

Les données massives et le Réseau mondial d'information en santé publique (RMISP)

Dion M¹, AbdelMalik P², Mawudeku A^{2*}

¹ Schulich School of Family Medicine and Dentistry, Université de Western Ontario, London (Ontario)

² Centre de mesures et d'interventions d'urgence, Agence de la santé publique du Canada, Ottawa (Ontario)

* Correspondance : Abla.Mawudeku@phac-aspc.gc.ca

Résumé

Contexte : La mondialisation et le risque de propagation rapide des maladies infectieuses émergentes ont rendu de plus en plus nécessaires la surveillance continue et la détection précoce de ces maladies. Le Réseau mondial d'information en santé publique (RMISP) a été établi pour améliorer la connaissance de la situation et accroître la capacité de détection précoce des événements de santé publique émergents.

Objectif : Décrire comment le RMISP utilise les données massives comme technique efficace de détection précoce des éclosions de maladies infectieuses et déterminer les orientations futures possibles pour le RMISP.

Constatations : Chaque jour, le RMISP analyse plus de 20 000 nouvelles en ligne (provenant de plus de 30 000 sources) en neuf langues dans le monde entier. Un programme sur Internet compile les données à partir d'un algorithme qui repère les signaux potentiels d'événements de santé publique émergents, lesquels sont ensuite examinés par une équipe multilingue et pluridisciplinaire. Une alerte est envoyée si un risque potentiel est détecté. Ce processus a fait ses preuves durant l'éclosion du syndrome respiratoire aigu sévère (SRAS) et a été adopté peu après par un certain nombre de pays afin de satisfaire au nouveau Règlement sanitaire international (RSI), qui exige que chaque pays soit capable de détecter et de signaler de façon précoce les maladies. Le RMISP a repéré l'éclosion précoce de SRAS en Chine, s'est vu attribuer le mérite de la première alerte du coronavirus du syndrome respiratoire du Moyen-Orient (CoV-SRMO), et a joué un rôle important dans le suivi de l'éclosion du virus Ebola en Afrique occidentale. De futures évolutions sont envisagées pour faire progresser la capacité du RMISP à tirer parti d'autres sources de données massives, telles que les médias sociaux, ainsi que sa capacité analytique en matière d'élaboration d'algorithmes.

Conclusion : L'adoption précoce des données massives par le RMISP a permis d'accroître la capacité mondiale à détecter les éclosions internationales de maladies infectieuses et les autres événements de santé publique. L'intégration de sources supplémentaires de données massives et les progrès en matière de capacité analytique pourraient renforcer encore davantage la capacité du RMISP à détecter les maladies et à lancer l'alerte rapidement.

Introduction

Avec l'avancée de la mondialisation, les maladies transmissibles et les événements de santé publique émergents se propagent de plus en plus rapidement. Par conséquent, leur surveillance continue et leur détection précoce sont encore plus importantes pour prévenir ou atténuer la propagation internationale des maladies infectieuses et pour laisser aux pays le temps nécessaire pour s'y préparer et y faire face. Les données massives désignent des ensembles de données extrêmement volumineux provenant de sources telles que les médias sociaux ou les journaux, qui nécessitent des méthodes de calcul puissantes pour révéler des tendances, des schémas ou des prédictions de probabilité d'un événement (1,2). Les données massives sont utilisées pour optimiser les processus de vente et d'affaires, orienter les échanges de joueurs entre les équipes sportives, ou encore améliorer l'urbanisme. Elles deviennent rapidement essentielles à

divers aspects de la santé, allant de l'administration des soins de santé à l'application Suivi de la grippe de Google et à la pharmacosurveillance (3).

Le Canada a été l'un des premiers à adopter les données massives pour le repérage initial des infections émergentes, et ce, dès 1997 avec la mise en place du Réseau mondial d'information en santé publique (RMISP), fruit d'un effort de coopération entre (à l'époque) Santé Canada et l'Organisation mondiale de la Santé (OMS) (4,5). Le RMISP continue à être géré par l'Agence de la santé publique du Canada (l'Agence) et rassemble un réseau mondial de professionnels de la santé publique et d'organisations (p. ex. ministres de la Santé) dans une optique de connaissance de la situation et de détection précoce des événements de santé publique émergents. Le RMISP s'appuie sur un système automatique sur Internet qui parcourt les journaux et les autres communications dans le monde entier pour repérer les indicateurs potentiels (ou « signaux ») d'éclosions, qui sont ensuite analysés et rapidement évalués par une équipe multilingue et pluridisciplinaire à l'Agence. Lorsqu'un risque est détecté, les analystes diffusent les renseignements pertinents et alertent les hauts fonctionnaires et les intervenants pour que ceux-ci prennent une décision. Bien qu'il ait été initialement conçu pour repérer les éclosions de maladies transmissibles, le système est également utilisé pour surveiller les dangers chimiques et radionucléaires potentiels (4,6).

Le présent article a pour objectif de décrire le fonctionnement du RMISP dans le contexte des données massives, de l'illustrer par des exemples récents du RMISP en action et d'explorer les possibles orientations futures.

Le RMISP et les données massives

Les données massives sont définies par trois V : volume, vitesse et variété (7,8). Le volume décrit la quantité de données recueillies, la vitesse correspond à la fréquence à laquelle les données sont recueillies et diffusées, et la variété désigne la multiplicité des sources utilisées pour compiler les données (7).

Le volume et la variété du RMISP sont illustrés par l'utilisation de fonctions de recherche et d'agrégateurs de nouvelles (sociétés donnant accès à des milliers de sources d'actualité dont le contenu est automatiquement indexé) qui rassemblent de grandes quantités d'ensembles de données provenant de multiples sources différentes. Une application sur Internet du système du RMISP balaye et analyse en continu les sources de nouvelles acquises en neuf langues (arabe, anglais, persan, français, portugais, russe, chinois simplifié, espagnol et chinois traditionnel) (4). La quantité de données générées dépend des critères, des variables et des algorithmes définis pour les agrégateurs (6). Ces algorithmes repèrent les signaux potentiels d'événements de santé publique émergents et filtrent les données non pertinentes considérées comme du « bruit » (**Figure 1**) (7). Chaque jour, en moyenne, le RMISP traite 3 000 nouvelles (9). Le volume augmente lorsque les sources d'actualité élargissent leur couverture des événements de santé publique émergents, comme dans le cas de la récente éclosion du virus Ebola en Afrique occidentale.

Le RMISP en action

Détection précoce

Le RMISP s'est avéré une ressource efficace pour la détection précoce des éclosons de maladies infectieuses. Son utilité a été initialement démontrée pendant l'écllosion du syndrome respiratoire aigu sévère (SRAS) en 2003, lorsque les premières alertes ont été lancées dans les actualités des journaux chinois. La première nouvelle en anglais concernant une écloson atypique en Chine a été mentionnée par une société pharmaceutique dans la rubrique financière d'un journal qui avait fait état d'une hausse des ventes de médicaments antiviraux (11). Cela a permis non seulement de signaler l'émergence de l'écllosion, mais également de fournir des renseignements supplémentaires sur l'utilisation locale des antiviraux pour contenir la propagation du virus.

À la suite de l'écllosion de SRAS, l'importance de l'utilisation des médias d'actualité en complément des systèmes nationaux de surveillance de santé publique plus traditionnels a été reconnue par l'OMS et ses États membres (12,13). L'écllosion de SRAS a abouti à des révisions du Règlement sanitaire international (RSI) (14) exigeant que les pays signalent et contrôlent les éclosons potentiellement préoccupantes au niveau international afin de renforcer la sécurité mondiale en matière de santé publique. Le Règlement note que l'OMS peut tenir compte de rapports émanant de sources autres que les notifications et les consultations officielles dans son évaluation d'un événement de santé publique potentiel émergent (14). Après l'écllosion de SRAS, de nombreux pays ont adopté les résultats du RMISP dans le but d'élargir leur capacité de surveillance (4,15).

Au fil des ans, le RMISP a continué à détecter les signaux précoces d'éclosons préoccupantes à l'échelle internationale, comme la grippe pandémique H1N1 en 2009 (16). Les premiers rapports en espagnol à ce sujet faisaient état d'une écloson d'une maladie respiratoire inhabituelle dans l'État de Veracruz, au Mexique, qui avait provoqué deux décès.

En avril 2012, le RMISP a repéré huit cas d'une maladie respiratoire inconnue et un décès en Jordanie. Le RMISP a envoyé une alerte pour informer ses partenaires, incluant l'OMS, de ces cas. Suite à une enquête plus poussée et les résultats d'une analyse rétrospective de laboratoire, une écloson du coronavirus du syndrome respiratoire du Moyen-Orient (désormais appelé CoV-SRMO) fut confirmée. Un avis du RSI fut publié en novembre 2012. Le RMISP s'est vu attribuer le mérite d'avoir été le premier à envoyer une alerte concernant ce nouveau virus émergent.

Surveillance en continu

Le RMISP s'est avéré utile aussi bien pour la détection précoce que pour la surveillance en continu. La surveillance en continu des événements est cruciale pour bien connaître la situation, en ce qui concerne l'évolution d'une écloson et les stratégies d'intervention et d'atténuation mises en œuvre par les communautés locales, nationales et internationales. Parmi les exemples de connaissance de la situation en matière de stratégies d'atténuation, on peut citer la capacité du RMISP à repérer les annulations de vols ou de croisières, les nouveaux avis aux voyageurs, les procédures de dépistage sanitaire aux frontières ou les interdictions commerciales. Ce processus a été beaucoup plus efficace que de communiquer individuellement avec les sociétés de transport commercial, les agences de voyage et les aéroports.

Par exemple, pendant la grippe pandémique H1N1, le RMISP a été utilisé comme source de renseignements par l'Organisation mondiale du commerce afin de surveiller l'ampleur et l'effet des interdictions commerciales (17). De même, durant l'intervention récente contre l'écllosion du virus Ebola en Afrique occidentale, le RMISP a permis de connaître la situation sur le plan des annulations de vol, des avis aux voyageurs et des procédures de dépistage sanitaire aux frontières.

Prochaines étapes

Nouvelles sources de données potentielles

Internet, le courrier électronique, les téléphones intelligents et les médias sociaux se sont développés rapidement depuis la création du RMISP. Par conséquent, de nouvelles sources potentielles de données massives ont émergé qui peuvent être analysées pour détecter les signaux précoces d'éclosions de maladies infectieuses. Les médias sociaux (tels que Twitter et Facebook) ont enregistré une croissance exponentielle au cours des dix dernières années, et ces plateformes créent d'énormes quantités de contenu et de données générés par les utilisateurs (18).

Ces divers médias sociaux constituent de nouvelles sources de données potentielles pour le RMISP. De plus, d'autres organisations ont commencé à exploiter les ressources des médias sociaux pour améliorer la surveillance des maladies (18). Par exemple, l'application Suivi de la grippe de Google surveille les comportements de recherche en ligne pour détecter des signes précurseurs de la grippe (19), des chercheurs ont utilisé Facebook pour contribuer à prévoir les résultats en matière de santé au niveau de la population locale (20), Twitter a été utilisé comme une vaste source de données pour surveiller les tendances de la santé pendant une éclipse de grippe aviaire (21), et les téléphones cellulaires ont servi à mesurer les schémas de mobilité humaine dans le contexte de la transmission du paludisme dans les pays en développement (22).

Les médias sociaux ont contribué à l'amélioration des interventions d'urgence en permettant de saisir des données en temps réel sur la santé des communautés (23) et la réaction du public à un événement (24). Par exemple, l'utilisation de téléphones intelligents et de Twitter au Nigeria pendant l'éclosion d'Ebola en Afrique occidentale a aidé à repérer une éclipse dans une nouvelle région trois jours avant qu'elle soit annoncée par l'OMS (25).

Au nombre des autres nouvelles applications figurent les systèmes d'externalisation à grande échelle, qui saisissent les symptômes soumis de façon volontaire par le grand public à l'aide d'Internet ou des réseaux de téléphone cellulaire et qui les compilent rapidement et fournissent un retour sur ces données en temps quasi réel. Ces systèmes ont été utilisés par des applications participatives de surveillance des maladies infectieuses telles que Flu Near You (26) et DoctorMe (27).

Cependant, l'utilisation des médias sociaux comme sources de données pose quelques problèmes inhérents. L'une des principales difficultés associées aux données massives en général et au contenu des médias sociaux en particulier est le « rapport signal/bruit » qui peut augmenter considérablement le nombre potentiel de faux positifs et de faux négatifs. Avec l'abondance de discussions et de gazouillis autour de l'éclosion du virus Ebola en Afrique occidentale, par exemple, il était difficile de faire la distinction entre les véritables signaux préoccupants et la pléthore de messages à laquelle on pourrait autrement s'attendre pendant un tel événement. En outre, certains médias sociaux, notamment les gazouillis qui sont limités à 140 caractères, peuvent ne pas fournir assez d'information contextuelle pour permettre de repérer un signal fiable (28).

Une autre difficulté associée à l'utilisation des médias sociaux est la représentativité. Tout le monde n'a pas accès à un téléphone intelligent et, par conséquent, les données des plateformes de médias sociaux ne peuvent refléter que la partie de la population qui en utilise un (28). La technologie mobile se répand considérablement, ce qui pourrait contribuer à résoudre les préoccupations relatives à la représentativité (29).

Enfin, l'utilisation des médias sociaux soulève des considérations éthiques associées aux droits des individus, notamment des problèmes liés à la protection des renseignements personnels (2).

Amélioration de l'analyse des données

Le RMISP pourrait non seulement élargir ses sources de données, mais également faire progresser ses capacités d'analyse des données. Des méthodes avancées de calcul et de vérification pour améliorer la sensibilité et la spécificité des signaux détectés sont actuellement envisagées (30). La question de savoir si

un meilleur traitement des données pourrait réduire la nécessité d'avoir recours à une équipe multilingue et pluridisciplinaire est également à l'étude. Le RMISP évalue et perfectionne en permanence les agrégateurs et les algorithmes utilisés, ce qui pourrait potentiellement aboutir à des formes plus avancées d'intelligence artificielle. La poursuite de l'amélioration de la capacité analytique du RMISP permettra la gestion, l'intégration, l'analyse et l'interprétation solides de volumes de données de plus en plus importants et complexes (31).

Conclusion

Le Réseau mondial d'information en santé publique du Canada a été parmi les premiers à adopter les données massives et, en tant que ressource mondiale en continu, il aide les pays à satisfaire les exigences en matière de capacité de surveillance des événements pour la détection précoce et le signalement des éclosions de maladies infectieuses et des autres événements préoccupants à l'échelle internationale. Les progrès continus des données massives, notamment l'utilisation des médias sociaux et des téléphones intelligents, ainsi que les améliorations de la capacité d'analyse constituent des occasions de renforcer encore davantage le RMISP. Dans l'ensemble, les approches fondées sur les données massives sont devenues une composante essentielle des efforts en matière de santé publique à l'échelle locale, nationale et internationale pour détecter, signaler et contrôler les éclosions émergentes.

Remerciements

Nous souhaitons remercier toute l'équipe du RMISP. Le Réseau doit son succès à leur travail assidu, à leur dévouement et à leur constante collaboration. Leur soutien et leurs conseils ont été grandement appréciés et ont fortement contribué à la rédaction de cet article.

Conflit d'intérêts

Aucun.

Financement

Le RMISP est financé par l'Agence de la santé publique du Canada.

Références

- (1) George G, Haas MR, Pentland A. Big Data and management. *Acad Manag J*. 2014;57(2):321-6.
- (2) Vayena E, Salathé M, Madoff LC, Brownstein JS, Bourne PE. Ethical challenges of Big Data in public health. *PLoS Comput Biol*. 2015;11(2):e1003904.
- (3) Muchaal P, Meganath K, Landry L, Aramini J. Evaluation of a national pharmacy-based syndromic surveillance system. *Can Comm Dis Rep*. 2015 41;9:204-210.
- (4) Keller M, Blench M, Tolentino H, Freifeld CC, Mandl KD, Mawudeku A, et al. Use of unstructured event-based reports for global infectious disease surveillance. *Emerg Infect Dis*. 2009 May;15(5):689-95.
- (5) World Health Organization. [Internet] Epidemic intelligence - Systematic event detection. Geneva: World Health Organization; 2015. Available from: <http://www.who.int/csr/alertresponse/epidemicintelligence/en/>.
- (6) Mykhalovskiy E, Weir L. The Global Public Health Intelligence Network and early warning outbreak detection: A Canadian contribution to global public health. *Can J Public Health*. 2006 Jan-Feb;97(1):42-4.
- (7) McAfee A, Brynjolfsson E, Davenport TH, Patil D, Barton D. Big Data. The management revolution. *Harvard Bus Rev*. 2012 Oct;90(10):61-7.
- (8) Hay SI, George DB, Moyes CL, Brownstein JS. Big Data opportunities for global infectious disease surveillance. *PLoS Med*. 2013;10(4):e1001413.
- (9) Mawudeku A, Blench M, Boily L, St John R, Andraghetti R, Ruben M. The Global Public Health Intelligence Network. In: *Infectious Disease Surveillance, Second Edition*. New York: John Wiley and Sons; 2013. pp. 457-69.
- (10) Heymann DL, Rodier G. Global surveillance, national surveillance and SARS. *Emerg Infect Dis*. 2004;10(2):173-5.

- (11) Mawudeku A, Blench M. Global Public Health Intelligence Network (GPHIN) In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: August 2006. Cambridge, MA, USA.
- (12) Davies SE. Nowhere to hide: Informal disease surveillance networks tracing state behaviour. *Global Change, Peace & Security*. 2012;24(1):95-107.
- (13) Davies SE, Youde JR. The Politics of surveillance and response to disease outbreaks: The new frontier for states and non-state actors. Burlington VT: Ashgate Publishing, Ltd.; 2015.
- (14) World Health Organization. International Health Regulations. 2008. Second Edition. Geneva: WHO; 2008.
- (15) Baker MG, Fidler DP. Global public health surveillance under new international health regulations. *Emerg Infect Dis*. 2006 Jul;12(7):1058-65.
- (16) Warren AP, Bell M, Budd L. Surveillance networks and spaces of governance: Technological openness and international cooperation during the 2009 H1N1 pandemic. Washington: Association of American Geographers; 2010.
- (17) Lamy P. Report to the TPRB from the Director-General on the financial and economic crisis and trade-related development. Geneva: World Trade Organization; 2009.
- (18) Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. *J Med Internet Res*. 2013 Jul 18;15(7):e147.
- (19) Davidson MW, Haim DA, Radin JM. Using networks to combine Big Data and traditional surveillance to improve influenza predictions. *Sci Rep*. 2015;5:8154.
- (20) Gittelman S, Lange V, Gotway Crawford CA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. *J Med Internet Res*. 2015 Apr 20;17(4):e98.
- (21) Fung IC, Wong K. Efficient use of social media during the avian influenza A (H7N9) emergency response. *Western Pac Surveil Response J*. 2013;4(4):1.
- (22) Buckee CO, Wesolowski A, Eagle NN, Hansen E, Snow RW. Mobile phones and malaria: Modeling human and parasite travel. *Travel Med Infect Dis*. 2013;11(1):15-22.
- (23) Merchant R, Elmer S, Lurie, N. Integrating social media into emergency-preparedness efforts. *N Engl J Med*. 2011 July 28;365(4):289-91.
- (24) Fung IC, Fu K, Ying Y, Schaible B, Hao Y, Chan C, et al. Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks. *Infect Dis Poverty*. 2013;2(1):1-12.
- (25) Odlum M, Yoon S. What can we learn about the Ebola outbreak from tweets? *Am J Infect Control*. 2015;43(6):563-71.
- (26) Wojcik OP, Brownstein JS, Chunara R, Johansson MA. Public health for the people: Participatory infectious disease surveillance in the digital age. *Emerg Themes Epidemiol*. 2014 Jun20;11:7,7622-11-7. eCollection 2014.
- (27) Susumpow P, Pansuwan P, Sajda N, Crawley AW. Participatory disease detection through digital volunteerism: How the DoctorMe application aims to capture data for faster disease detection in Thailand. Proceedings of the companion publication of the 23rd International Conference on World Wide Web Companion; International World Wide Web Conferences Steering Committee; 2014.
- (28) Guy S, Ratzki-Leewing A, Bahati R, Gwady-Sridhar F. Social media: A systematic review to understand the evidence and application in infodemiology. In: *Electronic Healthcare*. New York: Springer; 2012. p. 1-8.
- (29) e Geneva: International Telecommunications Union; 2013. Available from: <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013-e.pdf>.
- (30) Kostkova P. A roadmap to integrated digital public health surveillance: The vision and the challenges. Proceedings of the 22nd International Conference on World Wide Web Companion; International World Wide Web Conferences Steering Committee; 2013.
- (31) Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and internet-based data in global systems for public health surveillance: A systematic review. *Milbank Q*. 2014;92(1):7-33.