



HAL
open science

Régime de médiation des “ données ” en contexte professionnel

Ghislaine Chartron

► **To cite this version:**

Ghislaine Chartron. Régime de médiation des “ données ” en contexte professionnel. CIDE 22 22e Colloque International sur le Document Numérique, Dec 2021, Paris, France. hal-03500937

HAL Id: hal-03500937

<https://hal.science/hal-03500937>

Submitted on 22 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Régime de médiation des « données » en contexte professionnel

G. CHARTRON, Professeur du CNAM, DICEN-IDF, Directrice INTD
Ghislaine.chartron@lecnam.net

Colloque CIDE22, 9-10 dec 2021, CNAM-Paris

<http://europia.org/cide22/>

A paraître chez Europia

1. Introduction et contextualisation

Dans un contexte où les données numériques sont devenues inflationnistes, focalisant les projections d'innovations et de nouvelles croissances, soutenues par des prescriptions politiques affirmées notamment en Europe (open data, données de la recherche, sciences des données...), les sciences de l'information qui ont forgé leur identité originelle sur le document sont aujourd'hui questionnées par rapport à la croissance de toutes ces données hétérogènes. Dans les discours dominants, les données englobent désormais à la fois les documents textuels, les données factuelles, les données des objets connectés, les capteurs, les échanges libres des réseaux sociaux... toutes réunies dans l'espace numérique.

L'objectif de cette communication n'est pas de revenir sur les distinctions épistémologiques pertinentes entre « données, documents, information, connaissance » mais de prendre acte de la situation d'amalgame qui s'impose désormais dans le monde socio-économique. Documents, données structurées, non structurées constituent un tout global, assimilé à une matière première informationnelle pour laquelle il s'agit d'élaborer les médiations pertinentes permettant de construire du sens répondant à différents objectifs des organisations. Les frontières entre documents, données s'estompent dans le projet d'une intelligibilité globale. Prenons l'exemple d'une veille informationnelle, qualifiée désormais très souvent comme « veille à 360° » agrégeant les contenus des médias, les auto-publications des blogs, les commentaires des réseaux sociaux, les sources libres de l'open data (Mesguisch, 2021). Certains archivistes se demandent même s'il ne faut pas abandonner le terme d'archives au profit des données (Banat-Berger, Nougaret, 2014). La stratégie européenne des données semble également amalgamer tout type de données essentiellement au profit du développement de l'intelligence artificielle désormais (Commission, 2020).

Dans la suite de nos précédents travaux qui avaient contribué à éclairer la relation entre les humains et les données, sous l'angle de la création, de la médiation, de la décision et de la narration, travaux inscrits majoritairement sur des terrains de sciences humaines et sociales (Simonnot, Broudoux, Chartron, 2020), cette contribution vise à poursuivre le projet par la caractérisation de ce régime d'interactions avec les données dans le contexte professionnel, majoritairement le contexte des entreprises privées.

Le terrain d'observation est ici circonscrit à celui des missions de stage confiées aux apprentis du master MEDAS (Méga-données et analyse sociale¹) du CNAM pour les deux dernières promotions (soit 32 apprentis). Ce master a été débuté en 2017 au sein de l'INTD-CNAM pour former des *data-analyst*, la 5^{ème} promotion a débuté sa formation en septembre 2021. Le master vise à compléter l'offre de formation de l'Institut. L'INTD a été créé dans les années 50 pour former des spécialistes des techniques documentaires ; il a progressivement intégré les compétences renouvelées liées au traitement de tous les types de documents, aux enjeux de la GED, du KM et du web. Il entreprend de répondre désormais aux enjeux de l'inflation des données qui traversent tous les secteurs, en tentant de comprendre et d'accompagner au mieux le renouvellement des compétences (Arrubarrena, Kembellec, Chartron, 2019).

Nous avons analysé le mémoire de fin d'étude des 32 apprentis du master MEDAS, consultables en grande partie en ligne lorsqu'une clause de confidentialité n'a pas été demandée². L'analyse s'est appuyée sur une grille visant à relever pour chaque mission : le domaine d'activités, le service de rattachement de la mission, les types de données traitées, la finalité poursuivie par l'organisation, les outils utilisés, la génération de valeur pour l'organisation.

2. Secteurs d'activités et rattachement des « missions data » dans les organisations

Le corpus considéré concerne majoritairement des organisations du secteur privé (28/32 missions). Pour classifier le secteur d'activités de ces organisations, nous avons essayé de les ventiler selon la Nomenclature officielle d'Activités productives Françaises (NAF)³. Cette nomenclature pose quelques problèmes notamment en amalgamant dans une seule rubrique les télécoms et les médias (rubrique J), en réunissant dans une seule catégorie des industries « manufacturières » très diverses (rubrique C), en donnant peu de place aux entreprises de la transformation numérique (rubrique N). Toute nomenclature montrant des limites et des arbitraires, nous avons toutefois décider d'adopter cette nomenclature officielle.

	Classe d'activités NAF	Nbre de missions	Organisation d'accueil
A	Agriculture, sylviculture et pêche	1	Ministère Pêche
B	Industries extractives	2	Total Energie, BRGM
C	Industrie manufacturière	3	Pharmacie(Sanofi) , Automobile (PSA), Dassault
D	Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné		
E	Production et distribution d'eau ; assainissement, gestion des déchets et dépollution		
F	Construction		

¹ <https://formation.cnam.fr/rechercher-par-discipline/master-mega-donnees-et-analyse-sociale-medas--1085595.kjsp>

² <https://portaildoc-intd.cnam.fr/ListRecord.htm?list=table&table=3&compl=104>

³ <https://www.insee.fr/fr/information/2406147>

G	Commerce ; réparation d'automobiles et de motocycles		
H	Transports et entreposage	3	SNCF, DHL, Air France-KLM
I	Hébergement et restauration		
J	Information et communication	5	Jeu Video, Canal+, SFR, SFR, agence com
K	Activités financières et d'assurance	8	Allianz, BforBank, BPCE, Aviva, Generali, BNP, GMF, Generali
L	Activités immobilières	1	CAFPI
M	Activités spécialisées, scientifiques et techniques	2	IPSOS, Publicité (Solocal)
N	Activités de services administratifs et de soutien	3	Marketing (OPEN), entrepreneuriat (UP'n BIZ), Numérique (HARDIS)
O	Administration publique	2	Injep, Armées
P	Enseignement		
Q	Santé humaine et action sociale	2	Startup Qare, Pôle emploi
R	Arts, spectacles et activités récréatives		
S	Autres activités de services		
T	Activités des ménages en tant qu'employeurs ; activités indifférenciées des ménages en tant que producteurs de biens et services pour usage propre		
U	Activités extra-territoriales		

Tableau 1 : Secteurs d'activités de production pour les missions des apprentis

Le premier constat pourrait être celui d'une certaine diversité des secteurs concernés. Ce qui traduit le fait que les développements de services autour de la data sont bien transversaux aux secteurs, le tertiaire étant plus représenté dans notre échantillon. Cette transversalité tend à confirmer que les données sont un élément majeur partagé de la transformation digitale des entreprises en général (Nesvijevskaia, 2019), (Morelon, 2020).

Deux secteurs dominant toutefois le corpus des 32 missions :

- Activités financières et d'assurance : 25%
- Information et communication : 16%

Vient ensuite les 3 secteurs suivants avec 9% de représentation :

- Industrie manufacturière,
- Transports et entreposage
- Activités de services administratifs et de soutien

Une autre question d'ordre organisationnel peut être soulevée : A quels services sont rattachées ces « missions data » au sein des organisations ? Peut-on observer des équipes et des services data constituées et autonomes au même titre que les DSI par exemple ? Le tableau 2 nous donne des éclairages au prisme du corpus étudié.

Service de rattachement de la mission	Nbre de missions	Secteurs d'activités
Service data, service spécifique	8	2 Banque, 1 Assurance (K) 1 Ministère Pêche (A) 1 Ministère Education nationale (O) 1 Ministère armées (O) 1 Startup santé (Q) 1 Dassault (C)
Service Marketing	8	2 Info/com media, teleco) (J) 2 Service numérique (N) 2 enquêtes, publicité (M) 1 Automobile (B) 1 Total (B)
Service développement stratégie, pilotage	6	4 Assurance (K) 1 Air France (H) 1 Immobilier (L)
Service Contrôle , gestion	3	1 Transport (H) 1 Infocom (J) 1 Banque (K)
Autres	8	

Tableau 2 : Rattachement des « missions data » dans les organisations

Une maturité croissante de la donnée dans les organisations est observable avec l'identification de structures dédiées transversales particulièrement dans les secteurs de la banque et de l'assurance et de certains Ministères. Ce constat marque une évolution par rapport aux observations de Karoui, Devauchelle et Duzert en 2014 (Karoui et al., 2014).

Le second constat est l'affectation majeure des missions data de nos apprentis aux services Marketing (connaissance client, marketing profilé) et aux services Pilotage et Stratégie (calcul d'indicateurs de performances).

Les services de contrôle de gestion sont en 3^{ème} position pour accueillir des « missions data ». Enfin, les « missions data » restent aussi très diversifiées et peuvent être rattachées au plus près des métiers comme le service RH, le service financier ou le service de gestion des documents d'activités.

Pour donner l'exemple d'une maturité avancée de la data dans le secteur bancaire, nous rapportons les propos d'un apprenti accueilli dans un groupe bancaire :

« le groupe a mis sur pied un programme de gouvernance et de démocratisation de la data structuré autour de 2 pôles :

- Un pôle gouvernance et démocratisation de la data :

pôle chargé de l'animation et du déploiement de la gouvernance et de la culture data au sein du groupe. La gouvernance met en place un certain nombre de postes d'encadrement comme le CDMO, DMO, DPO, Data owners respectivement Chief data management office, Data management officer et Data protection officer. Le DMO est un pivot de la data au sein d'un établissement, les CDMO est celui qui chapeaute les métiers de la Data dans le groupe, et le DPO est le responsable de la sécurité physique et juridique des données.

- Un pôle usages avancés de la data et IA :

pôle responsable de la coordination des cas d'usage business et du centre d'expertise et d'animation data science. Ce pôle met en avant les technologies de la data science, l'intelligence artificielle et la machine learning. »

3. Catégorisation des données en contexte professionnel

La catégorisation des données conduit souvent à considérer une dimension technique qui distingue les données structurées (tableaux, bases de données comme par exemple celles des CRM-Customer Relationship Management), les données non structurées sans format prédéfini (les commentaires, les données des réseaux sociaux, les données de capteurs...). Une autre dimension relève de la stabilité des données dans le temps en introduisant un qualificatif de température (données froides (vs) données chaudes).

D'autres catégorisations réfèrent au codage (csv, json, xml ...), d'autres aux environnements de collecte (données du web, données d'objets connectés, open data, ...).

Les terrains professionnels des missions analysées permettent d'apporter une contribution à ces efforts de catégorisation. L'ensemble des données traitées dans les missions de ce corpus pourraient être classées en 3 grands types de catégories : les données d'activités des organisations, les données des personnes (employés et clients), les données sur l'environnement de l'organisation concernée (Tableau 3)

Données d'activités de l'organisation
<ul style="list-style-type: none">- Données financières- Données commerciales : achats et ventes- Données de transaction en contexte (interventions, logistique, sinistres, message bancaire, prestations...)- Données de contrôle- Données des délégations, partenaires- Documents d'activités (GED)
Données sur les employés et les clients
<ul style="list-style-type: none">- Données sur le client (CRM, consultations Web, mails, avis...)- Données sur le personnel interne (RH)
Données sur l'environnement (externe)
<ul style="list-style-type: none">- Open data- Données commerciales sectorielles

Tableau 3 : Catégorisation des données traitées en contexte professionnel

Nous n'irons pas plus loin dans cette catégorisation mais des questions intéressantes se posent pour les données d'activités principalement : comparaison du cycle de vie des données d'activités et du cycle de vie des documents d'activités, question de l'apport des pratiques des records managers au traitement des données d'activités élargies (Maday, 2015) ?

4. Finalités poursuivies par les organisations

Nous nous intéressons désormais à préciser les enjeux, les finalités des missions confiées aux apprentis dans le cadre de leur contrat d'apprentissage en tant que *data-analyst*.

Concernant le document, nous pouvons rappeler que trois objectifs génériques majeurs soutiennent le traitement des documents :

- des documents pour construire une mémoire, un patrimoine (dimension historique) (Tricot, Lemarié, Sahut, 2016)
- des documents pour formaliser des savoirs (dimension cognitive) (Callenge, 2015)
- des documents pour soutenir des actions telles que décider, sécuriser, innover (dimension action) (Zacklad, 2005).

Si nous voulons confronter ces cadres au contexte du corpus étudié, le constat est une convergence générale avec les documents pour soutenir les actions de l'entreprise. Le contexte professionnel explique en grande partie cette orientation par rapport à d'autres contextes comme celui de la science ouverte par exemple où l'enjeu de mémoire, de capitalisation des savoirs est aussi prégnant.

La génération de valeur projetée dans les missions des 32 apprentis est centralement celui de la performance (performance opérationnelle, financière, marketing,).

En 2018, Bouafa et Jausaud, explorant 46 projets en entreprises avaient également identifié les finalités suivantes majoritairement inscrites dans cette recherche de performance : amélioration d'études, gain de productivité, aide à la décision, création de services (Bouafia, Jausaud, 2018). Brasseur (2013) nous donne une grille de lecture intéressante pour distinguer trois niveaux de recherche de performance en valorisant leurs données : des indicateurs calculés sur des événements passés (le reporting), des indicateurs en temps réel (le monitoring), des indicateurs pour l'avenir (la prédiction) (cf. Figure 1).

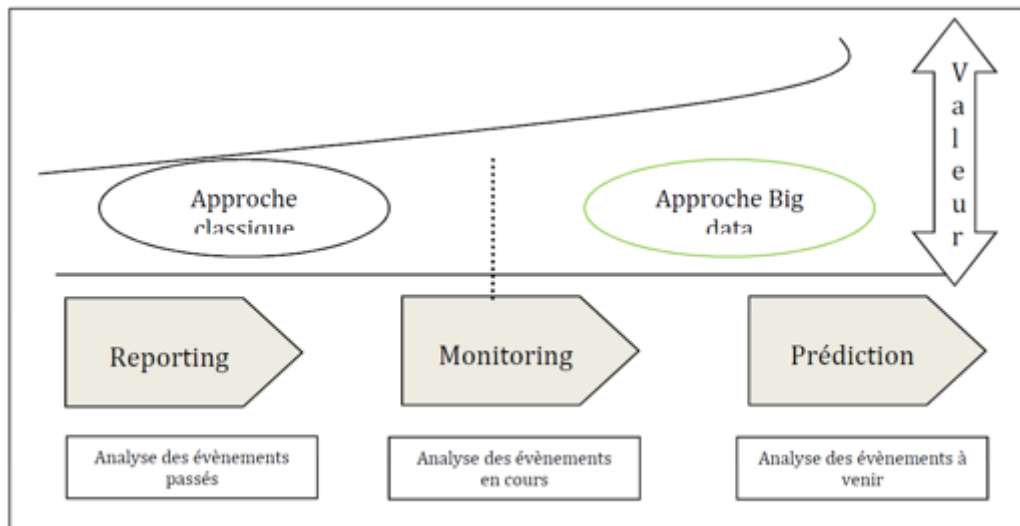


Figure 1 : Evolution de l'analyse des données (Brasseur, 2013)

La fonction de *data-analyst* visée par le master MEDAS désigne majoritairement des missions de tableaux de bord (reporting) associés à des datavisualisations. Cette finalité s'accompagne également de calculs d'indicateurs variés. Le secteur de la « business intelligence » y est associé, prolongeant en quelque sorte les fonctions de veille stratégique et d'intelligence économique qui étaient mises en avant jusqu'à présent.

Pour cette finalité de *reporting*, on peut distinguer dans le corpus traité la mise en avant des objectifs suivants :

- Aide à la décision
- Contrôle des activités de production/ventes
- Evaluation des comportements des usagers, expérience du client
- Optimisation marketing, connaissance client, catégorisation profilée
- Performance des agents
- Automatisation, optimisation de l'analyse des données textuelles (mails, avis)

Aucune mission relevée ne concerne le monitoring avec des analyses de flux en temps réel. L'industrie 4.0 qui intègre cette temporalité n'est, par exemple, pas représentée.

Par contre plusieurs missions intègrent également des objectifs de prévision, associées à des besoins d'anticipation de risques :

- Anticipation des risques de fraude
- Fidélisation des clients, (prévision de l'attrition-client)
- Prévision des attritions des employés
- Prévision pour le pilotage

Les compétences demandées intègrent alors des modélisations statistiques et des processus d'apprentissage sur les données passées. Ce constat met en exergue que les frontières des métiers généralement identifiés pour la data sont poreuses car la modélisation prédictive est plutôt rattachée au métier de *data-scientist* qu'à celui de *data-analyst*.

5. Dispositifs de médiation : activités et outillage correspondant

La compilation des activités et des outils mentionnés dans les mémoires permet de dresser une cartographie de l'environnement de travail d'un *data-analyst* (Tableau 4). Même partielle, cette cartographie éclaire l'offre logicielle en évolution continue dans un secteur très concurrentiel.

Activités	Outils
Extraction des données	SQL, SAS-SQL requêtage Langage R Python(Webscrapping)
Nettoyage, intégration des données	Logstash/Elasticsearch Langage R Excel Pyspark logiciel interne PowerBI
Tableaux de bord et indicateurs, datavisualisation	Package-R Kibana/ElasticSearch Tableau PowerBI Sales force Qlik SAS Excel
Web analyse	Google analytics Big Query Contentsquare Outils d'analyse facebook, youtube Adobe analytics
Analyses statistiques descriptives (classifications) Et text-mining	Python Package-R SPAD Logiciel text mining : Qwam, Text analytics Azure MS
Modélisations statistiques pour la prévision	Python Package-R SPAD
Certification GED qualité, documentation	SAE

Tableau 4 : Activités et outillage associé

Cette cartographie amène plusieurs remarques pouvant éclairer les programmes de formation :

- Les organisations sont outillées différemment selon la maturité de leur stratégie sur les données, différence que l'on pourrait résumer ainsi « entre industrialisation et bricolage »
- Certains logiciels dominent le corpus traité : Tableau, PowerBI, SAS-SQL et les packages R (logiciels libres)
- Logiciels propriétaires et logiciels libres sont présents
- La classification des outils selon les activités s'appuie sur les pratiques des apprentis. Or, l'étendue des fonctionnalités de ces outils peut être plus large et offrir une solution complète pour la « Business Intelligence »
- La web analyse est dominée par des géants du marketing digital
- La proximité avec la gestion des documents est rare (1 seul cas).

Enfin, cette éclairage plus fin de la médiation sur les données en contexte d'entreprises nous donne des précisions sur les tâches réalisées et les outils mobilisés. Par contre, il ne dit rien sur le dialogue avec les métiers, les réajustements, les négociations conduites comme le développe d'autres travaux comme ceux d'A. Nesvijevskaia (2019) ayant proposé un nouveau modèle de gestion des projets data fondé sur le gain de valeur pour les métiers (modèle Brizo_DS).

L'offre croissante des logiciels dédiés repérés contribue à « démocratiser » la *data-analyse*, à simplifier considérablement les compétences nécessaires en statistiques, en programmation et en datavisualisation.

Il n'en reste pas moins qu'une bonne utilisation de ces logiciels nécessite une prise de recul critique qui ne peut s'acquérir qu'en comprenant les statistiques sous-jacentes. Le choix des visualisations prédéfinies nécessite une bonne adéquation avec les messages à transmettre (le sens des données) et cette compétence relève d'une maîtrise de la sémiotique visuelle. Concernant les langages de programmation (R et python), s'ils s'effacent progressivement au profit de fonctionnalités encapsulées dans des logiciels métiers ou dans les bibliothèques de programmation dédiées, au regard des rapports de nos apprentis, ils demeurent toutefois nécessaires pour « bricoler » les données et pour s'avancer sur le terrain de la modélisation statistique à des fins de prédiction.

En conclusion

Ce travail exploratoire devra être amplifié par l'analyse des autres mémoires disponibles (environ 50 concernant le master Medas du Cfa-Paris et celui du Cfa-cnam-Nantes).

Il est une contribution à la compréhension du vaste champ de la médiation de la donnée dans les organisations. Son intérêt réside en grande partie par son terrain : le contexte est majoritairement celui d'entreprises privées. La littérature scientifique intègre peu d'observations dans ce domaine, au contraire des terrains comme l'open data ou les données académiques. Ces investigations permettent aussi un feed-back constructif pour les programmes d'enseignement.

Enfin, il nous paraît important de tracer les lignes de continuum pour les métiers de l'information qui ont développé une maîtrise du traitement du document, de l'information, et

des connaissances et qui doivent désormais se positionner sur le traitement des données dans une société numérique.

Références

Arruabarrena B., Kembellec G., Chartron G. (2019), « Data littératie & SHS : développer des compétences pour l'analyse des données ». *CODATA*, Marne la Vallée, France. [\(hal-03154432\)](#)

Banar-Berger F., Nougaret C. (2014), « Faut-il garder le terme archives ? Des archives aux données », *Gazette des archives*, vol. 233, pp. 7-18.

Bouafia S., Jaussaud J. (2018), « Projets big Data des entreprises : quelles transformations organisationnelles ? ». *Revue COSSI*, N°4-2018 , <https://revue-cossi.info/numeros/n-4-2018-big-data-thick-data/702-1-2018-revue-bouafia-jaussaud>

Brasseur C. (2013), *Enjeux et usages du Big Data : Technologies, méthodes et mise en œuvre* , Hermès, Coll. Management & informatique, 203 pages, ISBN 978-2746245204,

Calenge B. (2015), , Document, information, collection, savoirs, connaissance. Dans : B. Calenge, *Les bibliothèques et la médiation des connaissances* (pp. 19-28). Paris: Éditions du Cercle de la Librairie.

Commission européenne (2020), *Une stratégie européenne pour les données*, <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>

Karoui M., Devauchelle G., Dudezert A. (2014), « Big Data : Mise en perspective et enjeux pour les entreprises », N° Spécial "Big Data", *Revue Ingénierie des Systèmes d'Information*, 19 (3), 73-92 Hermès.

Maday, C. (2015). L'apport de la gestion des documents d'activité (*records management*) à l'ouverture des données: Réflexions basées sur les pratiques en France. *Les Cahiers du numérique*, 11, 149-166. <https://doi.org/>

Mesguich V. (2021), *Rechercher l'information stratégique sur le Websourcing, veille et analyse à l'heure de la révolution numérique*, De Boeck supérieur.

Morelon C. (2020), *La transformation par les données : cas de PME de la région Occitanie. Gestion et management*. Université Paul Valéry - Montpellier III, Thèse sous la dir. de . Florence Noguera, [\(NNT : 2020MON30021\)](#). [\(tel-03132644\)](#)

Nesvijejskaia, A. *Phénomène Big Data en entreprise : processus projet, génération de valeur et Médiation Homme-Données*. Sciences de l'information et de la communication. Conservatoire national des arts et métiers, 2019. Thèse sous la dir. de . Ghislaine Chartron, [\(NNT :](#)

2019CNAM1247. {tel-02970702}

Simonnot B., Broudoux E. , Chartron G., *Humains et données : création, médiation, décision, narration*. Actes du colloque "Document numérique et société", Nancy, octobre 2020. 7ème conférence Document numérique & société (doc-soc), De Boeck Supérieur, 2021, 9782807332638. {hal-03154436}

Tricot A., Lemarié J., Sahut G. (2016), *Le document : communication et mémoire*, De Boeck Supérieur.

Zacklad M. (2005), « Processus de documentarisation dans les Documents pour l'Action (DopA) : statut des annotations et technologies de la coopération associées ». Actes du colloque "*Le numérique : Impact sur le cycle de vie du document pour une analyse interdisciplinaire*", Montréal, Québec., 2005. Éditions de l'ENSSIB , http://archivesic.ccsd.cnrs.fr/sic_00001072/