Data science and informatics: when it comes to biomedical data, is there a real distinction?

doi:10.1136/amiajnl-2013-002368

Lucila Ohno-Machado, *Editor in chief*

Recent initiatives from the National Institutes of Health (NIH; http://bd2k.nih.gov) and other federal agencies emphasize that health-relevant big data present unique challenges. Professionals who specialize in handling these challenges, that is, data scientists, are in very high demand in all disciplines. Although the name 'data scientist' could imply a significant difference between those who handle data versus those who handle information (ie, processed data), when biomedical, healthcare, and health behavioral data are concerned, there is no distinction: biomedical informatics is biomedical data science. It is great to see that agencies now recognize the value of our discipline and rightfully place it at the top of their priority list. An increasing number of academic institutions are following suit.

There are many reasons why the awareness of biomedical informatics as a scientific discipline has raised in the past few years. Biomedical and behavioral research have been deeply impacted by technologies that enable rapid and relatively inexpensive data collection. New collection sources that generate big data are becoming more and more important and are beginning to merge with more traditional biomedical data sources such as electronic health records. However, expertise and tools to transform these data into knowledge are not being developed at the same speed. Large collections of un- or sub-annotated data can now be collected seamlessly from a variety of sources, but the interpretation of these data are grossly lagging behind their acquisition. Automated, real-time annotation works relatively well for a few types of data (eg, geographical location), but not for others (eg, clinical condition and imaging findings). There is a pressing need to train more experts in biomedical data science/biomedical informatics and to invest in biomedical informatics research so that we can develop tools that make full use of the health-related data that are now being collected in large volumes. Not doing so would put the investments that were made in collecting these data at risk, and miss an opportunity to leverage resources in times where budget constraints threaten current and future generations of scientists. Additionally, we must acknowledge that some of these individuals need to be trained in sub-specialties of informatics that require particular skills.

This issue of the journal focuses on imaging informatics, as well as patient- and provider-centered studies of health information technology. Guest editors for imaging informatics Hsu, Markey, and Wang (***see page 1010***) start this issue with a brief commentary on the key developments in imaging informatics in the past decade as well as the challenges and opportunities of including imaging in the arsenal of tools for precision medicine. As opposed to image processing as the focus of attention, imaging informatics focuses on the use of imaging data for decision making. For example, articles from this issue describe image retrieval technology (***see pages 1014 and 1076***), imaging-based databases (***see pages 1053 and 1046***) including a consumer-based imaging portal (***see page 1028***), imaging tools for prognosis (***see page 1059***) and imaging tools for diagnosis of specific conditions such as glaucoma (***see page 1021***) and intervertebral disc degeneration (***see page 1082***). Since each imaging modality has pros and cons, combinations (eg, magnetic resonance (MR) and positron emission tomography (PET)) can offer diagnostic advantage over each modality alone (***see page 1037***). Studies at a higher level of abstract are also presented in this issue. Determining how radiologists interact with an image information system to make diagnostic decisions can facilitate the design of future systems (***see page 1067***). Radiology is certainly not the only medical specialty that makes extensive use of images. Pathology informatics has become an important sub-field of biomedical informatics that also has important intersections with imaging, and this issue presents some applications in this area (***see pages 1091 and 1099***).

In addition to imaging, this issue also presents articles focused on the effects of health information technology on patients and providers. Patient-provider communication is essential for the success of systems, so it is important to evaluate the impact of electronic resources in particular segments of the population, such as the elderly. North (***see page 1143***) describes the security in patient portals that handle e-visits. Lyles (***see page 1128***) describes the impact of patient-provider communication through a patient portal for the Diabetes and Age Study. Vedel (***see page 1109***) reviews the literature on health information systems in geriatric and gerontology. As we have previously commented in past issues of the journal, a large portion of health-related communications rely on narrative text. MacLean (***see page 1120***) uses crowd sourcing to identify medical terms in patient-authored text, and Huang (***see page 1168***) describes a natural language processing technique to handle ill-formed sentences.

Health information technology can have profound impact on provider workflows and actions. Stubig (***see page 1132***) describes improvements in work efficiency in a trauma center using a real-time WLAN-based location system, and Ross (***see page 1137***) reports on changes in ambulatory testing rates after the implementation of a health information exchange system. Westbrook describes the impact of electronic prescribing on physicians' and nurses' workflow (***see page 1150***), and compares the safety of two e-prescribing systems (***see page 1159***).

The articles in this issue of *JAMIA* represent a broad spectrum of work done by biomedical data scientists, that is, biomedical informaticians who are dedicated to promote health and alleviate the burden of disease through clever transformation of data into knowledge. It is time we celebrate our achievements and the external recognition that our discipline is now receiving.