractérisé par l'utilisation massive des technologies de l'information et de la communication et vu aussi les enjeux considérables, le micro-ciblage était le choix stratégique des équipes de campagnes d'Obama. Les zones cruciales (parfois réduites à un simple comté) vont faire l'objet de campagnes de télévision spécifiquement ciblées, diffusées sur les chaînes locales, mais aussi d'actions militantes personnalisées, individualisées (Bille, 2015). Les bénévoles faisant le porte-à-porte ou la prospection par téléphone, disposaient de données spécifiques (préférence politique, situation familiale et patrimoniale, dossier médical, comportements d'achat, loisirs, historique web...) sur les électeurs visités ou appelés. Chacune de ces interactions a produit des données complémentaires sur la cible et qui sont réinjectée dans les serveurs dédiés à la campagne d'Obama pour affiner les modèles de micro-ciblage, orientant les bénévoles vers la porte suivante à frapper ou le numéro de téléphone adéquat à appeler. L'efficacité et l'ampleur de ce processus ont permis aux démocrates d'être en avance en matière d'établissement de profil des électeurs. Au fond, ces modèles innovants de micro-ciblages ont révolutionné les techniques anciennes de prospection tels que le porte-à-porte et la prospection téléphonique, les rendant ainsi plus efficaces et efficaces.

En plus de ces actions traditionnelles revisitées, d'autres actions en ligne ont été entreprises, comme la création des pages Facebook pour soutenir le candidat et promouvoir la collecte de fonds, ou le compte officiel d'Obama sur Twitter pour interagir directement avec ses partisans et *followers*.

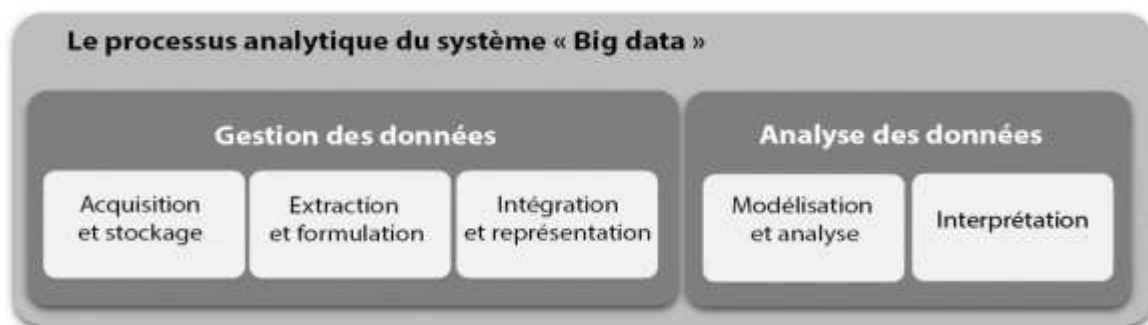
Ces formes de communication actives instaurent des relations tangibles entre le candidat, ses équipes de campagne et les électeurs. Ces formes produisent des informations complémentaires de plus en plus précises et ciblées. Des chercheurs ont constaté et même signalé le fait que la communication des politiciens à travers Facebook et Twitter a pratiquement pris le pas sur la communication traditionnelle de presse (Monnoyer-Smith, 2011 ; Chibois, 2014 ; Roginsky et Perrier, 2014). Ils précisent que grâce à ces plateformes en ligne, l'acteur politique publie ses messages, partage ses convictions et valeurs sans passer par des intermédiaires. Ses publications sont parfaitement assimilées, sans risque de contradiction ou de dérapage, établissant ainsi un lien direct avec son public.

En somme, toutes ces technologies et techniques déployées se consolident dans un « système de gestion de la relation électeur » ou comme nommée dans nos recherches « Elector Relationship Management » et se base sur des ensembles de masses de données quantifiées ingénieusement organisées par de nouveaux spécialistes (Politique intelligence manager, Data scientist, Data analyst, Data miner...) dans les organisations politiques, permettant ainsi aux experts en sciences sociales et politiques d'atteindre des résultats et des performances irréalisables auparavant (Boussaid, Azdimousa, 2017).

3. LE PROCESSUS ANALYTIQUE DU SYSTEME « BIG DATA » :

Pour transformer des volumes de données élevés, diversifiées et rapidement périssables en connaissances significatives procurant des avantages concurrentiels, les décideurs politiques ont besoin d'un processus efficace d'analyse se basant sur des informations émanant des sources vérifiées, afin de piloter avec succès des projets comme les campagnes des élections. L'exploitation du Big data pour ces fins politiques a apporté déjà ses fruits comme constaté précédemment.

Le processus global d'extraction des données significatives et valorisables du Big data peut-être divisé en cinq étapes (Labrinidis & Jagadish, 2012 ; Gandomi & Haider, 2015), illustrées ci- dessous.



Source : A. Gandomi, M. Haider / International Journal of Information Management 35 (2015) 137–144

Ces cinq étapes constituent les deux sous-processus analytiques principaux du Big data: la gestion des données et l'analyse des données. D'une part, la gestion des données implique des processus et des supports technologiques pour acquérir et stocker des données, les formuler et les intégrer afin de les analyser. D'autre part, l'analyse des données, se fonde sur des techniques utilisées pour analyser et interpréter les informations acquises à partir du Big data.

Ceci dit, les techniques analytiques utilisées pour traiter des données structurées et non structurées, stockées dans le Big data, sont diverses et variées. Des techniques d'analyse de données de type texte, son ou vidéo issues des réseaux et médias sociaux en lignes seront exposées ci-dessous. Ces techniques représentent un sous-ensemble pertinent des outils disponibles pour aborder et traiter les données du Big data.

- **L'analyse des données de type : Texte**

L'exploration de texte s'effectue à travers des techniques qui extraient des informations à partir de données textuelles stockées dans des bases de données. Des organisations géantes comme Facebook, Google, Microsoft ou autres détiennent d'énormes données textuelles telles que les flux de données des réseaux sociaux en ligne, les courriels, les documents formels et informels relatifs aux administrations, entreprises et organisations politiques. Lesdits multinationales disposent aussi des réponses et résultats concernant les enquêtes et études réalisées à l'échelle mondiale pour des fins politiques, économiques et même géostratégiques.

Les techniques d'analyse de texte impliquent une analyse statistique des données, une pensée computationnelle à la compréhension des ensembles informationnels et un apprentissage automatique permettant le développement d'une intelligence artificielle.

Les analyses de texte permettent aux partis politiques de convertir de grands volumes de textes générés par le citoyen en données significatifs. Lesdites données soutiendront la prise de décision fondée sur des données concluantes. L'exemple d'Obama en matière de micro-ciblage aux élections présidentielles de 2008 et 2012 ou celui de Trump par rapport à l'utilisation des réseaux sociaux lors des élections présidentielles de 2016, illustrent parfaitement le processus d'analyse basé sur les données de types textes. (Boussaid et Azdimousa, 2017)

Ceci dit, les méthodes d'analyse de texte utilisées sont souvent les techniques d'extraction d'informations, les techniques de synthèse du texte, les techniques des réponses aux questions, Les techniques d'exploration d'opinion. Lesdites techniques seront présentées ci-dessous :

Les techniques d'extraction d'informations permettent l'extraction des données structurées à partir de textes non structurés. Par exemple, les algorithmes utilisés par ces

techniques peuvent extraire des informations structurées telles que le nom du candidat, sa popularité et la fréquence de sa visibilité sur les médias. Deux sous-processus s'effectuent dans le processus d'extraction d'informations selon Jiang (2012), à savoir, la reconnaissance des entités et l'extraction de la relation entre les entités. Le premier trouve les identifiants dans le texte et les catégorise dans des entités prédéfinies telles que la personne, l'organisation et les médias, tandis que le deuxième identifie et extrait des relations sémantiques entre ces entités dans le texte.

Les techniques de synthèse de texte produisent automatiquement un résumé succinct d'un ou plusieurs documents. Le résumé obtenu offre des informations clés pour déchiffrer le texte original. Ces techniques sont utilisées pour assimiler rapidement des articles scientifiques, des articles publicitaires, des publicités, des courriels et des blogs. D'une manière générale, le résumé élaboré suit deux approches : l'approche extractive et l'approche abstractive (Gandomi & Haider, 2015). Dans le cadre d'une approche extractive, le résumé est créé à partir des unités de texte d'origine (habituellement des phrases). L'approche extractive implique l'identification des unités de texte les plus importantes puis les assembler pour formuler un résumé sous forme d'un sous-ensemble homogène à partir du document original. L'importance des unités de texte est évaluée en analysant leur emplacement et leur fréquence dans ledit texte. En revanche, l'approche abstractive nécessite l'extraction d'informations sémantiques du texte. Le résumé contient des unités de texte qui ne sont pas toujours présentes dans le document source. Afin d'analyser le texte original et de générer le résumé, la synthèse abstractive intègre les techniques avancées de traitement du langage naturel (PNL). En conséquence, les systèmes abstraits ont tendance à générer des résumés plus structurés et adéquats que les systèmes extractifs (Hahn & Mani, 2000). Cependant, vu que le volume des données dans le Big data est consistant, les techniques d'analyses extractifs sont plus faciles à utiliser.

Les techniques de réponses aux questions sont des techniques d'analyse des données fondées sur la capacité à répondre à des questions formulées en un langage parlé par l'homme, en s'appuyant sur l'outil complexe du PNL. L'exemple de l'application Siri d'Apple illustre parfaitement l'utilisation de ces techniques dans une approche commerciale. Les techniques de réponses aux questions peuvent être appréhendées de trois manières : l'approche basée sur la récupération de l'information, l'approche basée sur les connaissances et l'approche hybride. L'approche basée sur la récupération de l'information est souvent composée de trois étapes. Premièrement, les questions posées seront traitées pour déterminer les détails, tel que le type de question, l'objectif souhaité de la question et le type de réponse. Une fois ces étapes effectuées, une requête sera formulée.

Deuxièmement, en utilisant la requête formulée dans le traitement des questions, une opération de récupération des passages pertinents, pré-écrits, à partir d'un ensemble de documents existants et abondants dans le Big data, sera réalisée et traitée dans un document unique. La troisième étape est le traitement des réponses. L'extraction des réponses *ad hoc* après réalisation des deux étapes précédentes, permettra l'élaboration d'une liste de choix de réponses exhaustives. Lesdites réponses seront classées en ordre de pertinence. La réponse la mieux classée sera proposée à la personne qui a posé la question en premier.

L'approche basée sur les connaissances génère une description sémantique de la question, qui est ensuite utilisée pour interroger les données structurées de type texte. Cette approche est particulièrement utile pour des domaines restreints, tels que la médecine où de grands volumes de documents pré-écrits n'existent pas, vu le manque de redondance des données.

Dans l'approche hybride, la question est analysée sémantiquement mais les réponses choisies sont générées en utilisant les méthodes de récupération de l'information.

Les techniques d'exploration d'opinion analysent les textes exprimant des opinions par des clients ou citoyens sur des entités établies comme vu précédemment, telles que les produits, les organisations, les individus et les événements. Le marketing, les finances et les sciences politiques et sociales sont les principaux domaines d'application de l'analyse du sentiment. Les techniques d'analyse du sentiment sont divisées en trois sous-groupes, à savoir les techniques d'analyse des opinions au niveau du document en entier, au niveau d'une phrase dans le document et par rapport à l'aspect du document. Les techniques utilisées au niveau du document déterminent si le document en complet, exprime un sentiment négatif ou positif. Les techniques d'analyse effectuée au niveau de phrase tentent de déterminer la polarité d'un seul sentiment concernant une entité connue exprimée en une seule phrase. Les techniques adoptées au niveau d'une phrase doivent d'abord distinguer les sens d'ordre objectifs de ceux d'ordre subjectifs. Les techniques d'analyse basées sur les aspects reconnaissent tous les sentiments dans un document et identifient les aspects de l'entité à laquelle chaque sentiment se réfère. Par exemple, les critiques des citoyens par rapport aux politiques contiennent généralement des opinions sur différents aspects. En utilisant des techniques d'exploration d'opinion basées sur les aspects, l'acteur politique pourra obtenir des informations précieuses sur les différents aspects de son image et évaluer ainsi la portée des messages qu'il émet.

- **L'analyse des données de type : Son**

Les analyses des données de type audio extraient des informations valorisables à partir des données audios non structurées. L'analyse des données audio est aussi appelée analyse de la parole lorsqu'elle est appliquée sur le langage humain.

L'analyse de la parole suit deux approches technologiques communes : l'approche basée sur la transcription (largement connue par les anglo-saxons sous le nom de « large-vocabulary continuous speech recognition », LVCSR) et l'approche phonétique.

L'approche LVCSR suit un processus en deux étapes : l'indexation et la recherche. Dans la première phase, cette approche tente de transcrire le contenu vocal de l'audio. Ceci est effectué à l'aide d'algorithmes de reconnaissance automatique de la parole qui correspondent aux sons et aux mots. Les mots sont identifiés sur la base d'un dictionnaire prédéfini. Si cette technique d'analyse ne parvient pas à trouver le mot exact dans le dictionnaire, alors le mot qui semble le plus proche est choisi. Le résultat de l'analyse est un fichier d'index consultable qui contient des informations sur la séquence des mots prononcés dans la piste audio. Dans la deuxième phase, les méthodes d'analyse basées sur le texte, comme citées précédemment, sont utilisées pour trouver le terme de recherche dans le fichier d'index.

L'approche phonétique fonctionne avec des phonèmes. Les phonèmes sont des éléments sonores du langage humain, articulés de façon distinctives et considérés sur le plan physiologique et acoustique comme des constructions de pair minimales, c'est-à-dire des paires de mots de sens différents mais qui ne diffèrent que par un seul son dans leur forme sonore.

L'approche phonétique se compose également de deux phases : l'indexation et la recherche phonétique. Dans la première phase, contrairement à l'approche LVCSR, cette technique d'analyse traduit l'ensemble des paroles prononcées dans la piste audio étudiée en une séquence de phonèmes et non en une séquence de mots. Dans la deuxième phase, une représentation phonétique des termes recherchés est élaborée et présentée sous une forme de fichier d'index comme résultat pour ainsi la traiter.

- **L'analyse des données de type : Vidéo**

L'analyse de contenu vidéo nécessite une variété de techniques pour analyser, établir un système de veille et extraire des informations significatives à partir des flux de vidéos en temps réel et préenregistrées. La popularité florissante des sites Web de partage de vidéos comme Youtube et le développement continu des technologies de production visuelle sont les deux principaux facteurs qui ont mis en avant l'intérêt croissant d'analyser des vidéos par des approches automatiques. Le

troisième facteur qui à son tour, a contribué au développement de l'analyse des données vidéos est l'émergence du Big data. Ce système d'information gigantesque a permis de filtrer automatiquement les données de type vidéo et extraire de l'information précieuse pour les décideurs, à partir de milliers d'heures de vidéo. Pour mesurer l'ampleur de ces exploitations facilitées, il faut prendre en considération qu'une seconde de vidéo en haute résolution génère 2000 fois plus d'octets nécessaire pour son stockage que la rédaction d'une seule page de texte (Manyika et al., 2011)

Indexer et récupérer des vidéos automatiquement constitue un autre domaine d'applications des techniques d'analyse de contenu vidéo, surtout, vue l'émergence généralisée de vidéos en ligne et hors ligne. D'où la nécessité d'indexer le contenu multimédia pour faciliter la recherche et la récupération. L'indexation d'une vidéo peut être effectuée selon les différents niveaux d'informations disponibles dans une vidéo. Lesdits niveaux incluent les métadonnées, la bande son, les transcriptions et le contenu visuel de la vidéo. Dans l'approche fondée sur les métadonnées, les systèmes de gestion de bases de données relationnelles (SGBDR) sont utilisés pour la récupération et la recherche de vidéos. Les techniques d'analyse des données de type audio et texte peuvent être appliquées pour indexer une vidéo se basant sur ses bandes sonores et ses transcriptions associées.

Il existe deux approches de l'analyse de contenu vidéo, à savoir une approche d'analyse appliquée sur les vidéos stockées dans des serveurs et une autre appliquée sur les vidéos sources.

Approches d'analyse basée sur des vidéos stockées dans des serveurs : Les vidéos enregistrées sont renvoyées vers un serveur centralisé. Ledit serveur est dédié pour la réalisation des analyses desdits vidéos. En raison des débits de connexions parfois limités, la vidéo générée par la source est habituellement comprimée pour réduire sa taille. Ces manœuvres impactent souvent la résolution des vidéos. La perte d'information qui en résulte peut affecter la précision de l'analyse. Cependant, l'approche basée sur le stockage des vidéos dans des serveurs offre des économies d'échelle et facilite la maintenance du matériel informatique.

Approche d'analyse basée sur les vidéos sources : L'analyse des vidéos est effectuée sur les données brutes enregistrées par la caméra. Par conséquent, l'ensemble du contenu vidéo est disponible pour l'analyse, souvent en haute résolution. Cette approche permet une analyse de contenu plus efficace. Cependant, ladite approche peut s'avérer plus coûteuse par rapport à l'approche d'analyse basée sur les vidéos hébergées dans des serveurs, en matière de matériels informatiques déployés et sa maintenance.

4. BIG DATA ET CREATION DE VALEUR ELECTEUR-POLITIQUE

L'accumulation des données rendue possible par la baisse des coûts de stockage couplée aux nouvelles capacités de traitement de l'information donne une valeur économique potentiellement très forte aux données numérisées, à la condition d'être capable de donner du sens à des volumes de données très importants (Chamaret, 2014). Ainsi, le Big data induit une révolution des modes de pensée. Cette révolution n'est pas technologique, elle est plutôt sociétale : il ne s'agit plus de savoir comment les données seront traitées techniquement, mais dans quel but les informations exploitées seront valorisées (Mayer-Schonberger et Cukier, 2013). Par ailleurs, la valorisation de l'information peut aussi concerner la politique comme dans le cas étudié d'Obama. Ceci dit, pour extraire une valeur politique pour des électeurs-politiques, à partir des données traitées, il faut en premier lieu que les partis politiques collectent un nombre très important de données. Puis, en deuxième lieu, lesdits partis disposent des compétences nécessaires pour le traitement des données pour les exploiter et en tirer profit. En dernier lieu, il faut que les partis politiques conçoivent de nouvelles formes de valorisation des données pour se procurer des avantages compétitifs.

4.1. LE PROCESSUS DE CREATION DE VALEUR

Afin de déterminer les activités clés dans une entreprise qui peuvent lui procurer un avantage concurrentiel, Porter dans ses travaux (1980, 1985) a développé un outil d'aide à la décision, nommé **chaîne de valeur**. Ladite chaîne est un modèle qui décrit une série d'activités à valeur ajoutée reliant l'offre d'une entreprise (matières premières, logistique d'entrée et processus de production) à sa demande (logistique sortante, marketing et ventes). En analysant les étapes d'une chaîne de valeur, les gestionnaires peuvent repenser leurs processus internes et externes pour améliorer leur efficience et leur efficacité (Rayport et Sviokla, 1995).

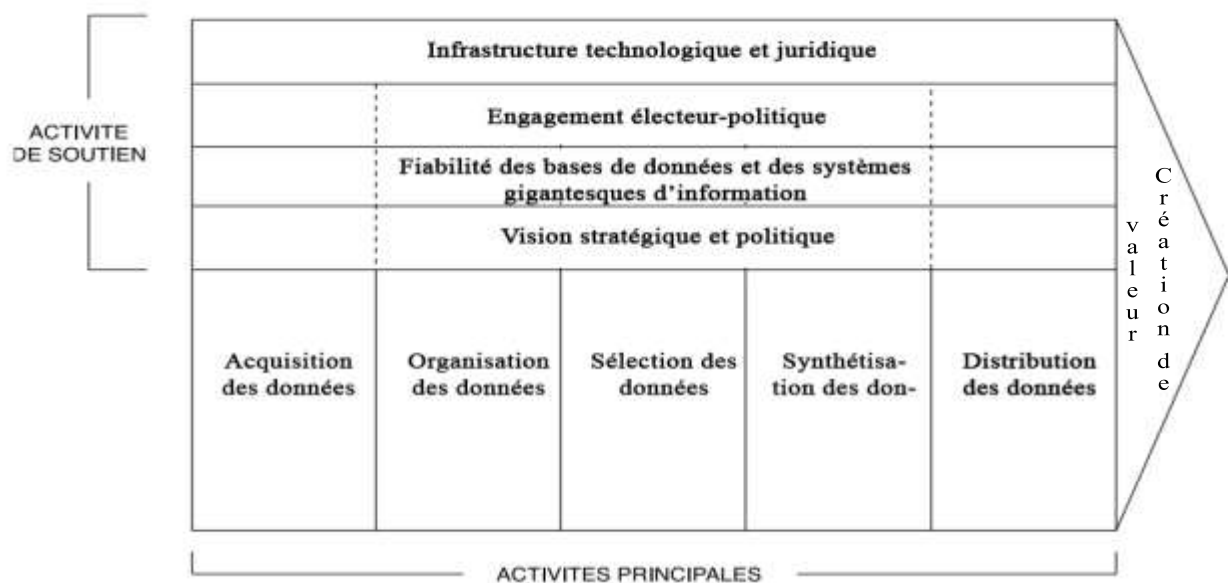
Rayport et Sviokla (1995) étant parmi les premiers chercheurs à transposer le modèle de la chaîne de valeur des activités physiques aux systèmes d'information virtuels. Ils ont constaté que pour créer de la valeur à n'importe quelle étape d'une chaîne de valeur virtuelle, il faut impliquer une séquence de cinq activités : acquérir, organiser, sélectionner, synthétiser et distribuer des informations. Tout comme quelqu'un prend la matière première et l'affine en quelque chose d'utile - comme dans l'enchaînement des tâches impliquées dans l'assemblage d'une automobile sur une ligne de production - un gestionnaire recueille aujourd'hui des informations brutes et ajoute de la valeur à travers ces étapes.

S'inspirant de leurs travaux, mais aussi des travaux de Curry et al. (2016), nous proposons une chaîne de valeur du Big data (Figure 1) qui contribue à modéliser les activités principales et de

soutien indispensables à la création de valeur électeur-politique. Cette valeur ne peut être créée que par l'exploitation de la chaîne de valeur virtuelle.

Les acteurs politiques doivent accéder aux données collectées par ces systèmes gigantesques d'informations en se basant sur les étapes de la chaîne de valeur. Ceci pourrait les aider à mieux planifier, exécuter et évaluer les résultats avec plus de précision et de rapidité. Cela signifie aussi que ces acteurs doivent prendre des décisions à chaque étape de la chaîne de valeur virtuelle, tout en ciblant les données à collecter mais aussi choisir la meilleure manière de les distribuer et d'analyser leurs impacts.

Figure 1 : Chaîne de valeur du Big Data à usage politique



L'acquisition de données : est le processus de collecte des données en relation avec la politique, de types structurées, non structurées. C'est aussi un processus de filtrage des flux de données en temps réel avant d'être stockés dans des serveurs dédiés, afin de procéder à leur organisation. L'étape de l'acquisition de données nécessite un investissement important en infrastructure. Comme cité auparavant, l'infrastructure nécessaire pour supporter l'acquisition de données importantes doit fournir la possibilité de traiter les données collectées par rapport à des séries d'événements. Lesdites infrastructures technologiques et mesure de gérer des volumes de données issus d'une multimodalité de sources.

L'organisation des données : consiste à exploiter les données brutes acquises, les analyser sémantiquement et développer un apprentissage automatique. Après extraction des informations, cette étape contribue à découvrir les données à potentiel élevé et tisser des liens entre elles, afin de pouvoir analyser les données communautaires et les données intersectorielles.

La sélection des données : est l'étape qui suit celle de l'organisation, classification et préservation des données. Ces données, après transformation, validation et annotation, consistent à être automatisées et conservées pour que l'utilisateur, à travers une computation des ensembles informationnels, puissent choisir des données de qualité.

La synthétisation des données : concerne l'étape de standardisation des bases de données avec un développement de capacité d'évolutivité et de performance par rapport à l'accessibilité rapide aux données ainsi que leurs disponibilités et cohérences.

La distribution des données : porte essentiellement sur les activités de visualisation, modélisation des données. Les données partagées peuvent aider à prendre des décisions, prétendre des comportements ou des résultats mais aussi établir des visions stratégiques pour améliorer les performances des politiques.

CONCLUSION

Dans cet article, après exposition d'une revue de la littérature sur le Big data et ses différents concepts, nous avons étudié en profondeur le cas des élections présidentielles d'Obama aux États Unis (2008-2012) qui a permis de dégager des éléments clés pour la mise en œuvre d'une campagne politique efficace, à savoir, prédiction des comportements des électeurs, le micro-ciblage des citoyens-électeurs... Chose qui nous a conduit en premier lieu à décrire le processus analytique du système « Big data », puis en second lieu, d'examiner le processus de création de valeur électeur-politique, à travers lequel nous avons proposé un modèle de chaîne de valeur virtuelle du Big data permettant son exploitation au service de la relation électeur-politique.

Le modèle de la chaîne de valeur, comme conçu par Porter (1980) traite l'information comme un élément de soutien du processus d'ajout de valeur, et non comme une source de valeur elle-même. Par exemple, les décideurs utilisent souvent les informations qu'ils saisissent sur l'inventaire, la production ou la logistique pour surveiller ou contrôler ces processus, mais ils utilisent rarement les informations elles-mêmes pour créer une nouvelle valeur pour le client.

Cependant, la chaîne de valeur virtuelle proposée dans cet article, s'appuyant sur une infrastructure technologique, permet aux décideurs politiques de mieux cerner leurs actions politiques dans un monde submergé par l'information. En conséquent, ils peuvent acquérir de avantages compétitifs par une bonne gestion des différentes étapes de la chaîne de valeur dans son ensemble, ainsi créer de la valeur par l'information et pour l'électeur-politique.

BIBLIOGRAPHIE

- Agrawal, D., Abbadi, A. El, Arora, V., Budak, C., Georgiou, T., Mahmoud, H. A. ... Wang, S. (2015). Mind your Ps and Vs: A perspective on the challenges of big data management and privacy concerns. In *Big Data and Smart Computing (BIGCOMP)* (pp. 1–6)
- Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*, 109(2), 155–172
- Boussaid, J., Azdimousa, H. (2017), E-participation, cyber-citoyen-électeur et cyber-partis : vers une digitalisation des pratiques politiques, 5^{ème} édition du Forum International sur la Recherche en Marketing (FIRM 2017), ENCG Oujda, Octobre 2017
- Camacho, J., Macia-Fernandez, G., Diaz-Verdejo, J., & Garcia-Teodoro, P. (2014). Tackling the big data 4 vs for anomaly detection. In *Proceedings - IEEE INFOCOM* (pp. 500–505).
- Cécile Chamaret, « La révolution big data », *Annales des Mines - Gérer et comprendre* 2014/2 (N° 116), p. 94-96
- Chadwick, A. and Stromer-Galley, J. (2016) 'Digital Media, Power, and Democracy in Parties and Election Campaigns: Party Decline or Party Renewal?' *The International Journal of Press/Politics* 1 –11.
- Danah boyd & Kate Crawford (2012): CRITICAL QUESTIONS FOR BIG DATA, *Information, Communication & Society*, 15:5, 662-679
- Donald P. Green et Alan S. Gerber, *Get Out the Vote: How to Increase Voter Turnout*, Brookings Institution Press, 2008, 225pp
- Earley, S. (2014). Agile Analytics in the Age of Big Data. *IT Professional*, 16(4), 18–20
- Edmunds, A., & Morris, A. (2000). The problem of information overload in business organisations: a review of the literature. *International Journal of Information Management*, 20(1), 17–28.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144
- Hahn, U., & Mani, I. (2000). The challenges of automatic summarization. *Computer*, 33(11), 29–36.
- Jacques Bille, « Marketing politique et big data », *Commentaire* 2015/2 (Numéro 150), p. 307-314.
- Jiang, J. (2012). Information extraction from text. In C. C. Aggarwal, & C. Zhai (Eds.), *Mining text data* (pp. 11–41). United States: Springer.

Joe F. Hair Jr, (2007), "Knowledge creation in marketing: the role of predictive analytics", European Business Review, Vol. 19 Iss 4 pp. 303 – 315

Jonathan Chibois, « Twitter et les relations de séduction entre députés et journalistes. La salle des Quatre Colonnes à l'ère des sociabilités numériques », Réseaux 2014/6 (n° 188), p. 201-228.

José María Cavanillas, Edward Curry, Wolfgang Wahlster, "New Horizons for a Data-Driven Economy - A Roadmap for Usage and Exploitation of Big Data in Europe". Springer 2016, ISBN 978-3-319-21568-6

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 5(12), 2032–2033.

Laurence Monnoyer-Smith, « La participation en ligne, révélateur d'une évolution des pratiques politiques ? », Participations 2011/1 (N° 1), p. 156-185.

LE GOFF R., 2012, "Les nouvelles pratiques du marketing politique en France et aux Etats-Unis", journal lanetscouade, <http://journal.lanetscouade.com/article/VNX-IScAAGTEen9J/les-nouvelles-pratiques-du-marketing-politique-en-france-et-aux-etats-unis>

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, 156pp.

MAYER-SCHÖNBERGER, V., & CUKIER, K. (2013). Big data: a revolution that will transform how we live, work and think. London, John Murray

McGuire, T., Manyika, J., & Chui, M. (2012). Why Big Data is the New Competitive Advantage. Ivey Business Journal Online.

Michael Porter, Competitive Strategy. Techniques for Analysing Industries and Competitors, Free Press, 1980.

Michael Porter, M.E. Competitive Advantage. Creating and Sustaining Superior Performance, Free Press, 1985.

Nenad Jukić, Abhishek Sharma, Svetlozar Nestorov & Boris Jukić (2015) Augmenting Data Warehouses with Big Data, Information Systems Management, 32:3, 200-209

Patrice Flichy, « Rendre visible l'information. Une analyse sociotechnique du traitement des données », Réseaux 2013/2 (n° 178-179), p. 55-89.

Rayport, J. F., & Sviokla, J. J. (1995). Exploiting the virtual value chain. Harvard Business Review, 73, 75–85.

Ruet H.,2013,"La victoire d'Obama : cas d'étude concret d'utilisation des Big Data", journal du net, <http://www.journaldunet.com/ebusiness/expert/53760/la-victoire-d-obama---cas-d-etude-concret-d-utilisation-des-big-data.shtml>

Sandrine Roginsky, Valérie Jeanne Perrier, « La fabrique de la communication des parlementaires européens. « Tweet ton député » et les « ateliers du député 2.0 » », Politiques de communication 2014/2 (N° 3), p. 85-124.

Uma, G. G. & Ashok, G. (2016). Vision: A Missing Key Dimension in the 5V Big Data Framework. Journal of International Business Research and Marketing, 1(3), 46-52.

WATERTON C., 2010, "Experimenting with the Archive: STS-ers as Analysts and Co-constructors of databases and Other Archival Forms", Science Technology Human Values, n° 1.