ID News Briefs: Big Data

Ioannidis JPA, Khoury MJ. **Big data meets public health** Science 28 November 2014: Vol. 346 no. 6213 pp. 1054-1055. http://www.sciencemag.org/content/346/6213/1054.full?sid=775d7875-0e89-4618-bd28-04fb6f240c6e

The term "Big Data" refers to volumes of large, complex, linkable information. Beyond genomics and other "omic" fields, Big Data includes medical, environmental, financial, geographic, and social media information. Most of this digital information was unavailable a decade ago. This swell of data will continue to grow, stoked by sources that are currently unimaginable. Big Data stands to improve health by providing insights into the causes and outcomes of disease, better drug targets for precision medicine, and enhanced disease prediction and prevention. Moreover, citizen-scientists will increasingly use this information to promote their own health and wellness. Big Data can improve our understanding of health behaviors … and accelerate the knowledge-to-diffusion cycle… But the promise of Big Data is also accompanied by claims that "the scientific method itself is becoming obsolete" (and) "Big Error" can plague Big Data.

Pigott DM, Kraemer MU. **Enhancing infectious disease mapping with open access resources**. Euro Surveill. 2014; 19(49): pii=20989. Available online: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20989

The Spatial Ecology and Epidemiology Group (SEEG) at the University of Oxford … has collated a number of globally comprehensive and up-to-date databases from… three sources: (i) comprehensive PubMed searches, (ii) information from unpublished health surveys and entomological field studies made available by collaborators, and (iii) internet disease surveillance systems such as HealthMap….Disease and vector specific databases are then made openly available through online depositories which cover the diseases mentioned above…Future efforts include developing the Atlas of Baseline Risk Assessment for Infectious Diseases (ABRAID), an automated mapping platform which integrates the framework described above to generate spatially comprehensive, iteratively improving, evidence based maps of disease risk at the global level for a prioritised number of infectious diseases.

Iwasaki Y[1], Abe T, Wada Y, Wada K, Ikemura T. **Novel bioinformatics strategies for prediction of directional sequence changes in influenza virus genomes and for surveillance of potentially hazardous strains**. BMC Infect Dis. 2013 Aug 21; 13:386. doi: 10.1186/1471-2334-13-386. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3765179/

With the remarkable increase of microbial and viral sequence data obtained from high-throughput DNA sequencers, novel tools are needed for comprehensive analysis of the big sequence data. We have developed "Batch-Learning Self-Organizing Map (BLSOM)" which can characterize … millions of genomic sequences on one plane. .. Important issues for bioinformatics studies of influenza viruses are prediction of genomic sequence changes in the near future and surveillance of potentially hazardous strains… Millions of genomic sequences from infectious microbes and viruses have become available because of their medical and social importance, and BLSOM can characterize the big data and support efficient knowledge discovery.