

Le « Big Data RH » : vers une nouvelle convention de quantification ?⁽¹⁾

Par Clotilde CORON
IAE Paris, GREGOR (EA 2474)

La mobilisation du Big Data dans différentes sphères de la société modifie aujourd'hui nos façons de consommer, de nous informer, de prendre des décisions. Cependant, peu de travaux s'intéressent à l'influence du Big Data sur la manière d'utiliser les chiffres et plus généralement sur la quantification dans les Ressources humaines. En mobilisant le cadre théorique de la sociologie de la quantification, notre travail cherche à identifier si et dans quelle mesure l'utilisation du Big Data contribue à créer de nouvelles conventions de quantification dans les RH. Pour cela, nous nous fondons sur de l'ethnographie organisationnelle grâce à une immersion longue en tant que chef de projet Big Data RH d'une grande entreprise, complétée par huit entretiens semi-directifs. Nous montrons qu'en effet le Big Data contribue à créer une nouvelle convention de quantification en RH, reposant sur une approche prédictive et non plus explicative. Cette convention se caractérise par quatre paradigmes : exhaustivité du modèle, corrélation, personnalisation et amélioration continue.

Introduction

Le Big Data est devenu dans les dernières années un phénomène important. Né tout d'abord de l'inflation du volume de données disponibles du fait de l'émergence des réseaux sociaux et des progrès de la téléphonie mobile, il s'est peu à peu étendu à toutes les sphères de la société (Mayer-Schönberger et Cukier, 2014 ; Menger, 2017). Ses applications dans le domaine du management sont ainsi de plus en plus évoquées par les entreprises et certains travaux académiques. Le Big Data est présenté comme venant modifier en profondeur les organisations (Sivarajah *et al.*, 2017 ; Raguseo, 2018), mais aussi les stratégies de marketing (Erevelles, Fukawa, et Swayne 2016). Le domaine des Ressources humaines (RH) est également touché par le Big Data (Angrave *et al.*, 2016). Ainsi, la presse managériale se fait régulièrement l'écho de dispositifs de RH mobilisant des volumes importants de données, issues des réseaux sociaux par exemple, ou non structurées⁽²⁾ (CV, commentaires libres sur un questionnaire...) : algorithmes de pré-sélection de CV, algorithmes de suggestions de postes et de candidats à partir de données issues de LinkedIn, algorithmes de prédiction des démissions à partir des comportements sur les réseaux sociaux... Cependant, si certains

travaux académiques se penchent sur l'utilisation de la quantification en RH (Madsen et Slåtten, 2017 ; Huselid, 2018), peu s'intéressent spécifiquement au Big Data en RH, et à la manière dont cette émergence du Big Data influe sur l'utilisation des chiffres et des méthodes statistiques par les acteurs RH.

La sociologie de la quantification s'intéresse justement à l'utilisation des chiffres et méthodes statistiques par différents acteurs. Plus précisément, elle s'intéresse aux conventions sociales sous-jacentes à l'usage de la quantification, qu'il s'agisse d'indicateurs, de chiffres, ou de méthodes statistiques. Ainsi, les travaux fondateurs de Desrosières (1993 ; 2008a ; 2008b) se sont penchés sur les opérations de taxinomie sous-jacentes à la production d'indicateurs (Gilles, 2013). Ils montrent les enjeux politiques et sociaux qui sous-tendent ces choix en apparence guidés par des raisons scientifiques ou méthodologiques. Espeland et Stevens (1998) se sont inscrits dans la même démarche en s'intéressant aux opérations de « commensuration », processus grâce auquel on parvient à comparer des objets à partir d'une même métrique. Chiapello et Walter (2016) s'intéressent quant à eux à la diffusion de conventions de quantification dans le monde de la finance. Par « conventions de quantification », ils entendent une « méta-convention, [qui] recouvre une configuration ou un ensemble cohérent d'opérations à la fois cognitives et normatives, incluant la sélection des éléments à prendre en compte, les critères de jugement pertinents,

⁽¹⁾ Je remercie chaleureusement les deux rapporteurs anonymes ainsi que l'équipe de *Gérer & Comprendre* pour l'ensemble du processus de relecture, révisions et publication.

⁽²⁾ Des données non structurées, comme du texte, des images, ne peuvent pas être rangées dans un tableur classique, contrairement aux données structurées.

les choix de schémas mathématiques, etc. » (p. 159)⁽³⁾. Nous souhaitons nous engager dans la même voie que Chiapello et Walter, mais en appliquant leur démarche au Big Data RH.

Nous nous demandons donc dans quelle mesure l'essor du Big Data produit de nouvelles conventions de quantification dans le domaine des Ressources humaines. Cette question prend une acuité particulière au regard du faible nombre de travaux sur le sujet du Big Data RH, qui contraste avec l'essor important du sujet dans la presse managériale. Elle tente en effet d'ouvrir vers une compréhension plus approfondie de ce phénomène et de ses implications pour les RH.

Pour traiter cette question, nous avons mobilisé un mélange de matériaux. Nous avons mené une approche ethnographique dans le cadre de notre expérience professionnelle de 20 mois en tant que chef de projet Big Data RH dans une grande entreprise. Nous avons complété cette approche par des entretiens facilitant la prise de distance avec le matériau ethnographique. Du fait de notre inscription dans la démarche de Chiapello et Walter et de notre objectif de caractérisation des conventions de quantification, nous avons fait preuve d'un intérêt particulier envers les données, méthodes et indicateurs de résultats utilisés dans les projets étudiés.

Nous présentons tout d'abord des éléments de contexte et de cadrage théorique, avant de décrire le cas mobilisé et notamment les deux projets principaux sur lesquels notre analyse repose. Nous présentons ensuite les résultats avant de les discuter. Nous montrons ainsi que les projets étudiés correspondent à des évolutions de paradigmes, qui en fait peuvent s'inscrire dans l'émergence d'une nouvelle convention de quantification en RH, correspondant au passage des modèles explicatifs aux modèles prédictifs.

Des usages variés de la quantification en RH à la notion de convention de quantification

L'utilisation de la quantification dans les RH

De nombreux travaux en sciences de gestion s'intéressent à l'utilisation des chiffres en RH. Plusieurs types d'utilisation des données sont valorisés par ces travaux, notamment la définition d'indicateurs de suivi qui permettent d'identifier des grandes tendances et l'utilisation de méthodes statistiques plus sophistiquées (Davenport, Harris et Shapiro, 2010 ; Rasmussen et Ulrich, 2015 ; Angrave *et al.*, 2016 ; Madsen et Slåtten, 2017). Ce second usage est généralement qualifié de « Analytique RH » (Rasmussen et Ulrich, 2015 ;

Huselid, 2018 ; Levenson, 2018). Il vise une meilleure compréhension de phénomènes RH complexes (absentéisme, engagement, par exemple) au moyen de modèles statistiques permettant d'établir des liens de causalité entre différentes variables. De nombreux travaux académiques relient l'utilisation de la quantification en RH à une approche EBM (*Evidence-based management*) (Boudreau et Ramstad, 2004 ; Lawler, Levenson et Boudreau, 2010 ; Cossette, Lépine et Raedecker, 2014), c'est-à-dire une approche dans laquelle les données sont utilisées pour améliorer la prise de décision. Finalement, les travaux sur le sujet avancent que la quantification est mise par les RH au service d'une meilleure compréhension du réel grâce à une approche causale et une identification des grandes tendances.

Cependant, ces travaux n'abordent pas le sujet du Big Data RH, qui donne pourtant lieu à de nombreux contenus dans la presse managériale⁽⁴⁾ : algorithmes de pré-sélection de CV, de suggestions de formations ou de postes, de prédiction des démissions, de prédiction de l'absentéisme, par exemple.

Le Big Data, un terme souvent caractérisé par des aspects techniques

Le Big Data reste encore sujet à une variété de définitions (Ollion et Boelaert, 2015). Un rapport de Gartner (2001) a défini le Big Data au moyen de 3 « V » (McAfee et Brynjolfsson, 2012 ; Kitchin, 2014 ; Boullier, 2017 ; Raguseo, 2018). Tout d'abord, le *Volume* de données doit être important, nécessitant par exemple de travailler sur des plateformes de stockage spécifiques. Ces données sont de plus caractérisées par leur *Variété*, au sens où l'on peut croiser des sources de données hétérogènes (internes et externes), et où elles peuvent être structurées et non structurées (Gandomi et Haider, 2015). Ces données sont dynamiques, c'est-à-dire mises à jour très fréquemment, ce que l'on appelle la *Vélocité*. Deux autres « V » ont été ajoutés plus récemment (Bello-Organ, Jung et Camacho, 2016 ; Erevelles, Fukawa et Swayne, 2016) : *Véracité*, renvoyant à la question de la qualité des données, et *Valeur*, renvoyant à l'idée de tirer un bénéfice à partir de ces données (Raguseo, 2018). Cette définition met donc l'accent avant tout sur les caractéristiques techniques des données mobilisées (volume, variété, mise à jour, qualité).

D'autres travaux s'intéressent plutôt à la manière d'utiliser les données. Ainsi, Cardon (2015) et O'Neil (2016) évoquent la notion d'algorithme (Bello-Organ, Jung et Camacho, 2016), et l'importance croissante accordée aux algorithmes dans la vie et les services quotidiens. Cochoy et Vayre (2017) rappellent d'ailleurs qu'une des promesses du Big Data se situe dans l'automatisation de certaines tâches, permise par la construction d'algorithmes adaptés (Paye, 2017).

⁽³⁾ Traduction par nos soins. Texte original : "Quantification convention is more like a meta-convention: its name covers a configuration or a coherent set of operations both cognitive and normative, including selection of the items to take into account, relevant judgement criteria, choices of mathematical schemas, etc."

⁽⁴⁾ Voir par exemple : <http://talentplug.com/le-big-data-au-service-des-rh/>, ou <https://www.exclusiverh.com/dossiers/big-data-quelles-applications-pour-les-rh.htm>.

Les conventions, un concept permettant d'appréhender la quantification

Prolonger ces travaux en les appliquant au domaine des RH nécessite de se pencher sur la façon dont le Big Data est mobilisé dans les RH et sur ce qu'il change aux caractéristiques des usages de la quantification dans les RH exposées ci-dessus.

La sociologie de la quantification s'intéresse aux pratiques de quantification et s'attache à montrer en quoi elles sont socialement construites (Diaz-Bone, 2016). Les travaux de Desrosières (Desrosières 1993 ; 2008a ; 2008b) déconstruisent le « mythe de la quantification » en montrant que les opérations de quantification ne produisent pas des reflets neutres de la réalité. D'autres travaux du même champ s'intéressent à la quantification comme enjeu de pouvoir et d'influence (Espeland et Stevens, 1998) ou aux effets que la quantification peut produire sur les individus (Lave, 1984 ; Sunstein, 2000 ; Stevens, 2008 ; Juven, 2016).

Ce cadre analytique repose notamment sur le concept de conventions (Diaz-Bone et Thévenot, 2010) : les conventions permettent de partager un cadre d'interprétation et d'évaluation des objets, des actes et des personnes, et donc d'agir en situation d'incertitude (Eymard-Duvernay, 1989 ; Cartier et Liarte, 2010 ; Diaz-Bone, 2016).

L'originalité des travaux de Desrosières consiste à mobiliser le concept de conventions pour analyser les opérations de quantification, ce qui revient à étudier les « conventions de quantification » (Desrosières, 2008a), à savoir un ensemble de représentations sur la quantification qui vont permettre de coordonner « les comportements, mais aussi les représentations » (Chiapello et Gilbert, 2013, p. 114). La particularité des conventions de quantification réside dans le fait qu'elles reposent entre autres sur des arguments et sur des techniques scientifiques, ce qui renforce l'illusion de leur objectivité (Salais, 2016). Chiapello et Gilbert (2013) puis Chiapello et Walter (2016) aident à préciser ce que peut être une convention de quantification : un ensemble d'opérations incluant par exemple la sélection des données, le choix des méthodes, les critères de pertinence et d'évaluation de l'outil.

Plusieurs travaux mobilisant ce cadre s'intéressent à la diffusion de techniques statistiques, comme autant de conventions de quantification. Ainsi, Desrosières (1993) se penche sur la diffusion des techniques et méthodes statistiques les plus courantes (moindres carrés ordinaires, corrélation). Salais (2016) s'intéresse à l'influence réciproque entre les conventions de quantification et les conventions sociales. Chiapello et Walter (2016) étudient la diffusion de systèmes de calcul en finance.

C'est cette dernière voie que nous souhaitons emprunter pour étudier le Big Data RH. Nous nous demandons donc si l'introduction du Big Data fait émerger de nouvelles conventions de quantification dans les RH.

Design de recherche

Pour répondre à cette question, nous étudions tout particulièrement deux projets dans le domaine du Big Data RH : Shortlist, un projet visant à présélectionner les candidatures dans le cadre du recrutement, et Formazon, un projet visant à fournir des suggestions personnalisées de formations aux salariés (sur le modèle d'Amazon par exemple). Ces deux projets ont été menés au sein de Multinum, une grande multinationale du secteur du numérique ayant son siège et la majorité de ses salariés (environ 90 000) en France. Du fait de son positionnement dans le secteur du numérique, Multinum cherche à développer l'usage du Big Data, à la fois dans le marketing et dans d'autres secteurs comme les Ressources humaines, ce qu'illustrent Shortlist et Formazon.

Un matériau mixte pour appréhender un sujet nouveau

Nous avons pris le parti de nous intéresser aux données utilisées et aux méthodes sous-jacentes aux algorithmes mobilisés. Pour cela, nous avons mobilisé principalement de l'ethnographie organisationnelle (Grosjean et Groleau, 2013), que nous avons complétée par des entretiens.

Nous avons été chef de projet Big Data RH chez Multinum durant 20 mois et avons notamment piloté le projet Formazon. Durant ce pilotage, nous avons travaillé entre autres avec des *data scientists*, des acteurs RH (recrutement et formation), des équipes de gestion des données RH et des juristes. Nous avons ainsi amassé un matériau important d'échanges informels, de mails, de comptes rendus de réunions, de présentations dans des comités de suivi, de documents internes. Nous avons lors de cette période aussi eu la charge de récolter des informations sur les différents projets de Big Data en RH de l'entreprise. C'est lors de cette mission que nous avons pu recueillir du matériau, à la fois sous forme d'échanges informels, de mails, et de documents de présentation, sur Shortlist.

Il s'agit donc ici d'une immersion longue et d'une approche ethnographique. La difficulté de cette approche réside dans le fait qu'elle nécessite à la fois une intégration au sein d'un groupe ou d'un sujet, mais aussi une distance intellectuelle et une prise de recul (Schouten et McAlexander, 1995). Dans notre cas, pour garantir cette prise de distance, nous n'avons commencé à analyser les matériaux qu'après avoir quitté Multinum et notre rôle de chef de projet.

Comme le soulignent Schouten et McAlexander (1995), cette méthode donne accès à des matériaux inaccessibles par des entretiens ou l'administration d'un questionnaire. En l'occurrence, les réflexions qui ont guidé la sélection des données et la construction des algorithmes, tout comme les difficultés et hésitations qui ont émaillé la conduite des deux projets évoqués, nous auraient été inaccessibles autrement que par l'approche ethnographique. Or, ce sont justement les éléments qui nous intéressaient et que nous avons retenus dans cette masse d'informa-

tions et de matériaux : choix guidant la construction des algorithmes, échanges sur le sujet, hésitations éventuelles, données mobilisées, définition des indicateurs de réussite.

Pour prolonger la prise de distance, après avoir quitté Multinum, nous avons effectué huit entretiens autour de ces projets, ainsi que d'autres projets mobilisant des données et se réclamant soit du Big Data RH, soit d'une utilisation plus classique de la quantification chez Multinum et d'autres entreprises (voir Tableau 1 ci-dessous), de façon à pouvoir comparer les différents types d'utilisation. Le but de ces entretiens consistait à obtenir des informations sur des projets de quantification en RH. Les entretiens portaient donc sur un projet en particulier et demandaient aux interviewés d'explicitier le but du projet, les données mobilisées, les méthodes utilisées, les différentes étapes de mise en œuvre, ainsi que les difficultés et les limites.

Nous nous intéressions particulièrement aux données et méthodes mobilisées. Le codage des entretiens a donc consisté avant tout à lister les sources de données (souhaitées, et réelles, quand les acteurs avaient dû revoir leurs ambitions à la baisse comme cela a été le cas pour Formazon), et les méthodologies, ainsi que les différents arguments donnés pour justifier l'emploi de telle ou telle méthode. Par ailleurs, les entretiens venant après la séquence longue d'ethnographie organisationnelle, nous avons déjà identifié des pistes de recherche, notamment autour des nouveaux paradigmes introduits par le Big Data RH. Cela a influencé notre analyse des entretiens, venant comme un matériau confirmatoire et non exploratoire.

Les deux sous-sections suivantes, fondées sur le matériau ethnographique et les entretiens, sont consacrées à décrire les deux principaux projets de Big Data RH qui nous intéressent, Shortlist et Formazon.

Shortlist : un algorithme de présélection des candidatures

Shortlist, dont l'objectif consistait à mettre en place un algorithme de présélection des CV, a débuté en février 2016. Multinum reçoit plus de 100 000 CV par an à analyser, pour une fourchette allant de 6 000 à 10 000 offres d'emploi selon les années, sachant que le volume de candidatures est beaucoup plus important sur certains métiers que sur d'autres. Ce projet a nécessité un partenariat entre la Direction du recrutement (notamment la directrice du recrutement et les dix chargés de pré-sélection de CV) et une équipe interne de *data scientists*.

Le projet proprement dit s'est déroulé en trois phases. De février à juillet 2016, l'équipe de *data scientists* a bâti un premier algorithme sur un petit nombre d'offres et de CV. Plus précisément, l'algorithme ainsi conçu fonctionne sur de l'analyse sémantique, qui, à partir de tous les CV lus, crée des nuages de mots en fonction du positionnement des mots dans le CV (deux mots positionnés sur la même ligne peuvent être considérés comme proches). Ensuite, il compare les nuages de mots des CV avec les nuages des mots des offres, à la fois en fonction de leur présence et de leur positionnement dans le CV et l'offre. Ainsi, le même mot ne se trouvera pas dans le même nuage s'il se trouve dans un paragraphe lié au savoir-faire ou un paragraphe lié au savoir-être. Enfin, l'algorithme attribue un score de pertinence à chaque CV, en fonction de la proximité des nuages de mots du CV avec ceux de l'offre concernée. Ce score de pertinence permet alors de classer les CV et donc de les présélectionner.

Ensuite, au second semestre de 2016, les *data scientists* ont mobilisé de plus grands volumes de données (toutes les offres et tous les CV sur 2 mois, ce qui correspond à environ 1 000 offres et 10 000 CV), de manière à améliorer l'algorithme, qui apprend à partir

Tableau 1 : Matériaux recueillis

Projets	Matériaux
Formazon (Big Data RH)	<ul style="list-style-type: none"> – Ethnographie organisationnelle – Documents internes, réunions, échanges de mails – 1 entretien avec un <i>data scientist</i>
Shortlist (Big Data RH)	<ul style="list-style-type: none"> – Ethnographie organisationnelle – Documents internes – 2 entretiens : 1 avec un <i>data scientist</i>, 1 avec une chef de projet
Autres projets de Big Data RH chez Multinum	<ul style="list-style-type: none"> – 1 entretien avec une chef de projet de suggestions de postes en interne (partenariat avec une start-up externe)
Autres projets d'usage plus classique de la quantification en RH chez Multinum	<ul style="list-style-type: none"> – 2 entretiens : 1 avec une <i>data scientist</i> sur une étude de l'absentéisme, 1 avec un chef de projet d'analyse des verbatims
Acteur externe dans le domaine du Big Data RH	<ul style="list-style-type: none"> – 1 entretien avec un entrepreneur sur une solution de suggestions de postes en fonction du profil de compétences des salariés
Acteur externe dans le domaine de la quantification plus classique en RH	<ul style="list-style-type: none"> – 1 entretien avec un entrepreneur sur une solution d'analyse clé-en-main de l'absentéisme

des CV et des offres. Le score de pertinence, calculé comme le taux de similarité entre les 20 premiers CV retenus par l'algorithme et les 20 premiers CV retenus par un chargé de recrutement sans l'aide de l'algorithme, a été estimé à 65 % à l'issue de cette étape.

Enfin, au premier semestre 2017, l'algorithme a été amélioré « à la main ». Alors que les deux premières étapes n'avaient pas donné lieu à une collaboration spécifique avec l'équipe de recrutement, des *data scientists* ont collaboré avec deux chargés de recrutement volontaires (sur les dix de l'équipe) lors de cette dernière étape. Cette collaboration visait à analyser les cas de désaccord entre algorithme et chargé de recrutement : comprendre pourquoi le chargé de recrutement retenait ou ne retenait pas un CV, et pourquoi l'algorithme ne donnait pas le même résultat, à partir d'une offre et d'un ensemble de CV. Cette étape a permis d'améliorer la pertinence des nuages de mots et donc de l'algorithme globalement. Cela a fait passer le score de pertinence de 65 à 80 %.

Formazon : un algorithme de suggestions de formations

Formazon, dont l'objectif consistait à construire un système de suggestions de formations aux salariés, a débuté en février 2016. Ce projet a réuni une chef de projet rattachée à la DRH (nous-même), une équipe de quatre *data scientists* internes mais non rattachés à la DRH, la directrice de la formation et des représentants de son équipe. Les premières réunions avec l'ensemble des participants ont servi à cadrer le projet et l'objectif. Par exemple, il a été décidé dans cette première phase de se concentrer sur des recommandations de formations en e-learning uniquement, pour éviter de devoir passer par l'accord managérial – ce qui aurait été le cas si l'algorithme avait suggéré des formations en présentiel.

À cette première phase caractérisée par des ambitions assez élevées de la part des différents acteurs (définition d'un dispositif final reliant l'entretien annuel, les souhaits de mobilité, les itinéraires déjà existants des salariés...) a succédé une phase de confrontation aux données disponibles. Il s'est avéré que de nombreuses données qui auraient pu être utiles pour suggérer des formations étaient, soit inexistantes, soit de mauvaise qualité. Par exemple, l'entreprise ne disposait pas de données sur les compétences individuelles des collaborateurs. Finalement, les données suivantes, identifiées à la fois par les *data scientists* et les acteurs de la formation, ont pu être collectées et mobilisées pour le projet : historique de formation, données d'identification (statut managérial, classification, domaine métier), communautés et personnes suivies sur le réseau social interne.

La phase suivante a consisté à construire la base de données. Pour cela, la chef de projet a envoyé un mail à 10 000 salariés, leur décrivant le but du projet et leur demandant leur accord pour utiliser leurs données à des fins de suggestions de formations. 1 700 salariés ont répondu positivement en mai 2016. Leurs données ont ainsi pu être rassemblées par l'équipe de gestion

des données RH et concaténées dans une base fournie aux *data scientists*. De plus, les *data scientists* ont dû à ce stade pallier l'absence de données sur certains sujets par la création de nouvelles données à partir de l'existant. Par exemple, ils ont estimé que l'absence d'informations sur la satisfaction des apprenants à l'égard des formations suivies pouvait nuire à la qualité des recommandations finales. De ce fait, ils ont mis au point une méthode permettant d'extrapoler cette satisfaction à partir des données existantes (temps passé sur la formation, nombre de connexions, pourcentage de réussite au quizz final le cas échéant) en se fondant sur des hypothèses comportementales. Par exemple, ils ont supposé qu'une personne ayant passé moins de cinq minutes sur une formation ne l'avait pas appréciée, et qu'à l'inverse se connecter plusieurs fois pour suivre une formation témoignait d'une bonne appréciation de celle-ci. Cela a finalement permis aux *data scientists* de produire une sorte d'indicateur de satisfaction à l'égard des formations, certes pas totalement satisfaisant, mais qui palliait l'absence initiale de données.

Les mois suivants ont été consacrés à la construction de l'algorithme, qui a donné lieu à de nombreux tâtonnements et échanges entre les différents acteurs (chef de projet, *data scientists*, acteurs de la formation). L'algorithme final conçu par les *data scientists* était composé de trois sous-algorithmes. Premièrement, un algorithme de filtrage collaboratif, équivalent à ceux utilisés par Netflix ou Amazon, a été mobilisé. Le filtrage collaboratif fonctionne sur le principe suivant : si deux individus A et B ont le même historique (d'achat de produits, de formations suivies), mais que B en a un peu plus que A (un autre produit acheté, une autre formation suivie), alors ces choix supplémentaires de B peuvent intéresser A et lui seront donc recommandés. Comme l'ont souligné les *data scientists* pour arguer de l'efficacité de ce type d'algorithme, ce principe très basique fonctionne plutôt bien pour Amazon et Netflix. Cependant, cet algorithme n'a permis de suggérer des formations qu'aux salariés ayant un historique de formation suffisamment riche pour être rapproché de l'historique d'autres salariés (environ deux tiers des participants). Pour les salariés n'ayant pas de suggestions à ce stade, un algorithme de filtrage thématique a été utilisé. Il s'agissait de rapprocher les formations suivies (historique de formation) de formations non suivies du catalogue au moyen d'outils d'analyse sémantique. Le rapprochement par mots-clés permet de suggérer des formations ayant des contenus proches de formations déjà suivies par l'utilisateur. Cette seconde méthode a permis de suggérer des formations à une centaine de salariés. Les *data scientists* ont rapproché les salariés restants de salariés ayant eu des suggestions grâce aux variables RH. Plus précisément, le dernier sous-algorithme qu'ils ont conçu considérait comme « proches » des salariés du même domaine métier, même niveau de classification, même statut managérial, suivant les mêmes groupes sur le réseau social interne, et attribuait alors à un salarié n'ayant pas encore de suggestions les suggestions envoyées à un salarié proche. Ce fonctionnement a permis finalement d'envoyer des suggestions à l'ensemble des salariés inscrits, en janvier 2017.

Les différents acteurs se sont mis d'accord pour évaluer la pertinence de l'algorithme sur plusieurs éléments. Tout d'abord, un questionnaire non anonyme (de façon à pouvoir rattacher un répondant aux suggestions reçues) a été envoyé à l'ensemble des participants, leur demandant notamment ce qu'ils pensaient des suggestions qui leur avaient été faites et s'ils pensaient les suivre. Ensuite, les données liées au fait de suivre des formations en ligne ont été récupérées, de manière à évaluer si les participants avaient suivi les suggestions, et si cela avait conduit à une hausse du recours à la formation en ligne.

Les nouveaux paradigmes de quantification introduits par le Big Data en RH

Ces deux projets ainsi que, dans un souci de comparaison, les informations obtenues sur les autres projets, nous permettent à présent d'identifier des éléments propres au Big Data RH, que nous qualifions à ce stade de « paradigmes de quantification » avant de les rattacher à la notion de convention dans la discussion. Nous suivons l'ordre logique des éléments évoqués *supra* : sélection des données, choix des méthodes, critères d'évaluation des algorithmes, et nous ajoutons une dimension correspondant aux modes de fonctionnement. Pour la sélection des données, nous observons un passage du paradigme de l'économie du modèle à celui de l'exhaustivité du modèle. Sur le choix des méthodes, nous observons d'une part le passage du paradigme de la causalité à celui de la corrélation, en lien avec le passage d'un modèle explicatif à un modèle prédictif, et un passage du paradigme de la segmentation à celui de la personnalisation. Sur les modes de fonctionnement, nous observons un passage du paradigme du « meilleur modèle » à celui de l'amélioration continue. Les critères d'évaluation des algorithmes apparaissent de façon transversale dans les différentes sections. Pour chaque paradigme, nous commençons par évoquer les caractéristiques de l'usage classique des statistiques, avant de passer à ce que le Big Data change, en tout cas dans les projets étudiés.

Du paradigme de l'économie du modèle à celui de son exhaustivité

La statistique a longtemps mis en avant la nécessité de l'économie du modèle, autrement dit, le fait d'utiliser le moins de données possible pour expliquer le plus d'informations possible. Cela permet de favoriser l'interprétation que l'on pourra faire du modèle, un modèle avec de très nombreuses variables étant plus difficile à interpréter qu'un modèle plus économe. C'est d'autant plus vrai dans le domaine des RH, où la capacité à interpréter le modèle sera sans doute plus importante que le modèle lui-même, du fait du souhait déjà évoqué de pouvoir en tirer des pistes d'action. Le projet interne d'étude de l'absentéisme illustre bien cette volonté : alors que la base initiale contenait plus d'une centaine de variables, le modèle final produit par la *data scientist*

et présenté au DRH, aboutissement de comparaisons entre différents modèles et de sélection de variables, faisait ressortir une vingtaine de variables explicatives.

Dans les cas de Big Data RH observés, l'observation et les entretiens révèlent que la préoccupation des *data scientists* consiste plutôt à utiliser le plus grand nombre de variables possible. Les *data scientists* déplorent ainsi régulièrement l'absence de telle ou telle donnée.

« Après, [sur la] construction de l'algo, on a eu une très grosse difficulté, on avait des données pourries, on n'avait pas assez qualifié les données en amont. Deux difficultés : [des données issues du réseau social interne] très dures à exploiter, et un catalogue [de formations] qui bougeait en permanence. Avec très peu de notes, peu d'usages, un catalogue mouvant... »
(*Data scientist*, Formazon)

Les autres acteurs (Directrice du recrutement, de la formation...), n'étant pas experts sur les méthodologies quantitatives mobilisées, remettent peu en cause cette préoccupation. Cela incite finalement les acteurs à collectivement identifier et mobiliser toutes les sources de données disponibles : SIRH, mais aussi CV et offres (dans le cas de Shortlist), ou encore communautés suivies sur le réseau social et descriptifs des formations (dans le cas de Formazon). Cette quête s'accompagne donc d'une hétérogénéité des données, à la fois sur le plan des sources, et sur le plan des types de données (structurées et non structurées).

Selon les *data scientists*, les données non structurées (textes notamment) offrent ainsi de nombreuses potentialités encore sous-exploitées. En effet, elles se présentent comme des données brutes, non formatées à la base. Les méthodes de traitement de ces données non structurées, par exemple l'analyse sémantique mobilisée dans Shortlist, n'imposent de plus pas de formatage initial ou relevant d'un parti pris humain. Les nuages de mots sont construits à partir des textes des CV et des offres, par apprentissage automatique de la machine. À ce titre, ces données peuvent sembler plus exhaustives aux yeux des acteurs, RH ou *data scientists*, selon l'argument que l'expression libre est plus riche que la donnée formatée.

« Le gros problème des données RH, c'est que soit elles sont structurées à l'avance, mais du coup elles sont très globales, soit tu es confronté au mur de la donnée non structurée, l'expression libre, ça demande pas mal de ressources pour traiter cette info. Donc le gros enjeu RH, [sur] toutes ces données-là liées à ce que sont et ce que peuvent faire les gens, c'est de capter la donnée puis de traiter cette captation massive d'infos, pour ne pas être débordé. » (Entrepreneur, solution sur les compétences et les suggestions de postes)

Cette recherche d'exhaustivité dans les données mobilisées nécessite par ailleurs dans certains cas la création de nouvelles données. On a ainsi évoqué le cas de l'évaluation de la formation par les apprenants, donnée jugée importante par les *data scientists* pour la bonne qualité du modèle, et approximée par le temps passé sur la formation et le nombre de connexions.



Photo © The Quilters' Guild of the British Isles / BRIDGEMAN IMAGES

Penrith Quilt, créé par John Wharton, 1900. Quilt Museum and Gallery, York.

« La recherche se heurte à l'hétérogénéité des données, à la fois sur le plan des sources et sur le plan des types de données, structurées ou non. »

Finalement, alors que les statistiques se fondent généralement sur un arbitrage entre qualité explicative et économie du modèle, ce dernier critère semble absent dans les projets de Big Data RH⁽⁵⁾.

Du paradigme de la causalité à celui de la corrélation : de l'explicatif au prédictif

Comme le soulignent Saunders, Lewis et Thornhill (2016), les méthodes quantitatives s'inscrivent généralement dans une posture positiviste, une démarche hypothético-déductive et un paradigme de la causalité : elles servent ainsi à identifier des liens de causalité entre des variables. Le raisonnement « toutes choses égales par ailleurs » est d'ailleurs tout à fait représentatif de cette démarche tendant à évaluer l'effet d'une variable sur une autre (Martin, 2012). Si cette recherche de causalité est si importante, c'est que le sens et l'interprétation du modèle sont clés. C'est notamment le cas en RH, où, comme on l'a vu, la démarche causale est régulièrement mobilisée, par exemple pour identifier l'effet du *turnover* sur la satisfaction des clients (Rasmussen et Ulrich, 2015).

Or, dans les projets de Big Data RH étudiés, la recherche de causalité semble secondaire, voire inexistante. Ainsi, dans Shortlist, le modèle proposé par les *data scientists* ne s'intéresse pas aux variables qui peuvent influencer sur la performance des candidats, ne cherche pas à identifier leurs effets. Il mobilise uniquement des corrélations (au sens de : liens entre deux variables, sans sens particulier), mesurées grâce à la cooccurrence de mots-clés entre des CV et des offres. Formazon s'inscrit dans une logique similaire. Le modèle produit par les *data scientists* se fonde en effet sur un rapprochement entre les individus en fonction de leur historique de formation (sous-algorithme 1) et de leurs caractéristiques (sous-algorithme 3), et sur un rapprochement entre les formations en fonction de leurs mots-clés (sous-algorithme 2). Les autres projets de Big Data RH étudiés grâce aux entretiens s'inscrivent également dans une logique de cooccurrence de mots-clés et de rapprochement.

On observe donc que les acteurs, *data scientists* comme RH, s'éloignent du paradigme de la causalité au profit du paradigme de la corrélation. Cela s'explique par le fait que le but de leurs projets et donc des modèles consiste non pas à expliquer un phénomène, mais plutôt à prédire de façon adéquate un événement (la performance d'un candidat, le goût d'un salarié pour une formation). Or, la causalité, nécessaire pour expliquer un phénomène, ne l'est pas pour en prédire un : la corrélation peut suffire. En effet, si deux variables sont liées entre elles, disposer d'une de ces variables suffit à prédire l'autre – et ce, même si le lien renvoie à une corrélation et non pas à une causalité. Par exemple, dans le cas de Formazon, l'algorithme de filtrage collaboratif, présenté comme le plus efficace par les *data*

scientists, repose sur une mesure de corrélation, et pas de causalité, mais vise bien à prédire un goût ou un comportement.

On peut supposer que cet éloignement vis-à-vis de la causalité traduit un éloignement vis-à-vis du sens et de l'interprétation du modèle. Par exemple, ici, le fait que le lien entre deux variables (être proche en matière de profil, et avoir des besoins en formation équivalents) soit très indirect a peu d'importance aux yeux des *data scientists* et des acteurs de la formation. Si cet éloignement est plus prononcé chez les *data scientists* que chez les acteurs RH (recrutement, formation), il n'en reste pas moins réel pour les deux catégories d'acteurs.

L'observation révèle d'ailleurs que la qualité des algorithmes de Formazon et Shortlist comme de celui du projet externe à Multinum de suggestions de postes a été mesurée en fonction de leur qualité prédictive, et non pas en fonction de ce que leur interprétation pouvait apporter aux politiques de formation et de recrutement.

« On a obtenu 84 % de matching. Je fais 2 exercices en parallèle : je prends une offre, je prends les candidatures, et ils [les chargés de recrutement] font le scoring (le classement des offres) à la main. Ils disent : "Parmi toutes les candidatures reçues, voici les 20 premiers". Et on compare ça avec Shortlist. » (*Data scientist, Shortlist*)

À l'inverse, les projets reposant sur une utilisation plus classique des méthodes quantitatives en RH sont centrés sur le paradigme de causalité et l'interprétation du modèle. Par exemple, les projets (interne et externe) d'étude de l'absentéisme, fondés sur des méthodes économétriques, reposaient essentiellement sur une interprétation du modèle d'explication de l'absentéisme : identifier les variables causales de l'absentéisme, et donc les leviers dont l'entreprise dispose pour le réduire.

« Il fallait garder en tête qu'on ne voulait pas prévoir qui serait absent, mais plutôt savoir pourquoi ils étaient absents. Donc il fallait qu'on puisse interpréter les résultats, c'est pour ça que je ne suis pas partie sur des algorithmes hyper prédictifs mais qu'on a du mal à interpréter, type réseau de neurones ou forêt aléatoire. » (*Data scientist, étude sur l'absentéisme*)

Du paradigme de la segmentation à celui de la personnalisation

Des techniques statistiques sont régulièrement utilisées pour faire de la segmentation, ou des typologies, c'est-à-dire identifier des groupes d'individus ayant un comportement homogène. Ainsi, un des résultats de l'étude sur l'absentéisme menée chez Multinum consistait à identifier six « profils-types » d'absentéisme, au moyen de techniques de segmentation.

En l'occurrence, les projets de Big Data RH étudiés vont plus loin que la segmentation en mentionnant la notion de personnalisation. Cette notion, qui apparaît de façon très marquée dans les discours et échanges entre les acteurs RH, et par exemple dans les noms donnés aux projets (« suggestions personnalisées de formations »),

⁽⁵⁾ Malgré les coûts que représentent les serveurs de stockage des données... Cela peut s'expliquer par la volonté affirmée de Multinum de soutenir l'émergence du Big Data RH, justifiant des efforts financiers.

renvoie à l'idée d'adapter un contenu à une personne et non plus à un groupe : suggérer la ou les formations qui intéresseront le plus un salarié, identifier à un niveau individuel qui pourrait constituer un bon candidat dans le cadre du recrutement, par exemple. Cela correspond à un propos apparemment paradoxal : parvenir à une personnalisation en mobilisant des données, c'est-à-dire en caractérisant les individus par des modalités communes.

« L'idée c'est de pousser de l'info ajustée, personnalisée aux collaborateurs. [...] Pour nous c'est ça la clé. Il ne faut rien de générique. Si tu commences à généraliser, forcément tu vas laisser des gens sur le côté. »
(Entrepreneur, solution sur les compétences et les suggestions de postes)

Sur le plan des pratiques et des méthodes mobilisées, le matériau d'observation montre que cela écarte les méthodes de segmentation usuelles (classification ascendante hiérarchique par exemple), qui permettent d'identifier des groupes et pas de descendre à un niveau individuel. Le filtrage collaboratif déjà évoqué constitue un exemple de méthode permettant la personnalisation. En effet, les différentes suggestions reçues par une personne peuvent provenir de rapprochements avec différentes personnes. Le filtrage collaboratif peut donc aboutir à une situation où chaque individu reçoit un ensemble unique de suggestions, ce qui correspond bien à l'idée de personnalisation.

Du paradigme du « meilleur modèle » au paradigme de l'amélioration continue

Dans un usage classique des méthodes quantitatives, les statisticiens présentent un ou plusieurs modèles identifiés au cours de la recherche comme étant les meilleurs modèles. C'est par exemple ce que fait le prestataire proposant des analyses de l'absentéisme : il effectue plusieurs tests de modélisation, puis présente un seul modèle à ses clients. C'est le cas aussi pour le projet d'analyse des verbatims mené chez Multinum : l'analyse permet de produire un ensemble de résultats, dont les plus intéressants sont présentés puis utilisés pour prendre des décisions.

Or, dans les projets de Big Data RH, la démarche itérative se prolonge bien après le premier envoi de résultats. Ainsi, dans le cas de Formazon, il y a eu de nombreux tests et itérations au sein de l'équipe projet avant l'envoi des premières suggestions, pour améliorer autant que possible leur qualité. Mais ces itérations ne suffisaient pas aux yeux des *data scientists*. Ils ont ainsi insisté fortement sur la nécessité de recueillir les retours des utilisateurs, par exemple en laissant aux salariés la possibilité de refuser une suggestion effectuée. Ces nouvelles données, produites par les utilisateurs, devaient permettre d'améliorer l'algorithme de suggestion. De la même façon, dans le cas de Shortlist, l'algorithme était pensé par les *data scientists* pour continuer à apprendre en continu, au fur et à mesure de la réception de nouveaux CV ou de nouvelles offres. Nous qualifions ce phénomène de paradigme de l'amélioration continue, au sens où, plutôt que de délivrer un

produit fini, les acteurs délivrent un produit (l'algorithme) qui devra s'améliorer constamment avec le temps – et notamment avec les nouvelles données produites par les utilisateurs.

Cependant, l'observation montre que ce paradigme d'amélioration continue est mal accepté par les salariés, qui ont plutôt l'habitude d'accéder à des produits finis et non pas à des produits en cours d'amélioration. Cela nécessite donc des efforts de pédagogie. Ainsi, dans le cas de Formazon, les retours des salariés sur les premières suggestions de formations révélaient souvent de très fortes attentes à l'égard de la qualité des suggestions, voire un mécontentement à l'égard de l'entreprise sur le sujet, alors qu'il leur avait bien été spécifié qu'il s'agissait d'une expérimentation, et que l'algorithme devrait s'améliorer au fur et à mesure de leurs retours. Ce phénomène illustre le difficile passage d'un fonctionnement où on fournit un produit fini à un fonctionnement où le produit (l'algorithme) continue à évoluer après avoir été fourni à ses destinataires, ici les salariés.

Discussion

Nous cherchions à identifier dans quelle mesure l'émergence du Big Data RH s'accompagne de nouvelles conventions de quantification. Nous avons donc identifié et étudié les nouveaux paradigmes introduits par le Big Data dans les RH, en lien avec les différents éléments que Chiapello et Walter (2016) présentent comme sous-jacents aux conventions de quantification : choix des données, sélection des méthodes, définition des critères d'évaluation, notamment. Nous constatons que l'émergence du Big Data s'accompagne de changements de paradigmes : passage de l'économie du modèle à son exhaustivité pour la sélection des données, passage du paradigme de la causalité à celui de la corrélation et passage de la segmentation à la personnalisation pour le choix des méthodes, et passage du « meilleur modèle » à l'amélioration continue pour les modes de fonctionnement.

Ces différents éléments nous permettent d'entamer une discussion avec la littérature, à la fois sur les statistiques et sur les RH.

Passage de l'économie du modèle à son exhaustivité. L'observation montre que cela nécessite la création de nouvelles données venant compenser l'absence ou le manque de fiabilité d'autres données. Il s'agit alors d'opérations de « mise en statistique » (Desrosières, 1993), que Juven (2016) décrit dans un autre contexte. Ces opérations consistant à créer de nouvelles données à partir de données existantes ne sont donc pas propres au Big Data. En revanche, le changement introduit par ce nouveau paradigme réside dans le fait que la question du nombre de variables mobilisées se pose toujours sous l'angle de son augmentation et jamais sous l'angle de sa réduction. Par ailleurs, cette dimension d'exhaustivité passe aussi par une volonté de recueillir autant de variables que possible, sans formuler d'hypothèse *ex-ante* sur leur pertinence.

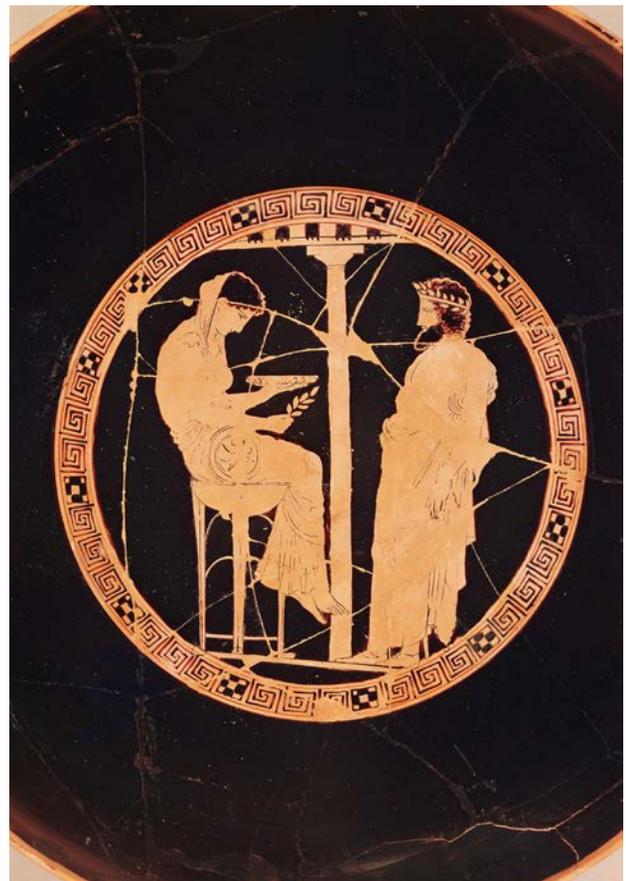
Passage de la causalité à la corrélation. Ce point avait été identifié par Mayer-Schönberger et Cukier (2014). Il se traduit ici par la perte d'hégémonie du raisonnement « toutes choses égales par ailleurs » au profit d'autres méthodes fondées sur des corrélations et des rapprochements (filtrage collaboratif, rapprochements de mots-clés...). Nous pouvons approfondir cette réflexion en supposant dans la lignée de Kitchin (2014) que cette évolution traduit aussi un changement de démarche et de paradigme épistémologique. En effet, la démarche hypothético-déductive, fondée le plus souvent sur des analyses causales, souffre possiblement de ces changements. Ainsi, l'idée de recueillir le plus de données possible, même des données dont on ne sait pas forcément à l'avance ce que l'on pourra en faire, tout comme la prégnance des corrélations, va à l'encontre de la démarche hypothético-déductive qui teste souvent l'effet d'un nombre précis de variables sur d'autres. Kitchin (2014) annonce alors l'émergence d'une *data-driven science*, « combinaison hybride d'approches abductive, inductive et déductive »⁽⁶⁾ (p. 5). Ici, nous voyons que les modélisations propres au Big Data RH, qui se fondent plutôt sur la découverte de rapprochements, et qui ne cherchent pas forcément à les expliciter mais se contentent de les utiliser dans une visée prédictive, contribuent à brouiller les contours de la posture positiviste. En effet, on y retrouve certains éléments propres à la posture positiviste, notamment l'hypothèse d'ontologie réaliste (il existe un réel indépendant de l'intérêt du chercheur). En revanche, le principe méthodologique d'hypothèse causaliste qui sous-tend ce paradigme (Gavard-Perret *et al.*, 2013) semble moins présent dans le cas du Big Data RH.

Passage du paradigme de la segmentation à celui de la personnalisation. Ce passage est important car il réfute un élément de base des statistiques, à savoir le fait de considérer des groupes d'individus, des moyennes, des fréquences, et non des individus spécifiques. Il remet en cause également une posture usuelle des RH, consistant à mener des politiques de gestion segmentées, et pas individualisées. Ce passage peut paraître d'autant plus étonnant que le Big Data repose sur l'idée d'une grande quantité de données et, entre autres, d'individus. Pour autant, dans les différents exemples étudiés, le projet nécessite bien *in fine* de revenir aux individus, pour leur suggérer des formations, ou les présélectionner dans le cadre d'un recrutement, par exemple. Enfin, ce passage renvoie à l'évolution de la GRH vers une plus grande personnalisation (Arnaud, Frimousse et Peretti, 2009).

Passage du paradigme du « meilleur modèle » à celui de l'amélioration continue. Cette évolution a des implications avant tout pratiques et opérationnelles pour les RH. Comme nous l'avons vu, cela nécessite une communication spécifique auprès des salariés, tout comme une conduite de projet adaptée.

Pour finir, nous pouvons nous demander si ces paradigmes propres au Big Data constituent autant de nouvelles conventions en RH, ou sous-tendent une

seule et unique convention regroupant ces différents changements. Il semble que la dimension de prédiction apparaisse comme transversale à ces évolutions et puisse constituer une convention de quantification propre au Big Data RH. Ainsi, c'est la recherche de la prédiction qui explique que l'économie du modèle ne constitue plus un critère de sa pertinence. En effet, c'est avant tout la nécessité de l'interprétation du modèle qui justifie le critère d'économie ; si l'interprétation disparaît au profit de la qualité prédictive, alors il faut mobiliser le plus grand nombre de variables possible. C'est aussi la recherche de la prédiction qui explique la fin du paradigme de causalité au profit de celui de la corrélation : comme l'écrivent Mayer-Schönberger et Cukier (2014, p. 80), « À l'ère des Big Data, ces analyses [par corrélations] conduiront au bout du compte à faire surgir des connaissances nouvelles et des prédictions utiles. ». La recherche de prédiction explique aussi le passage de la segmentation à la personnalisation. En effet, une segmentation ne peut pas prédire de façon précise ce qui plaira ou conviendra aux individus, puisqu'elle s'intéresse aux groupes et non pas aux individus, contrairement à la personnalisation. Enfin, la recherche de prédiction explique aussi la nécessité de l'amélioration continue : c'est en recueillant de façon continue les retours des utilisateurs que l'on



Peintre de Kodros, *Égée consultant l'oracle de Delphes*, coupe à figures rouges v. 430 av. J.C. Berlin, Antikensammlung der Staatlichen Museen.

« La recherche de prédiction explique aussi le passage de la segmentation à la personnalisation. En effet, une segmentation ne peut pas prédire de façon précise ce qui plaira ou conviendra aux individus, puisqu'elle s'intéresse aux groupes et non pas aux individus, contrairement à la personnalisation. »

⁽⁶⁾ Traduction par nos soins. Texte original : "A hybrid combination of abductive, inductive and deductive approaches"

peut améliorer les prédictions. Finalement, à l'inverse de Sivarajah *et al.*, (2017), nous considérons la dimension prédictive comme essentielle au Big Data, et non pas comme une simple dimension parmi d'autres des techniques de Big Data.

Conclusion

Dans cette recherche, nous avons donc identifié quatre évolutions propres au Big Data RH, que nous avons liées à l'apparition d'une nouvelle convention de quantification, celle de la prédiction. D'un point de vue théorique, cela permet une réflexion à la fois sur la posture de la fonction RH, sur les méthodes quantitatives mobilisées et leurs implications, et une discussion autour de la prégnance de la corrélation sur la causalité. Cela représente un enjeu important pour les recherches portant sur l'utilisation de la quantification au sein des organisations. Cela offre aussi l'opportunité de réfléchir à la définition d'une convention de quantification. On constate en effet que l'émergence d'une nouvelle convention de quantification, celle de la prédiction, induit des changements à la fois dans la collecte des données, les méthodologies utilisées, les modes de fonctionnement, et l'évaluation de la pertinence des modèles. Notre étude permet aussi de mieux définir le Big Data RH et d'identifier des évolutions à la fois méthodologiques et pratiques liées à son émergence. C'est d'autant plus important que les travaux sur le sujet sont encore peu nombreux.

D'un point de vue empirique, notre travail mobilise une ethnographie organisationnelle de longue durée et montre la pertinence de se pencher sur la construction des algorithmes de Big Data et leurs implications. Enfin, sur le plan managérial, notre travail ouvre des pistes de pratiques à développer pour les entreprises. En particulier, la diffusion d'une culture plus propice à l'amélioration continue pourrait constituer une action intéressante pour les entreprises souhaitant développer des démarches de Big Data, que ce soit en RH ou dans d'autres domaines.

Cependant, ce travail présente des limites qui ouvrent de nouvelles perspectives de recherche. Ainsi, les exemples utilisés dans cette étude de cas ne concernent que le Big Data en RH et limitent donc le potentiel de généralisation au Big Data en général. De plus, cette étude ne rend pas compte des possibles résistances à l'égard du Big Data RH. Il serait donc intéressant de prolonger cette étude par une recherche de plus grande ampleur sur le Big Data dans les organisations. Cela permettrait notamment d'étudier la question du partage de la convention de prédiction entre différents acteurs, en fonction des différents stades d'appropriation du Big Data. Cette question en appelle deux autres : comment les acteurs procèdent-ils pour gérer les éventuelles tensions entre deux conventions concurrentes ? Comment une convention qui nécessite un certain bagage méthodologique, comme celle du Big Data, peut-elle être partagée largement au sein d'une organisation ?

Bibliographie

- ANGRAVE D., CHARLWOOD A., KIRKPATRICK I., LAWRENCE M. & STUART M. (2016), "HR and Analytics: Why HR Is Set to Fail the Big Data Challenge", *Human Resource Management Journal* 26 (1), pp. 111. <https://doi.org/10.1111/1748-8583.12090>.
- ARNAUD S., SOUFYANE F. & PERETTI J.-M. (2009), « Gestion personnalisée des ressources humaines : implications et enjeux », *Management & Avenir* 8 (28), pp. 294314. <https://doi.org/10.3917/mav.028.0294>.
- BELLO-ORGAZ G., JUNG J. J. & CAMACHO D. (2016), "Social Big Data: Recent Achievements and New Challenges", *Information Fusion* 28 (mars), pp. 4559. <https://doi.org/10.1016/j.inffus.2015.08.005>.
- BOUDREAU J. W. & RAMSTAD P. M. (2004), "Talentship and the Evolution of Human Resource Management: From 'Professional Practices' To 'Strategic Talent Decision Science'", CEO Publication, Center for Effective Organizations.
- BOULLIER D. (2017), « Pour des sciences sociales de troisième génération (SS3G) : Des traces numériques aux répliques », in MENER P.-M. & PAYE S., *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, pp. 163184. Conférences. Paris, Collège de France. <http://books.openedition.org/cdf/5011>.
- CARDON D. (2015), À quoi rêvent les algorithmes : nos vies à l'heure des big data. Coll. La république des idées. Paris, Seuil.
- CARTIER M. & LIARTE S. (2010), « Timing d'entrée, incertitude et agglomération temporelle : le cas de l'industrie cinématographique hollywoodienne ». *M@n@gement* 13 (2), pp. 7098. <https://doi.org/10.3917/mana.132.0070>.
- CHIAPELLO E. & GILBERT P. (2013), *Sociologie des outils de gestion. Introduction à l'analyse sociale de l'instrumentation de gestion*. Paris, La Découverte.
- CHIAPELLO E. & WALTER Ch. (2016), "The Three Ages of Financial Quantification: A Conventionalist Approach to the Financiers' Metrology". *Historical Social Research* 41 (2), pp. 155177. <https://doi.org/10.12759/hsr.41.2016.2.155-177>.
- COCHOY F. & VAYRE J.-S. (2017), « Les big data à l'assaut du marché des dispositifs marchands : une mise en perspective historique », in MENER P.-M. & PAYE S., *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, pp. 2746. Paris, Collège de France. <https://doi.org/10.4000/books.cdf.4992>.
- COSSETTE M., LEPINE C. & RAEDECKER M. (2014), « Mesurer les résultats de la gestion des ressources humaines : principes, état des lieux et défis à surmonter pour les professionnels RH », *Gestion* 39 (4), pp. 4465. <https://doi.org/10.3917/riges.394.0044>.
- DAVENPORT T. H., HARRIS J. & SHAPIRO J. (2010), "Competing on Talent Analytics", *Harvard Business Review*, pp. 16.
- DESROSIERES A. (1993), *La Politique des grands nombres. Histoire de la raison statistique*. 2e éd. Paris, La Découverte.
- DESROSIERES A. (2008a), *Pour une sociologie historique de la quantification. L'Argument statistique I*, Paris, Presses des Mines. <https://doi.org/10.4000/books.pressesmines.901>.
- DESROSIERES A. (2008b), *Gouverner par les nombres. L'Argument statistique II*, Paris, Presses des Mines. <https://books.openedition.org/pressesmines/341>.
- DIAZ-BONE R. (2016), "Convention Theory, Classification and Quantification". *Historical Social Research* 41 (2), pp. 4871. <https://doi.org/10.12759/hsr.41.2016.2.48-71>.
- DIAZ-BONE R. & THEVENOT L. (2010), « La sociologie des conventions. La théorie des conventions, élément central des nouvelles sciences sociales française ». *Trivium*, n°5, pp. 116. <https://doi.org/10.4000/ifha.347>.
- EREVELLES S., FUKAWA N. & SWAYNE L. (2016), "Big Data Consumer Analytics and the Transformation of Marketing", *Journal of Business Research*, 69 (2), pp. 897904. <https://doi.org/10.1016/j.jbusres.2015.07.001>.

- ESPELAND W. N. & STEVENS M. L. (1998), "Commensuration as a Social Process", *Annual Review of Sociology*, 24 (1), pp. 313-43. <https://doi.org/10.1146/annurev.soc.24.1.313>.
- EYMARD-DUVERNAY F. (1989), « Conventions de qualité et formes de coordination ». *Revue économique*, 40 (2), pp. 329-60.
- GANDOMI A. & HAIDER M. (2015), "Beyond the Hype: Big Data Concepts, Methods, and Analytics". *International Journal of Information Management*, 35 (2), pp. 137-44. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
- GAVARD-PERRET M.-L., GOTTELAND D., HAON C. & JOLIBERT A. (2013), *Méthodologie de la recherche*, 2e éd., Pearson.
- GILLES M. (2013), « Compter pour peser ? La quantification de la « santé au travail » en entreprises : entre institutionnalisation et ambiguïtés conceptuelles », EHESS.
- GROSJEAN S. & GROLEAU C. (2013), « L'ethnographie organisationnelle aujourd'hui », *Revue internationale de psychosociologie et de gestion des comportements organisationnels*, n° supplément HS, pp. 13-23. <https://doi.org/10.3917/rips1.hs01.0011>.
- HUSELID M. A. (2018), "The Science and Practice of Workforce Analytics: Introduction to the HRM Special Issue", *Human Resource Management*, 57 (3), pp. 679-84. <https://doi.org/10.1002/hrm.21916>.
- JUVEN P.-A. (2016), *Une santé qui compte ? Les coûts et les tarifs controversés de l'hôpital public*, Paris, PUF.
- KITCHIN R. (2014), "Big Data, New Epistemologies and Paradigm Shifts", *Big Data & Society*, juillet, pp. 1-12. <https://doi.org/10.1177/2053951714528481>.
- KOVACH K. A., HUGHES A. A., FAGAN P. & MAGGITT P. G. (2002), "Administrative and Strategic Advantages of HRIS", *Employment Relations Today*, 29 (2), pp. 43-48. <https://doi.org/10.1002/ert.10039>.
- LAVE J. (1984), "The Values of Quantification", *The Sociological Review*, 32 (1), pp. 88-111. <https://doi.org/10.1111/j.1467-954X.1984.tb00108.x>.
- LAWLER E. E., LEVENSON A. & BOUDREAU J. W. (2010), "HR Metrics and Analytics: Use and Impact", *Human Resource Planning*, 2010.
- LEVENSON A. (2018), "Using Workforce Analytics to Improve Strategy Execution", *Human Resource Management*, 57 (3), pp. 685-700. <https://doi.org/10.1002/hrm.21850>.
- MADSEN D. & SLÄTTEN K. (2017), "The Rise of HR Analytics: A Preliminary Exploration", in *Global Conference on Business and Finance Proceedings*, 10.
- MARTIN O. (2012), *L'analyse quantitative des données : L'enquête et ses méthodes*, 3e éd. Paris, Armand Colin.
- MAYER-SCHÖNBERGER V. & CUKIER K. (2014), *Big data : la révolution des données est en marche*, Paris, Robert Laffont.
- MCAFEE A. & BRYNJOLFSSON E. (2012), "Big Data: The Management Revolution". *Harvard Business Review*, pp. 39.
- MENGER P.-M. (2017), « Introduction », in MENGER P.-M. & PAYE S., *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, pp. 7-23. Conférences. Paris, Collège de France. <http://books.openedition.org/cdf/4990>.
- OLLION E. & BOELAERT J. (2015), « Les sciences sociales et la multiplication des données numériques », *Sociologie*, 3 (6), pp. 1-20.
- O'NEIL C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. First ed. New York, Crown.
- PAYE S. (2017), « Postface : "Un travail de fourmi" », in MENGER P.-M. & PAYE S., *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, pp. 185-215. Conférences. Paris, Collège de France. <http://books.openedition.org/cdf/5013>.
- RAGUSEO E. (2018), "Big Data Technologies: An Empirical Investigation on Their Adoption, Benefits and Risks for Companies", *International Journal of Information Management*, 38 (1), pp. 187-95. <https://doi.org/10.1016/j.ijinfomgt.2017.07.008>.
- RASMUSSEN T. & ULRICH D. (2015), "Learning from Practice: How HR Analytics Avoids Being a Management Fad", *Organizational Dynamics*, 44 (3), pp. 236-42. <https://doi.org/10.1016/j.orgdyn.2015.05.008>.
- SALAI R. (2016), "Quantification and Objectivity. From Statistical Conventions to Social Conventions". *Historical Social Research*, 41 (2), pp. 118-34. <https://doi.org/10.12759/hsr.41.2016.2.118-134>.
- SAUNDERS M., LEWIS P. & THORNHILL A. (2016), *Research Methods for Business Students*, 7e ed. Harlow, Pearson.
- SCHOUTEN J. W. & MCALEXANDER J. H. (1995), "Subcultures of Consumption: An Ethnography of the New Bikers", *Journal of Consumer Research*, 22 (1), pp. 43-61.
- SIVARAJAH U., KAMAL M. M., IRANI Z. & WEERAKKODY V. (2017), "Critical Analysis of Big Data Challenges and Analytical Methods", *Journal of Business Research*, 70 (janvier), pp. 263-86. <https://doi.org/10.1016/j.jbusres.2016.08.001>.
- STEVENS M. L. (2008), "Culture and Education", *The Annals of the American Academy of Political and Social Science*, 619 (1), pp. 97-113. <https://doi.org/10.1177/0002716208320043>.
- SUNSTEIN C. R. (2000), "Cognition And Cost-Benefit Analysis", *The Journal of Legal Studies*, 29 (S2), pp. 1059-1103. <https://doi.org/10.1086/468105>.