

ÉDITORIAL

Pour ouvrir ce numéro de RM&I, nous proposons de mettre en débat la question des perspectives méthodologiques de recherche en sciences sociales. A cheval sur deux champs disciplinaires, sciences de gestion, sciences de l'information, il nous a été permis de constater au cours de ces dernières années le renouvellement créatif qu'incarne la vague des « digital methods » (méthodes numériques), adoptées par de nombreuses spécialités même si le mouvement nous semble beaucoup plus discret en sciences de gestion.

Au commencement était Richard Rogers, institué comme le père de ce courant dit aussi d'« Amsterdam »¹. Dans ses ouvrages de référence², il nous livre quelques réflexions qui permettent dans un premier temps de définir et conceptualiser l'objet : les digital methods y sont associées à des méthodes de recherche en sciences sociales utilisant des outils nativement numériques, nécessitant souvent des compétences en codage informatique³, pour analyser des données en ligne ou traces numériques (par exemple les commentaires, les "like", les émoticons, etc.) laissées par les internautes. L'approche est particulièrement recommandée pour les recherches ciblant les « données massives » ou Big data. Un chiffre permet de mesurer l'étendue du champ à explorer : le monde comptait en 2023, 4,7 milliards d'utilisateurs actifs de réseaux sociaux, soit près de 60% de la population de la planète (HootSuite-We Are Social, 2023)⁴.

¹ Par référence à l'université éponyme.

² Rogers, R. (2015). Digital methods for web research. Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource, 1-22, Rogers, R. (2019). Doing digital methods. Sage.

³ Le niveau d'exigence en compétences informatiques varie cependant en fonction de la complexité des problématiques traitées. Ces prérequis expliquent dans la foulée pourquoi le mouvement des digital methods a été marqué par l'arrivée de profils issus à la base des sciences informatiques et qui ont réussi à préempter une partie de l'agenda de recherche des sciences sociales.

⁴ Digital 2023 Global Overview Report de Datareportal; We are social and Hootsuite (janvier 2023). <https://datareportal.com/reports/digital-2023-global-overview-report>

Si cette manne ne peut être ignorée, le débat autour des digital methods appelle à certaines précautions clarificatrices, car il ne s'agit aucunement de s'enfermer dans un enthousiasme passionné ou naïf ni d'occulter certaines réalités. La première est épistémologique. La question invite à s'interroger sur la nature de la connaissance acquise, comparativement aux méthodes conventionnelles qualitatives ou quantitatives, et sur la manière dont les phénomènes en ligne étudiés sont en mesure de traduire les phénomènes sociaux dans le monde réel. Des interrogations qui se justifient d'autant plus que les algorithmes utilisés par ces plateformes pour orienter les flux de données sont loin d'être neutres et peuvent être biaisés en fonction des hypothèses et des préjugés de leurs concepteurs, ce qui peut affecter la validité et la fiabilité des résultats de la recherche. Bien documenté dans la littérature (Pariser, 2011)⁵, l'exemple le plus éloquent est sans doute celui de la « bulle de filtre »⁶, l'un des principaux accusés dans la problématique des fake news.

Les deux autres sont plutôt d'ordre opérationnel. S'agissant de la première, celle-ci renvoie à l'un des défis, et il est récent, auxquels les chercheurs sont confrontés, celui de l'accès, devenu plus limité, aux données. Traditionnellement, les API (acronyme anglais d'« application programming interfaces », un ensemble de protocoles et d'outils qui permettent à des applications informatiques de communiquer entre elles.) ont fourni aux chercheurs un moyen d'accéder gratuitement aux données massives des médias sociaux.

Cependant, deux événements ont montré au cours des cinq dernières les dangers de se fier exclusivement aux APIs, celles-ci pouvant être fermées sans avertissement. Le premier coup de semonce a été porté en avril 2018 par Facebook qui a fermé son API, impactant de manière significative l'extraction de ses données. L'API permettait à partir des pages publiques du réseau social d'extraire tous les messages, commentaires et métadonnées qui leur sont associés. Eclaboussé un mois auparavant par le scandale « Cambridge Analytica »⁷, Facebook se devait, il est vrai, de donner un gage de bonne foi sur la protection de données personnelles auprès de la communauté internationale.

Le deuxième coup d'estocade est plus récent et a concerné un autre réseau social prisé pendant longtemps par les chercheurs, X (anciennement Twitter). La plateforme a longtemps permis, moyennant la création d'un « compte développeur », où le chercheur justifie son projet académique, d'accéder gratuitement aux archives depuis la création du réseau en 2006. Cependant, l'arrivée en 2022 d'un nouveau propriétaire, Elon Musk, a mis fin à l'état de grâce, le magnat décidant de monétiser tout service potentiellement générateur de recettes à ses yeux, de la certification (l'ajout d'un petit badge bleu, autrefois gage d'authenticité, à côté du nom d'utilisateur) jusqu'à l'accès aux données, avec une grille de tarification qui a choqué la communauté académique. Par exemple, le niveau « Basic » de

⁵ Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. penguin UK.

⁶ Ce phénomène désigne les situations d'enfermement informationnels des utilisateurs de réseaux sociaux, les algorithmes ayant épié leurs habitudes de navigation précédentes grâce aux techniques d'analyses prédictives et leur proposant exclusivement un contenu proche de leurs centres d'intérêt.

⁷ Par référence à cette entreprise britannique, soupçonnée d'avoir soutenu Donald Trump pour la présidentielle américaine de 2016 aux Etats-Unis, en exploitant les données de millions d'utilisateurs

l'API est proposé à 100 \$ par mois, mais ne permet aux abonnés de collecter que 10.000 tweets par mois, ce qui est insignifiant par rapport à ce qui pouvait être par le passé collecté gratuitement en une journée. Même le niveau suivant, plus intéressant en termes de volumes de données accessibles, appelés « Entreprise », ne fournissait plus autant de données que la précédente option, dite « Decahose », avec dans la foulée des tarifs qui frôlent l'indécence : 42.000 \$ à 210.000 \$ par mois !

Dans ce qui est communément désigné d'« ère post-API », les chercheurs ont dû trouver des alternatives pour collecter des données, comme le web scraping (l'extraction automatique de contenu à partir de pages web et d'autres fichiers numériques), ou en faisant appel à des fournisseurs tiers de données, tels qu'« Octoparse », « ParseHub » ou « Import.io ». D'autres approches dites « user-centric », c'est-à-dire centrées sur l'utilisateur, consistent à collaborer directement avec les utilisateurs de la plateforme pour leur demander de partager leurs données personnelles avec les chercheurs.

L'autre contrainte concerne l'étape de validation des méthodes de collecte de données. Obtenir l'approbation du comité d'éthique pour des méthodes empiriques à l'instar des digital methods peut être difficile dans certaines institutions universitaires. Ces comités peuvent, à juste titre il faut en convenir, s'interroger sur les enjeux de protection de la vie privée et de sécurité des données et par conséquent hésiter à valider des approches susceptibles d'être perçues comme controversées ou porteuses de risques. D'ailleurs le piratage de données de Facebook dans le cadre du scandale « Cambridge Analytica », a échaudé plus d'un. L'affaire a mis à la fois en évidence la force de frappe ainsi que les enjeux, démocratiques dans le cas d'espèce, des recherches basées sur les API mais en même temps pointé du doigt la mauvaise utilisation des données personnelles collectées par les plateformes en ligne.

Pour conclure, sans aller jusqu'à acter la fin de l'eldorado des digital methods, les réponses aux nouveaux défis sont loin d'être évidentes. Celles-ci supposent avant tout un dialogue ouvert de manière à aligner les attentes entre chercheurs, plateformes de médias sociaux, utilisateurs et universités.

Mohamed Benabid

Université Mohamed 6 Polytechnique