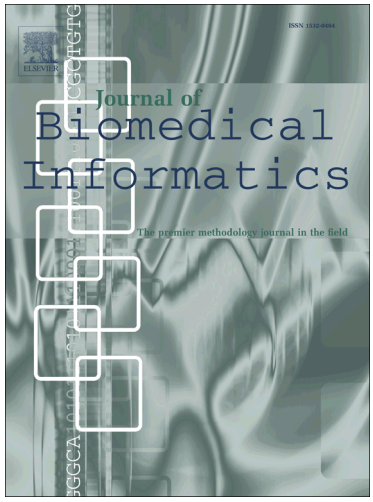

Envisioning the Future of 'Big Data' Biomedicine

Alex A.T. Bui, John Darrell Van Horn, the NIH BD2K Centers Consortium,

PII: S1532-0464(17)30070-9
DOI: http://dx.doi.org/10.1016/j.jbi.2017.03.017
Reference: YJBIN 2751

To appear in: *Journal of Biomedical Informatics*

Received Date: 13 February 2017
Revised Date: 17 March 2017
Accepted Date: 29 March 2017

Please cite this article as: Bui, A.A.T., Darrell Van Horn, J., the NIH BD2K Centers Consortium, Envisioning the Future of 'Big Data' Biomedicine, *Journal of Biomedical Informatics* (2017), doi: http://dx.doi.org/10.1016/j.jbi.2017.03.017

Envisioning the Future of 'Big Data' Biomedicine

Alex A. T. Bui[1], John Darrell Van Horn[2], and the NIH BD2K Centers Consortium*

[1] BD2K Centers Coordinating Center (BD2K CCC), www.bd2kccc.org, University of California Los Angeles, Los Angeles, CA USA, alexbui@bd2kccc.org

[2] BD2K Training Coordinating Center (BD2K TCC), www.bigdatau.org, University of Southern California, Los Angeles, CA USA, jvanhorn@usc.edu

Introduction

In our era of digital biomedicine, data take many forms, from "omics" to imaging, mobile health (mHealth), and electronic health records (EHRs). With the availability of more efficient digital collection methods, scientists in many domains now find themselves confronting ever larger sets of data and trying to make sense of it all (*1-4*). Indeed, data which used to be considered large now seems small as the amount of data now being collected in a single day by an investigator can surpass what might have been generated over his/her career even a decade ago (e.g., (e.g. *5*)). This deluge of biomedical information requires new thinking about how data are generated, managed, and ultimately leveraged to further scientific understanding and for improving healthcare. Responding to this challenge, the National Institutes of Health (NIH) has spearheaded the "Big Data to Knowledge" (BD2K) program (*6*). Data scientists are being engaged through BD2K to guide biomedical researchers through the thickets of data they are producing. NIH Director, Francis Collins, has noted, "Indeed, we are at a point in history where Big Data should not intimidate, but inspire us. We are in the midst of a revolution that is transforming the way we do biomedical research…we just have to devise creative ways to sift through this mountain of data and make sense of it" (*7*). The NIH is now taking its first major steps toward realizing biomedical science as an interdisciplinary "big data" science.

Maximizing the potential of data for everyone

The interplay between biomedical science and advancing technology drives a continuous cycle of data growth: as new technologies enable more and different varieties of data to be amassed, scientists exploit the potential of these technologies and the data being produced to uncover knowledge, and pose new questions that require novel technologies to probe further. Inherent to this accelerating cycle is the requirement to handle the growing data complexity and computational analyses. Towards this, the BD2K program involves the efforts of multiple Centers of Excellence (Table 1), two coordinating centers, and a set of focused individual research and training projects, considering the latest approaches in data science and their application to large-scale biomedical data.