Sharing data for the public good and protecting individual privacy: informatics solutions to combine different goals

Lucila Ohno-Machado, Editor in chief

doi:10.1136/amiajnl-2012-001513

The bioethics advisory committee to the President has recently issued a report emphasizing the importance of protecting health information, particularly the data about an individual’s genome. The report does not prescribe how to balance the need for sharing information to accelerate discoveries with the potential risk of privacy breach. However, it does mention the potential benefits of data sharing and calls for the development of solutions that minimize the risk of privacy breach. Similarly, a recent report from the NIH’s ‘Workshop on Establishing a Central Resource of Data from Genome Sequencing Projects’ recommends that ‘sequence/phenotype/exposure data sets (be) deposited in one or several central databases.’ Studies on human genomes require good characterization of individual phenotypes, and some of these data may be retrieved from electronic health records. Lessons learned from over a decade of research in privacy technology can help guide solutions to the problem of combining phenotype and genome data in a way that preserves confidentiality. This issue of JAMIA, in addition to several articles we have been publishing in the past few years on technology and policy, explains regulatory constraints, presents a collection of the latest research results on privacy technology, and displays diverse perspectives on the topic of reusing clinical data for research, healthcare quality improvement, and public health.

I am grateful to guest associate editors Malin, O’Keefe, and El Eman for organizing a call for papers and the subsequent reviews for submissions on privacy technology and policy for this special focus issue. These articles represent a variety of subtopics and approaches, ranging from differential privacy (a technical framework that guides safe data disclosure by quantifying the risk of privacy breach to an individual) to discussions on legal aspects of protecting privacy. Malin and his co-guest editors discuss particular articles in their comprehensive editorial.

This issue of the journal also contains reports on how different institutions have been approaching the utilization of clinical and genomic data for research and public health. Hirpszatk discusses how EHRs need to evolve to fulfill current need for comprehensively phenotyping individuals, and Marsolo describes lessons learned when integrating a commercial EHR with research systems. In a provocative article, Witten and Tibshirani express concern about the traditional publishing models, particularly with regards to their inability to ensure that experiments are reproducible. The authors also discuss how the pre-publication review model may be antiquated in an era in which readers have an opportunity to comment on any published article. Farley proposes a platform for biomedical knowledge computing, and Cusack reports on AMIA recommendations for data capture and documentation.

Data sharing requires an environment in which the professionals who handle the data adhere to the highest ethical standards and implement systematic processes that (a) measure data quality, (b) respect consumer preferences, (c) successfully identify research cohorts, and (d) are scalable. The articles by Weiskopf and Weng, Ancker, Ge, and Natter address each of these issues, respectively. Cumin provides an excellent example of data sharing, describing two available datasets for anesthetics records. Avilach describes a European experience for harmonization of the process involved in the identification of medical events in healthcare databases. Jones reuses administrative claims data to supplement a state disease registry. Sharing the knowledge obtained from the analyses of the shared data is an equally important endeavor. Kawamoto discusses a potential framework for knowledge sharing in the context of clinical decision support.

It is exciting to start the new year with an issue of JAMIA that will certainly generate a lot of discussion and lead to a potential re-examination of several current practices and paradigms. My goal is to continue to promote this discussion throughout the year and to keep an open mind to adapting the journal to new trends in scholarly publishing. New publishing models may serve not only our diverse informatics community, but also the scientific and lay communities at large, extending our journal beyond its current boundaries.

Finally, as I complete 2 years of service to JAMIA, I remain indebted to an outstanding editorial team, authors, reviewers, and readers who continue to provide valuable feedback so we can further improve the journal.

Funding The author is partially funded by NIH grant U54HL108460.

Competing interests None.

Provenance and peer review Commissioned; not peer reviewed.

REFERENCES
2. Report on the Workshop on Establishing a Central Resource of Data from Genome Sequencing Projects. June 5–6, Rockville, MD.