

# Health surveillance using the internet and other sources of information

doi:10.1136/amiajnl-2013-001812

Lucila Ohno-Machado, *Editor-in-chief*

The rapid accumulation of large amounts of data from Internet searches, social network sites, and self-monitoring devices, in addition to the widespread adoption of electronic health records (EHR) in the past few years have increased opportunities as well as challenges in effectively integrating and using these data. *JAMIA* has recently issued requests for papers (RFPs) for two special issues: one is dedicated to EHR Phenotype Extraction, and another one is dedicated to Big Data. The first RFP addresses the problem that, even though they are electronic, the data in EHRs are not always easy to structure and to harmonise across different institutions. Scalable and reliable approaches to extract meaningful phenotypes that can be integrated with genetic and environmental data are urgently needed, and we expect to publish the best work in this area in this special issue. The second RFP relates to 'meaningful use' of big health-related data. We will feature the most innovative approaches for efficient storage, pre-processing, analysis, and sharing of data from 'omics', EHRs, imaging, and Internet sources of information.

In this issue of *JAMIA*, we also present innovative work that is related to these topics. These articles will intrigue and motivate readers to apply existing algorithms and tools to new problems, as well as to develop new solutions to derive knowledge from data originating from very heterogeneous sources. White (*see page 404*) uses data originated from a log of web searches, Harpaz (*see page 413*) and Xu (*see page 420*) use data from EHRs, and Avillach (*see page 446*) uses the literature to detect and validate

adverse drug events. El Eman (*see page 453*) proposes a method to combine distributed sources to detect rare adverse drug events while preserving privacy, which is a topic also covered by Mohammed (*see page 462*). Articles from Phansalkar (*see page 489*) and Duke (*see page 494*) cover drug-drug interaction alert systems, while Olsho (*see page 470*) and Galanter (*see page 477*) report on the impact of computerised provider order entry (CPOE) systems in the rate of medication errors. Also relating to medications, Fung (*see page 482*) and McDonald (*see page 499*) describe systems to extract information from narrative text in drug labels and to automate the medication regimen complexity index, respectively.

The Internet has been used as a vehicle to collect information, but also to deliver interventions. Samwald (*see page 409*) describes a prototype system for patients to carry their pharmacogenomics information in a way that is easily interpretable by clinicians. Patient empowerment is also addressed by Turner-McGrievy (*see page 513*), who compares traditional versus mobile app self-monitoring of physical activity and dietary intake. Osborn (*see page 519*) describes patient experiences using web portals and secure messaging for diabetes management, and Tang (*see page 526*) reports on the impact of engaging diabetic patients in online disease management. Mathieu (*see page 568*) reviews the quality of Internet-based randomised controlled trials, and Silverstein (*see page 535*) describes an innovative web-based stereoscopic visualisation system that helps clinicians inter-actively collaborate at various sites.

Today, EHRs constitute a major source of data for biosurveillance. Maslove (*see page 427*) uses data from EHRs to monitor acquired infections using social network techniques, and Cheng (*see page 435*) uses these type of data to detect outbreaks using real-time structural models. Surveillance techniques are used by Mahajan (*see page 441*) to identify cases of Hepatitis B and by Ong (*see page 506*) to monitor health information system failures. Clinical data warehouses derived from EHRs are the primary sources of this information. Hruby (*see page 563*) reports on a centralised research data repository for outcomes research, while Tao (*see page 554*) describes the experience of using the Clinical Element Model to represent concepts for secondary use of EHR data. Maslove (*see page 544*) describes a method for discretization of continuous features in clinical data to facilitate the use of machine learning applications, among other purposes. Finally, a review by Dixon (*see page 577*) summarises the literature on communication of data from public health systems to clinicians via EHRs, and a review from Edworthy (*see page 584*) covers the topic of best practices for the design of medical audible alarm systems.

Healthcare and biomedical research are collecting data at a fast pace. Added to the relatively new sources of data originating from Internet use, this constitutes an ideal time to propose disruptive approaches for extracting knowledge from data. We expect to see a lot of highly innovative, and possibly unconventional, submissions to our special focus as well as to our regular issues in the upcoming months.