La science ouverte, les données massives et l'intelligence collective en médecine

Marcin Trzmielewski

▶ **To cite this version:**

Marcin Trzmielewski. La science ouverte, les données massives et l'intelligence collective en médecine. 2020. hal-04549494

HAL Id: hal-04549494
https://hal.science/hal-04549494

Submitted on 17 Apr 2024

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.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

La science ouverte, les données massives et l'intelligence collective en médecine

Marcin Trzmielewski
Laboratoire d'Etudes et de Recherches Appliquées en Sciences Sociales (LERASS)
Université Paul Valéry Montpellier 3
Publié le 11 mai 2020 sur le blog d'*AdviceMedica*

Aujourd'hui, face aux pratiques scientifiques traditionnelles, impactées par la marchandisation du savoir et par un manque de reproductibilité et de transparence, nous observons l'émergence du mouvement de la science ouverte. Ce mouvement peut être défini comme un ensemble de normes, de pratiques et d'outils de recherche proposant un modèle alternatif, s'appuyant sur la collaboration et sur la transparence. Elle réunit des sujets comme les données ouvertes (l'ouverture de données publiques), l'open source (la mise à disposition des logiciels avec un code source ouvert), l'open review (la transparence dans le processus de revue par les pairs), l'open access (le libre accès à la littérature scientifique) et les ressources éducatives ouvertes. Ce mouvement est fortement encouragé par les institutions publiques, telles que la Commission européenne, avec son programme Horizon 2020 [1], le Ministère français de l'enseignement, via le Plan national pour la science ouverte [2], et récemment le CNRS, à travers sa Feuille de Route du CNRS pour la science ouverte [3].

Le mouvement de la science ouverte impacte également le champ de l'information de santé. Les éditeurs de revues médicales ouvrent leurs archives d'anciens numéros et mettent en accès libre les numéros courants, en facilitant de cette façon l'accès aux données médicales issue de recherche clinique, nécessaires à l'exercice de la médecine factuelle [4]. A ce mouvement s'ajoute celui de données ouvertes. Selon son idée, les chercheurs qui bénéficient des aides à la recherche financées par l'argent public sont obligés de rendre leurs données disponibles sur le

Web à toute personne qui souhaiterait les utiliser. Il existe des infrastructures numériques pour déposer et échanger les données telles que les banques, les archives, les portails et les entrepôts de données [5] [9]. Les chercheurs en sciences biomédicales publient leurs travaux scientifiques (articles, thèses, rapports de recherche, diapositives, …) dans de grandes archives ouvertes comme HAL et dans les entrepôts institutionnels (ORCA de l'Université de Cardiff, Yale Medicine Thesis Digital Library, …). Ils publient également leurs jeux de données cliniques et d'enquêtes (enregistrées dans des fichiers Excel par exemple) dans les archives de données de la recherche telles que figshare, ZENODO et Dryad. Re3data.org, le registre de ce type d'infrastructures, en répertorie 125 en médecine et en santé publique [10]. L'objectif du dépôt des données de la recherche est de rendre la science plus collaborative et démocratique pour accélérer l'innovation et développer une croissance économique [5] [9].

Les gouvernements ouvrent aussi leurs données. Ils mettent à disposition des institutions scientifiques et privées les registres de morbidité, les entrepôts d'hôpitaux ainsi que les bases de données médico-administratives, les bases de grandes cohortes, d'études longitudinales et d'études cliniques [6]. Les autorités françaises proposent le Portail Epidémiologie France, un catalogue des bases de données de santé [7], alors les Etats-Unis offrent une archive des données HealthData. A ces données s'ajoutent les données produites depuis la transformation numérique du champ de l'information de la santé, initié il y a soixante-ans : les données provenant de bases de données bibliographiques, d'entrepôts d'hôpitaux et de Web 2.0 [8].
En conséquence de ces changements, les données de santé atteignent aujourd'hui des quantités impossibles à saisir par nos capacités cognitives. Le concept de données massives (big data) est très à la mode et désigne les données abondantes et hétérogènes, qui sont produites et échangées rapidement à tout instant [9] et par une variété d'acteurs : des professionnels de santé, par les associations et les mutuelles, aux patients et leurs familles [8]. Plusieurs acteurs veulent s'en

approprier et les exploiter, parmi lesquels les médecins, les chercheurs, les acteurs privés, les hôpitaux. Les enjeux sont multiples : améliorer la prise en charge des maladies, prédire des épidémies, améliorer la pharmacovigilance et mieux gérer les établissements hospitaliers [6].

De plus, la masse de données disponibles et le manque de maîtrise de techniques de recherche documentaire semble être souvent un obstacle pour les professionnels de santé à trouver une information pertinente, fiable, de qualité et dans un temps rapide. Les médecins signalent souvent le besoin des moteurs de recherche fédérés, permettant un accès direct aux ressources informationnelles utiles, à partir de leur poste informatique et de leur smartphone. Ils signalent aussi le besoin des formations à la recherche documentaire [10] [11] [12].

Face à cette masse de données, la question de l'organisation et de la curation des connaissances semble être cruciale pour permettre aux professionnels de santé d'accéder à une information fiable et dans un délai rapide. Pour cela, l'homme et les programmes informatiques effectuent différents processus tels que le catalogage, l'indexation et le marquage social (tagging) pour lesquels ils utilisent différents langages documentaires, comme les règles et formats de description bibliographique, les taxonomies, les folksonomies, les classifications, les thésaurus, les réseaux sémantiques et les ontologies, qui relèvent de l'organisation des connaissances [8] [13].

En outre, pour prendre une décision concernant la prise en charge des patients, les médecins en ville et ceux de centres hospitaliers universitaires demandent souvent l'avis à leurs confrères [4] [12] [14]. A présent, ce type de communication entre les spécialistes est facilitée, grâce aux outils techniques offerts par l'Internet. AdviceMedica est « une solution d'entraide collaborative

et intelligente pour des médecins souhaitant améliorer la résolution de cas difficiles ». Elle regroupe la communauté médicale des professionnels de différentes spécialités et fonctionne sur le principe d'intelligence collective, étant une sorte de collaboration des intelligences et des compétences des praticiens pour trouver la résolution concertée. L'un des exemples de communautés d'usage de cet outil est celle des allergologues français, qui l'utilisent d'ores et déjà dans leur pratique quotidienne [15]. AdviceMedica s'inscrit parfaitement dans le contexte de la science ouverte, en tant que modèle alternatif d'échange des savoirs médicaux. Il s'appuie sur la collaboration et sur l'intelligence collective, basées sur la confiance et sur la confraternité.

Sources

[1] HORIZON 2020 – Le portail français du programme européen pour la recherche et l'innovation. Publication d'un guide pour la gestion des données de recherche dans H2020 par la C.E. 2016.

[2] Ministère de l'Enseignement supérieur, de la recherche et de l'innovation. Science ouverte – Le plan national pour la science ouverte : les résultats de la recherche scientifique ouverts à tous, sans entrave, sans délai, sans paiement. 2019.

[3] CNRS. Science ouverte – Feuille de route du CNRS (18 novembre 2019). 2019.

[4] Gonod-Boissin, Florence. L'usage de l'information numérique en médecine générale : étude exploratoire en Rhône-Alpes. Thèse de doctorat, Université Claude Bernard Lyon 1, 2007.

[5] Trzmielewski, Marcin, Claudio Gnoli. Une classification interdisciplinaire pour l'échange et la médiation des données ouvertes de la recherche. 12ème Colloque international d'ISKO-

France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?

[6] INSERM. Big data en santé – Des défis techniques, humains et éthiques et relever. 2016.

[7] Portail Epidémiologie France.

[8] Trzmielewski, Marcin, Claudio Gnoli. Organiser les connaissances médicales : approche théoriques et techniques en information-documentation : Pratiques d'information et connaissances en santé.

[9] Leonelli, Sabina. La recherche scientifique à l'ère des Big Data. Cinq façons dont les Big Data nuisent à la science et comment la sauver. Sesto S. Giovanni : Editions MIMESIS, 2019.

[10] Léon, Elodie. Les pratiques de recherche documentaire des médecins généralistes : les freins et les difficultés pour l'accès à une information de qualité. Thèse pour le diplôme d'état de docteur en médecine. Université de Bordeaux, 2014.

[11] Dang, Suzanne. Utilisation des applications médicales smartphone par les médecins généralistes de Haute-Normandie en 2016.

[12] Letang, Tiphaine, Fabien Espitia. Recherche d'informations médicales actualisées et validées sur internet : pratiques, difficultés et attentes des médecins généralistes français.

[13] Hjørland Birger. Knowledge organization. Knowledge Organization, vol. 43, n° 6, 2016, p. 475-484.

[14] Tahamtan, Iman, Harnid Reza Baradaran, Askar S. Afshar. Drug information seeking behaviours of heath care professionals in Iran. New Library World, vol. 116, n° 3/4, 2015, p. 173-186.

[15] Demoly M, Bourrain JL, Demoly P. La complexité médicale appliquée à l'allergologie : causes et solutions. Rev Fr Allergol 2019; 59: 336-40.