



Mégadonnées et systèmes d'organisation des connaissances : de l'antagonisme épistémologique à l'interaction productive

Grant Campbell, Alexandre Fortier

► To cite this version:

Grant Campbell, Alexandre Fortier. Mégadonnées et systèmes d'organisation des connaissances : de l'antagonisme épistémologique à l'interaction productive. 12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?, Oct 2019, Montpellier, France. hal-02307930

HAL Id: hal-02307930

<https://hal.archives-ouvertes.fr/hal-02307930>

Submitted on 8 Oct 2019

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.

Mégadonnées et systèmes d'organisation des connaissances : de l'antagonisme épistémologique à l'interaction productive

D. Grant Campbell

Associate professor
University of Western Ontario
gcampbel@uwo.ca

Alexandre Fortier

Bibliothèque du Parlement du Canada
a.z.e.fortier@gmail.com

Résumé

Ce chapitre propose un modèle d'interaction productive entre les systèmes d'organisation des connaissances et les mégadonnées. À partir de principes provenant du domaine de l'acoustique, une analogie est suggérée entre les patrons de vibrations sonores qui servent à produire les sons musicaux et les patrons issus des analyses de mégadonnées qui peuvent servir à enrichir les systèmes d'organisation des connaissances.

Mots clés

Mégadonnées, ontologies, thésaurus, musique, acoustique

Title

Big data and knowledge organization systems: from epistemological antagonism to productive interaction

Abstract

This paper suggests a model of productive interaction for the relationship between professional knowledge organization systems and big data. Drawing on the principles of acoustics, we suggest that the patterns of sound vibrations and harmonics involved in producing musical tones provide a useful analogy for distinguishing patterns of enduring relevance in big data activity: patterns that merit incorporation into knowledge organization systems.

Keywords

Big data, ontologies, thesauri, music, acoustics

INTRODUCTION

Les traces que nous laissons au quotidien dans nos communications, déplacements et transactions, entre autres choses, sont de plus en plus nombreuses et ont profondément transformé la relation qu'entretiennent les individus et les organisations, que celles-ci soient gouvernementales, commerciales ou autres. Il faut désormais tenir pour acquis que nos moindres comportements sont enregistrés et amalgamés à d'autres pour créer de larges profils (Enserink et Chin, 2015). L'ubiquité des mégadonnées a également laissé émerger de nouvelles pratiques de recherche où la collecte et l'analyse reposent sur un paradigme étranger à l'établissement traditionnel des savoirs.

Viktor Mayer-Schönberger et Kenneth Cukier, à qui l'on doit l'un des ouvrages séminaux sur les impacts des mégadonnées, indiquent que la force des mégadonnées réside, paradoxalement, dans le rejet des qualités mêmes qui ont rendu les thésaurus et autres vocabulaires contrôlés si importants dans la recherche d'information. En lieu et place de standards de précision et d'uniformité pour la collecte de données et de leur organisation selon des relations prédéterminées, les partisans des mégadonnées proposent des collectes de données sur une grande échelle, avec une plus grande tolérance pour les erreurs, qui permettent de laisser émerger des patrons plutôt que de chercher à répondre à des questions déterminées au préalable (Mayer-Schönberger et Cukier, 2013).

L'émergence des mégadonnées remet également en question – une fois encore – la viabilité des systèmes d'organisation des connaissances humainement construits. En effet, la relation entre les mégadonnées et les systèmes formels d'organisation des connaissances se rapproche de la tension entre la recherche à l'aide de mots-clés et les vocabulaires contrôlés lors de l'apparition des premières bases de données dans les années 1980 : les thésaurus et les ontologies occupent un espace problématique face aux géants technologiques comme Google (Hjørland, 2016). Avec les mégadonnées, nous assistons non seulement à une résurgence de la fascination pour les données non structurées, mais également à un engouement pour les révélations qui viennent non pas de l'harmonisation et de la désambiguïsation des données, mais de la collecte de jeux de données toujours plus grands permettant des corrélations insoupçonnées (Wang et al., 2016).

Or, comme les possibilités de la recherche par mots-clés ont créé, en retour, une conscience des limites de la recherche par mots-clés, les pouvoirs des mégadonnées permettent également de soulever les problèmes nécessitant des outils plus traditionnels construits par l'humain. Décrites par le Forum économique mondial comme le « nouveau pétrole » (Glanz 2013) – une métaphore souvent reprise –, les mégadonnées représenteraient plutôt le nouveau plutonium pour certains : *« merveilleusement puissantes; dangereuses lorsqu'elles se répandent; difficiles à nettoyer; et avec des conséquences néfastes, lorsqu'utilisées improprement »* (Balsillie, 2019). La relation entre les professionnels de l'organisation et les mégadonnées n'est pas unilatérale. Nous présentons ici trois paradigmes – antagonisme, ignorance et partenariat – permettant de qualifier la nature de la relation.

L'antagonisme épistémologique entre les deux approches, d'abord, puis l'ignorance bienveillante où les deux approches continuent à évoluer isolément. Enfin, en combinant le concept de « classification expansible » de Charles Cutter (1891-1893) avec une analogie dérivée de l'acoustique musicale, un troisième paradigme présente un partenariat potentiellement productif entre les deux approches.

1 – PREMIER PARADIGME : UN ANTAGONISME ÉPISTÉMOLOGIQUE

De prime abord, la relation la plus naturelle, et l'inévitable pour certains, que l'on puisse établir entre les mégadonnées et les systèmes conventionnels d'organisation des connaissances serait l'antagonisme. Depuis Panizzi (1841), les professionnels chargés du contrôle bibliographique se sont appuyés d'une part sur des règles fondées sur des principes et des objectifs précis, et sur des outils standardisés d'autre part. Les professionnels de l'information apprennent que les métadonnées doivent être créées délibérément pour atteindre des objectifs explicites (Svenonius, 2000, p.15). Ils apprennent à établir un équilibre entre l'exhaustivité et la spécificité, et comment leurs décisions en la matière affecteront la précision et le rappel (Cleveland et Cleveland, 1983, p.81). Ils apprennent comment utiliser des vocabulaires contrôlés construits selon des principes validés pour indexer les documents, rendant ainsi possible l'exploitation par l'utilisateur des relations d'équivalence, de hiérarchie et d'association entre les concepts (Shiri, 2012, p.3).

La culture émergente des mégadonnées semble, en surface, démontrer soin, précision et expertise professionnelle. Dans leur apologie articulée et influente pour les mégadonnées Mayer-Schönberger et Cukier (2013, p.32) argumentent que la puissance des mégadonnées vient justement du fait qu'elles s'étendent au-delà de ces procédures établies. Les mégadonnées, disent-ils, n'ont ni la netteté ni la constance des données traditionnelles ; or, la redondance des données collectées rend leur désordre inoffensif. Les algorithmes de recommandation, comme celui de Netflix, par exemple, vont beaucoup plus loin que ne pourraient le faire des métadonnées de genre ou de sujet (Vanderbilt, 2013). Dans les médias sociaux, le profilage psychographique utilise des comportements d'utilisateurs pour déterminer des caractéristiques personnelles, comme l'affiliation politique ou l'orientation sexuelle, qui peuvent aller beaucoup plus loin que les informations soumises par les usagers à propos d'eux-mêmes (Kosinski, Stillwell et Graepel, 2013).

Un tel mépris des standards professionnels rend difficile l'acceptation de ces systèmes par les concepteurs et les utilisateurs de systèmes d'organisation des connaissances, comme les thésaurus et les ontologies. D'un côté, les professionnels de l'information pourraient jouer un rôle important de nettoyage de données pour faciliter la combinaison de jeux de données, par exemple la réconciliation de données structurées, semi-structurées et non structurées (Dzone, 2019, p.6). Ce genre de pratiques soulève d'un autre côté des questions éthiques. Le pouvoir des mégadonnées prend souvent sa source lorsque le désordre est mis en connexion avec des données très précises : des algorithmes qui peuvent par exemple prédire la propagation d'épidémies se basent sur des données géospatiales précises (Ginsberg et al., 2009). Cette interaction entre de grands ensembles de données et des individus spécifiques peut avoir des effets néfastes. Les organisations à la source de publicités ciblées, par exemple, analysent des quantités astronomiques de données brutes pour établir un algorithme, mais doivent se servir des traces laissées par un utilisateur pour être capables de distribuer leurs publicités aux personnes qui correspondent à un profil en particulier (Cameron, 2013). C'est ainsi que Cambridge Analytica a utilisé le pouvoir des mégadonnées pour cibler certains électeurs lors des élections états-uniennes de 2016 (Cadwalladr et Graham-Harrison, 2018). Les mégadonnées peuvent être désordonnées, mais combinées à des données précises, leur pouvoir devient décuplé. Il est dès lors compréhensible que certains professionnels de l'organisation de l'information demeurent circonspects à l'égard de ces développements.

2 - DEUXIÈME PARADIGME : UN IGNORANCE BIENVEILLANTE

La communauté des bibliothèques s'efforce depuis plusieurs années de rendre les catalogues interopérables avec les ressources se trouvant ailleurs sur Internet, notamment les moteurs de recherche, non pas en utilisant des mégadonnées, mais des données liées. Actuellement, la communauté des bibliothèques se prépare à adopter des systèmes d'organisation de l'information basés sur les principes des données liées grâce à l'adoption du Library Reference Model de l'IFLA, un modèle de haut niveau utilise le schéma conceptuel entités-association étendu pour décrire et structurer les métadonnées de l'univers bibliographique. La nouvelle mouture de la norme *Ressources : description et accès*, qui s'appuie sur ce modèle, facilitera l'implémentation des données liées dans les catalogues (Riva, Le Boëuf et Žumer, 2017, p.9).

Contrairement aux mégadonnées, les données liées, qui ont émergé de l'initiative du web sémantique du World Wide Web Consortium, possèdent des caractéristiques plus conventionnelles pour les professionnels de l'information. Les ontologies qui sous-tendent l'organisation à l'aide de données liées partagent plusieurs similarités conceptuelles avec les systèmes d'organisation des connaissances produits par les professionnels de l'information. Les deux communautés utilisent des vocabulaires contrôlés, sous la forme de thésaurus et d'espaces de noms. Les deux présupposent que les métadonnées, qu'elles soient entrées manuellement, moissonnées ou abstraites, sont créées délibérément dans des buts d'identification, de repérage ou d'inférence. En ce sens, les données liées offrent un moyen de migrer les pratiques traditionnelles d'organisation des connaissances dans les environnements interconnectés actuels, environnements qui conservent toutefois les principes inhérents à l'organisation des connaissances de recherche, d'accès et d'utilisation.

Dans ce scénario, les créateurs de thésaurus combinent leur aptitude à organiser des concepts aux demandes que nécessite la création d'ontologies pour des définitions précises qui permettent des inférences sophistiquées. La transition se fait cependant lentement et s'aligne avec l'engagement des bibliothèques envers la sphère publique en proposant une médiation entre les différentes communautés et une représentation des minorités tout en favorisant des sources d'information en accès libre (Widdersheim et Koizumi 2016). Les défenseurs des mégadonnées, en revanche, qui se présentent comme pourvoyeurs d'innovations radicales dans le domaine de l'intelligence artificielle, de systèmes « intelligents » et d'informatique ubiquiste qui s'intègrent rapidement dans nos quotidiens, négligent de mentionner que ces innovations ont un prix. Ainsi, ils omettent souvent d'indiquer les invasions pernicieuses dans la vie privée et les conséquences négatives qu'elles peuvent avoir sur les systèmes politiques et sociaux.

3 – TROISIÈME PARADIGME : UNE INTERACTION PRODUCTIVE

Malgré les questions pratiques et éthiques soulevées par l'invasion des mégadonnées dans de multiples domaines (voir, par exemple, Boyd et Crawford, 2012), les professionnels de l'information s'intéressent de plus en plus aux mégadonnées. Cet intérêt provient en partie de besoins grandissants pour des bibliothécaires de données qui sont capables d'exercer un contrôle bibliographique sur les quantités massives de jeux de données générés par les organisations, qu'elles soient publiques ou privées (Jones et van Selm, 2015, p.20 ; Guindon, 2013). Or, comment pouvons-nous agir sur ces besoins grandissants pour l'organisation et l'utilisation des mégadonnées et quel effet aura cette participation sur la manière avec laquelle nous développons et gérons les outils d'organisation des connaissances ?

Le troisième paradigme – et à nos yeux le plus intéressant – émane d'une réflexion liée à un projet de recherche sur les aspects informationnels des soins aux personnes atteintes de démence. L'analyse des interactions entre les personnes atteintes de démence et ceux qui leur prodiguent des soins révèle l'importance de la musique pour faciliter la communication (Levitin, 2006 ; Sacks, 2007 ; Campbell, 2015). L'extrapolation de ces résultats dans le domaine de l'organisation de l'information nous amène à réfléchir sur l'utilisation d'une analogie des propriétés acoustiques de la musique pour comprendre l'organisation des données.

La musique est reconnue comme un phénomène dense, complexe et puissant dont les effets se font sentir dans plusieurs champs de connaissance. Par exemple, les neuroscientifiques ont étudié les effets de la musique sur le cerveau (Levitin, 2006 ; Sacks, 2007) et les psychologues se sont attardés aux effets de la musique sur la psyché et les émotions (Tan, Pfordreshe & Harré, 2018 ; Powell, 2010 ; Guétin et al., 2009). Plus près des sciences de l'information, des informaticiens ont noté une connexion intellectuelle entre la composition musicale et la programmation informatique (Edwards, 2011).

Notre approche, en revanche, se veut différente. Nous utilisons ici les propriétés acoustiques de la musique, en particulier les propriétés physiques de la production sonore, comme analogie pour comprendre la relation entre les mégadonnées et la création d'outils d'organisation des connaissances, comme les thésaurus et les ontologies. Nous ne suggérons pas ici que les mégadonnées sont une forme de musique, mais simplement que cette analogie donne une clarification utile pour la relation potentielle entre ces deux champs de la gestion des données.

3.1 Son simple et son musical

Du point de vue des acousticiens, il existe une différence fondamentale entre un son simple et un son musical. Un son est à la base une vibration de l'air causée par un stimulus, comme le claquement d'une porte ou la chute d'un objet, qui fait vibrer le tympan. Pour la majorité des sons qui ne sont pas musicaux, cette vibration se dissipe rapidement. Le son musical, en revanche, est un son vibrant à une fréquence spécifique dans la durée. La hauteur du ton est déterminée par sa fréquence. Un orchestre, par exemple, accorde la note la à 440 Hz, alors que le do moyen du piano a une fréquence de 261,6 Hz (Powell, 2010, p. 27).

3.2 Harmonique

La réalisation d'une fréquence stable requiert généralement la production d'une vibration entre une paire de points d'ancrage stable : le mouvement de l'air entre les limites d'une colonne d'air dans le cas d'une flûte ou la vibration d'une corde suspendue entre le chevalet et le sillet au bout du manche pour le violon. Dans un son complexe, comme ceux produits par une corde de violon, la fréquence fondamentale est augmentée par des harmoniques, qui sont des résonances de la fréquence fondamentale (Powell, 2010, p. 30).

Les harmoniques sont nécessairement un multiple entier de la fréquence fondamentale, qui, elle, est à un niveau supérieur (voir Figure 1). Cela donne au son du violon une couleur différente de celui de la flûte qui produit peu d'harmoniques. Si elle est émise à la bonne position sur la corde, la note harmonique – même hauteur, mais à une octave différente – peut être entendue.

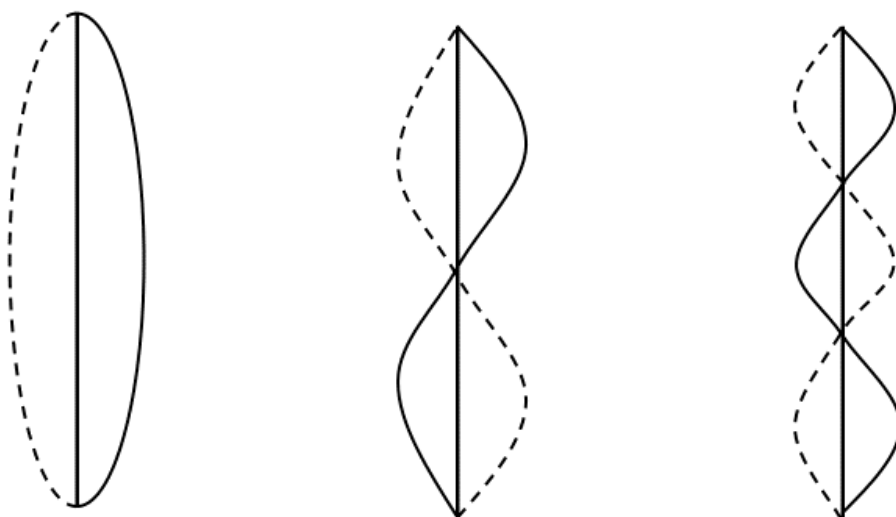


Figure 1 : La fréquence fondamentale et les harmoniques

Le caractère acoustique de la musique, alors, lorsqu'employé comme métaphore, nous donne deux concepts utiles pour comprendre la relation entre les systèmes d'organisation des connaissances et l'analyse des mégadonnées.

- Son musical : résonance consistante à une hauteur spécifique, obtenue à l'aide d'une vibration entre deux points stables ;
- Harmonique : existence de résonance à l'intérieur d'une fréquence fondamentale rendue possible grâce à la stabilité de deux points d'ancrage.

3.3 Musique et système d'organisation des connaissances

Si l'on examine des exemples pertinents de relations entre des concepts dans des vocabulaires contrôlés, on peut détecter des points d'ancrage stables à plusieurs endroits. Par exemple, dans l'entrée pour la démence dans les Medical Subject

Headings, des relations stables sont établies avec un ensemble d'autres concepts : « amentia », « démence génétique » et « démence sénile avec paranoïa », qui sont jugés équivalents (voir Figure 2). De la même manière, si l'on examine la structure hiérarchique, on peut voir que la démence est un terme spécifique sous les maladies du cerveau et un terme générique à la maladie d'Alzheimer et à l'aphasie primaire progressive (voir Figure 3).

Dementia MeSH Descriptor Data 2019

Details Qualifiers MeSH Tree Structures Concepts

MeSH Heading Dementia
Tree Number(s) C10.228.140.380
 F03.615.400
Unique ID D003704
Scope Note An acquired organic mental disorder with loss of intellectual abilities of sufficient severity to interfere with social or occupational functioning. The dysfunction is multifaceted and involves memory, behavior, personality, judgment, attention, spatial relations, language, abstract thought, and other executive functions. The intellectual decline is usually progressive, and initially spares the level of consciousness.
Entry Term(s) Amentia
 Familial Dementia
 Senile Paranoid Dementia
See Also Tauopathies
Public MeSH Note 1981; see DEMENTIA, SENILE 1979-91; see PSYCHOSES, SENILE 1963-78; see PSYCHOSES, PRESENILE 1963-72; see PICK DISEASE OF BRAIN 1981-91; see PICK'S DISEASE OF BRAIN 1966-80; for DEMENTIA, PRESENILE see DEMENTIA 1967-97; for DEMENTIA, SENILE see DEMENTIA 1966-97
History Note 1981(1963)
Date Established 1981/01/01
Date of Entry 1980/05/21
Revision Date 2015/06/30

Figure 2 : L'entrée pour la démence dans les Medical Subject Headings

Dementia MeSH Descriptor Data 2019

Details Qualifiers MeSH Tree Structures Concepts

Nervous System Diseases [C10]
 Central Nervous System Diseases [C10.228]
 Brain Diseases [C10.228.140]
 Acute Febrile Encephalopathy [C10.228.140.021]
 Akinetic Mutism [C10.228.140.042]
 Amblyopia [C10.228.140.055]
 Amnesia, Transient Global [C10.228.140.060]
 Auditory Diseases, Central [C10.228.140.068] +
 Basal Ganglia Diseases [C10.228.140.079] +
 Brain Abscess [C10.228.140.116] +
 Brain Damage, Chronic [C10.228.140.140] +
 Brain Death [C10.228.140.151]
 Brain Diseases, Metabolic [C10.228.140.163] +
 Brain Edema [C10.228.140.187]
 Brain Injuries [C10.228.140.199] +
 Brain Neoplasms [C10.228.140.211] +
 Cerebellar Diseases [C10.228.140.252] +
 Cerebrovascular Disorders [C10.228.140.300] +
Dementia [C10.228.140.380] ●
 AIDS Dementia Complex [C10.228.140.380.070]
 Alzheimer Disease [C10.228.140.380.100]
 Aphasia, Primary Progressive [C10.228.140.380.132] +
 Creutzfeldt-Jakob Syndrome [C10.228.140.380.165]

Figure 3 : La structure hiérarchique pour la démence dans les MeSH

Ces relations, jusqu'à preuve du contraire, sont stables dans le temps, et cette stabilité permet d'affirmer avec une certaine confiance que la démence et la démence génétique représentent le même concept, et que la maladie d'Alzheimer est un type de démence. La stabilité de ces relations provient ici d'un consensus scientifique issu de la recherche et de la pratique médicale, un consensus qui, en retour, permet d'établir des références prédictives qui ressemblent aux vibrations qui créent un spécifique à une fréquence donnée.

En outre, la stabilité de la structure permet de voir que la relation hiérarchique entre la démence, les maladies du cerveau et les maladies du système nerveux central joue un rôle d'harmonique. Dans un vocabulaire contrôlé où les hiérarchies sont

rigoureusement conçues selon les normes, les différents niveaux de spécificité interagissent pour exprimer et renforcer la structure existante, au même titre que les harmoniques à l'intérieur d'une fréquence fondamentale, étant des multiples entiers de la fréquence, contribuent à la richesse et à la complexité de la fréquence fondamentale sans l'ébranler.

Dans d'autres vocabulaires contrôlés, particulièrement ceux qui couvrent des sujets pour lesquels un consensus est absent, les relations sont moins stables. Les vedettes-matière de la Library of Congress, par exemple, contiennent des concepts qui émanent de contextes culturels complexes. Dans ce vocabulaire, la vedette « Prostitutes », par les relations d'équivalence qui renvoient à elle (« call-girls », « female prostitutes », « women prostitutes »), reflète l'idée que seules les femmes se prostituent. Cela peut représenter la caution bibliographique de la Bibliothèque du Congrès ; or, dans ce contexte, la relation hiérarchique entre « prostitutes » et « male prostitutes » indique que tous les prostitués masculins sont des « female prostitutes ». En terme musical, un tel vocabulaire, bâti au fil du temps sur des accumulations de sujets sans structures stables, ressemble à ensemble de sons sans son musical ni harmonique. Les mégadonnées pourraient ici venir pallier ce problème.

Les thésaurus et les ontologies qui tentent de couvrir des domaines hautement subjectifs et avec des particularités culturelles spécifiques pourraient bénéficier de l'apport d'analyses à l'aide de mégadonnées, si cette analyse suit l'analogie de l'acoustique musicale. Pour rester pertinents, les systèmes d'organisation des connaissances doivent se tourner plus que jamais vers une organisation flexible et inspirée d'analyses basées sur les données et non plus fondées sur une logique aristotélicienne (Hjørland, 2015 ; Ibekwe-SanJuan, 2017). En détectant des patrons relationnels stables entre des concepts dans de grands ensembles de données, les algorithmes propres aux mégadonnées pourraient informer les professionnels de l'information pour choisir les tons musicaux au travers du bruit ambiant. Une telle sélection de relations entre les concepts pourrait grandement améliorer les processus de révision de vocabulaires contrôlés dans de multiples domaines de la connaissance.

En retour, comme le mentionne Shiri (2014), les concepteurs de thésaurus et d'ontologies peuvent apporter leurs compétences en support à l'analyse des patrons qui émergent des ensembles de mégadonnées dans plusieurs domaines non conventionnels, comme la détection de maladie, l'analyse prédictive ou la résolution de problèmes complexes, utilisant ainsi leurs compétences d'analyse par facette et de construction de structures hiérarchiques pour détecter des harmoniques dans des ensembles abstraits.

CONCLUSION

Cathy O'Neil (2016) soutient que les algorithmes propres aux mégadonnées doivent être constamment enrichis de nouvelles données pour travailler avec précision et constance. Cet enrichissement leur permet de rester pertinents au fil du temps : « *C'est ainsi que les modèles exacts fonctionnent. Ils font des allers-retours constants avec les phénomènes qu'ils tentent de comprendre ou de prédire. Les conditions changent et les modèles doivent s'adapter* » (O'Neil 2016, p.18, traduction des auteurs).

Ainsi, un algorithme comme celui de Netflix peut donner à ses usagers des recommandations basées sur les tendances actuelles et non sur celles d'un mois qui précède.

Les outils d'organisation des connaissances pourraient jouer un rôle significatif dans cette mise à jour constante. Il y a plus d'un siècle, Charles Cutter a développé son *Expansive Classification System*, qui proposait un système de classification rudimentaire pour les petites collections et permettait le déploiement de structures classificatoires beaucoup plus complexes existant toutes au sein d'une structure plus large. Cependant, une telle interaction ne s'est pas encore avérée (Ibekwe-SanJuan et Bowker, 2017). Même si, comme le souligne à juste titre Hjørland (2013), les outils d'organisations des connaissances bénéficieraient des apports des techniques propres aux mégadonnées.

Les patrons répétés de données, pourrions-nous dire, atteignent dans notre troisième paradigme le statut d'une note de musique : les traces qui se produisent et se reproduisent suffisamment sont des patrons qui valent d'être conservés. De la même manière, les patrons contiennent souvent des sous-patrons, dont la constance avec leur patron crée un effet d'harmonique. La tâche des concepteurs d'outils d'organisation des connaissances n'est pas de créer des solitudes isolées des mégadonnées, mais de capturer, à partir de la multitude d'activités pour lesquelles les mégadonnées sont collectées, des patrons répétés dont la résonance peut inspirer des répétitions à l'intérieur des environnements informationnels.

REFERENCES BIBLIOGRAPHIQUES

BALSILLIE, Jim. (2019). Parlement du Canada. Chambre des Communes. Comité permanent de l'accès à l'information, de la protection des renseignements personnels et de l'éthique. *Témoignage à titre personnel livré au Grand Comité international sur les mégadonnées, la protection des renseignements personnels et la démocratie*. 42^e législature, 1^{re} session, réunion 152 (28 mai).

BOYD, Danah et CRAWFORD, Kate. (2012). Critical questions for big data. *Information, Communication & Society*, 2012, n° 15, p.662-679.

CADWALLADR, Carole. & GRAHAM-HARRISON, Emma. (2018, March 17). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *Guardian*. <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>

CAMERON, Darla. (2013, August 22). How targeted advertising works. *Washington Post*. <https://www.washingtonpost.com/apps/g/page/business/how-targeted-advertising-works/412/?noredirect=on>

CAMPBELL, D. Grant. (2015). Big data and the study of dementia : Epistemological promises and pitfalls. *Proceedings of the American Society for Information Science and Technology*. DOI : <https://doi.org/10.1002/meet.2014.14505101033>

- CLEVELAND, Donald B. & CLEVELAND, Ana D. (1983). *Introduction to indexing and abstracting*. Littleton,CO: Libraries Unlimited, 1983.
- CUTTER, Charles A. (1891–1893) *Expansive Classification: Part I: The First Six Classifications*. Boston, C. A. Cutter.
- DZONE. (2019). *DZone's 2019 guide to big data*. Research Triangle Park, NC: DZone. <https://dzone.com/guides/big-data-volume-variety-and-velocity>
- EDWARDS, Michael. (2011). Algorithmic composition: computational thinking in music. *Communications of the ACM*, 2011, vol. 54, n°7, p.58-67.
- ENSERINK, Martin et CHIN, Gilbert. (2015). The end of privacy. *Science*, 2015, vol. 347, n°6221, p.490–491.
- GINSBERG, Jeremy, MOHEBBI, Matthew H., PATEL, Rajan S., BRAMMER, Lynnette , SMOLINSKI, Mark S., & BRILLIANT, Larry. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 2009, vol. 457, n°19, p.1012-1014.
- GLANZ, James. (2013). Is big data an economic big dud? *New York Times* (18 août). Repéré : <https://www.nytimes.com/2013/08/18/sunday-review/is-big-data-an-economic-big-dud.html>
- GUÉTIN, S., et al. (2009). Effect of music therapy on anxiety and depression in patients with Alzheimer's type dementia: randomised, controlled study. *Dementia and Geriatric Cognitive Disorders*, 2009, vol. 28, n°1, p.36-46.
- GUINDON, Alex. (2013). La gestion des données de recherche en bibliothèque universitaire. *Documentation et bibliothèques*, 2013, vol.59, n°4, p.187–232.
- HJØRLAND, Birger. (2013). Theories of knowledge organization—Theories of knowledge. *Knowledge Organization*, 2013, n°40, p.169–181.
- HJØRLAND, Birger. (2015). Theories are knowledge organization systems (KOS). *Knowledge Organization* , 2015, n°44, p.113–128.
- HJØRLAND, Birger. (2016). Does the traditional thesaurus have a place in modern information retrieval? *Knowledge organization*, 2016, vol. 43, n°3, p.145-159.
- IBEKWE-SANJUAN, Fidelia. et BOWKER, Geoffrey. C. (2017). Implications of big data for knowledge organization. *Knowledge Organization*, 2017, n°44, p.187–198.
- JONES, Sarah et van SELM, Mariëtte. (2015). Training librarians for research data management support. *ALISS Quarterly*, 2015, vol. 10, n°3, p.20-23.
- KOSINSKI, Michael, STILLWELL, David, & GRAEPEL, Thore. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110/15, 5802-5805.
- LEVITIN, Daniel. (2006). *This is your brain on music: the science of a human obsession*. New York: Dutton, 2006, 320p.
- MAYER-SCHÖNBERGER, Viktor. & CUKIER, Kenneth. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt, 2013.
- O'NEIL, Cathy. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. New York: Crown, 2016.
- PANIZZI, Antoni. (1841). Rules for the compilation of the catalogue. *Foundations of cataloging: A sourcebook*. Ed. M. Carpenter & E. Svenonius. Littleton: Libraries Unlimited, 1985, p.1-14.
- POWELL, John. (2010). *How music works: the science and psychology of beautiful sounds*. New York: Little, Brown, 2010, .

- RIVA, Pat., Le BOEUF, Patrick, & ZUMER, Maya. (2017). *IFLA Library Reference Model : a conceptual model for bibliographic information*. As amended and corrected through December 2017. International Federation of Library Associations and Institutions.
- SACKS, Oliver. (2007). *Musicophilia: tales of music and the brain*. New York: Knopf, 2007.
- SHIRI, Ali. (2014). Linked data meets big data: A knowledge organization systems perspective. *Advances in Classification Research Online*, 2014, n°24, p.16–20.
- SHIRI, A. (2012). *Powering search: the role of thesauri in new information environments*. Medford, NJ: Information Today, 2012.
- SVENONIUS, Elaine. (2000). *The intellectual foundation of information organization*. Cambridge: MIT Press, 2000, 274p.
- TAN, Siu-Lan, PFORDRESHE, Peter & HARRÉ, Rom. (2018). *Psychology of music: from sound to significance*. Abingdon: Routledge, 2018.
- VANDERBILT, Tom. (2013, July 8). The science behind the Netflix algorithms that decide what you'll watch next. *Wired*. <https://www.wired.com/2013/08/qq-netflix-algorithm/>
- WANG, Chun, et al. (2016). Statistical methods and computing for big data. *Stat interface*, 2016, vol. 9, n°4, p.399-414.
- WIDDERSHEIM, Michael, & KOIZUMI, Masanori. (2016). Conceptual modeling of the public sphere in public libraries. *Journal of documentation*, vol. 72, n°3, p.591-610.