



March 15, 2013

Michelle Dunn, Ph.D.  
Program Director  
National Cancer Institute/NIH  
9000 Rockville Pike  
Bethesda, Maryland 20892

Re: Request for Information (RFI): Training Needs in Response to Big Data to Knowledge (BD2K) Initiative Notice Number: NOT-HG-13-003

Dear Dr. Dunn:

On behalf of AMIA (American Medical Informatics Association), I am pleased to submit these comments in response to the above-referenced request for information. AMIA is the professional home for biomedical and health informatics and is dedicated to the development and application of informatics in support of patient care, public health, teaching, research, administration, and related policy. AMIA seeks to enhance health and healthcare delivery through the transformative use of information and communications technology.

AMIA's 4,000 members advance the use of health information and communications technology in clinical care and clinical research, personal health management, public and population health, and translational science with the ultimate objective of improving health. Our members work throughout the health system in various clinical care, research, academic, government, and commercial organizations.

AMIA thanks the Department of Health and Human Services (the Department) and the National Institutes of Health (NIH) for issuing this request for information and applauds the NIH effort to address this critical and complex topic.

### **General Comments**

**Characteristics and Contents of Plans for Cross-Training Biomedical, Behavioral, Clinical, Computational, and Quantitative Scientists and Informaticians at All Career Levels.** A major AMIA undertaking of relevance to the NIH RFI is our successful multi-year initiative to elevate clinical informatics as a designated subspecialty by the American Board of Medical Specialties (ABMS), which was accomplished in 2012. We are also involved in developing a companion program – the Advanced Inter-professional Certification – which will be provided for informaticians who are not eligible for the ABMS board certification program. A significant

amount of work has been done on the detailed development of core competencies and training requirements.<sup>1 2 3</sup>

**Core Competencies.** AMIA believes the core competencies needed to guide graduate curriculum design for biomedical informatics, which includes the use of Big Data sets, includes the following:

- **Scope and Breadth of Biomedical Informatics:** investigates and supports reasoning, modeling, simulation, experimentation and translation across the spectrum from molecules to individuals to populations, from biological to social systems, bridging basic and clinical research and practice, and the healthcare enterprise.
- **Theory and Methodology:** develops, studies, and applies theories, methods, and process for the generation, storage, retrieval, use, management, and sharing of biomedical data, information, and knowledge.
- **Technological Approach:** builds on and contributes to computer, telecommunication, and information sciences and technologies, emphasizing their application in biomedicine.
- **Human and Social Context:** recognizes that the people are the ultimate users of biomedical information, and so draws upon the social and behavioral sciences to inform the design and evaluation of technical solutions, policies, and the evolution of economic, ethical, social, education, and organizational systems.

**Evaluation of Workforce Skills and Knowledge.** AMIA believes that the ultimate goal driving Big Data education should be to produce methods for supporting biomedical research and improving healthcare, both of which depend critically on human and social context issues. It is important that Big Data users include in their core competencies a familiarity with the fundamental insights of social, organizational, cognitive, and decision science, in addition to the technology competencies of networking, statistical algorithms, security, and databases. Ethical, legal, and social issues need to be addressed in regards to computational decision support, secondary use, and error reduction and management of Big Data. Traditional views of privacy and confidentiality are likely to be seen as inadequate to capture the values of individuals who both benefit by and fear discrimination from machine learning. Training initiatives should include a component that educates users not only about the benefits to be obtained from Big Data, but also the risks of data use practices that facilitate/promote health disparities and

---

<sup>1</sup> Defining the Medical Subspecialty of Clinical Informatics. Don E Detmer, John R Lumpkin, Jeffrey J Williamson. JAMIA 2009;16:167-168 external linkdoi:10.1197/jamia.M3094

<sup>2</sup> AMIA Board White Paper: Core Content for the Subspecialty of Clinical Informatics. Reed M Gardner, J Marc Overhage, Elaine B Steen, Benson S Munger, John H Holmes, Jeffrey J Williamson, Don E Detmer, for the AMIA Board of Directors. JAMIA 2009;16:153-157 external link doi:10.1197/jamia.M3045

<sup>3</sup> AMIA Board White Paper: Program Requirements for Fellowship Education in the Subspecialty of Clinical Informatics. Charles Safran, M Michael Shabot, Benson S Munger, John H Holmes, Elaine B Steen, John R Lumpkin, Don E Detmer, for the AMIA Board of Directors. JAMIA 2009;16:158-166

discrimination and steps to be taken to avoid such practices. The Institute of Medicine (IOM) has explored the need to accelerate research related to health care in the computer and social sciences and in health/biomedical informatics.<sup>4</sup>

The National Library of Medicine (NLM) informatics training programs are examples of programs and structures that could be expanded to help address the training and workforce needs, especially at the doctoral/postdoctoral level. The National Library of Medicine supports research training in biomedical informatics at selected educational institutions in the United States. These programs offer graduate education and postdoctoral research experiences in a wide range of areas including: health care informatics, translational bioinformatics, clinical research informatics, and public health informatics.<sup>5</sup>

### **Interprofessional Education Preparation to Support Effective Use of Complex Data Sets**

We believe that interprofessional education models should be established, implemented and evaluated across a broad array of groups with limited prior collaboration experience. Increasingly, technically- and clinically-oriented individuals need to communicate with each other and need to have some basic understanding of each others' perspective, language, and priorities. Employers need the workforce to demonstrate a greater convergence of skill sets across and within disciplines to use large-scale and diverse data sets to inform clinical decision-making. Implement team-based strategies to fuse data analysts with those who are prepared to translate data to knowledge. In summary, the future workforce needs to have a combined knowledge of clinical systems, a basic understanding of health care workflow, and basic familiarity with vocabulary at the level of a computer science graduate in the relevant areas. A closer linkage between "industry" (employer) needs and educational programs is recommended, emphasizing hands-on internships, fellowships and participation in "real-world" situations.

Other possible approaches include the following:

- Create virtual data laboratories to foster collaboration in the use and sharing of large data sets by colleagues globally.
- Create integrated curricula to address learning needs related to data mining, text retrieval, forecasting and machine learning techniques.
- Rapidly create and disseminate high-quality educational offerings and certification programs that are web-based and flexible to support ubiquitous access.
- Promote retention of entry-level employees by designing, implementing and evaluating career pathways and ladders.
- Design for diversity in the broadest sense in fostering the future workforce.

---

<sup>4</sup> Willam W. Stead and Herbert S. Lin, editors; Committee on Engaging the Computer Science Research Community in Health Care Informatics; National Research Council . Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions. [http://www.nap.edu/catalog.php?record\\_id=12572](http://www.nap.edu/catalog.php?record_id=12572)

<sup>5</sup> <http://www.nlm.nih.gov/ep/GrantTrainInstitute.html>

Further, the IOM and others have identified the increasing need for team based approaches.<sup>6 7</sup> We believe that health information systems design and development, and the implementation of big data systems increasingly require team-based approaches to user needs assessments, software engineering, design, and implementation. Talented candidates should be recruited from diverse disciplines and backgrounds such as (a) medicine, nursing, public health, allied health, or biological, (b) mathematical, economic, financial, computational, physical sciences, information sciences, or engineering, and (c) cognitive or social sciences. These diverse teams need to learn to communicate on all levels consistently and adequately, which involves use of a common language as well as team-based strategies that fuse data analysis with translation of data to knowledge. The Electronic Data Methods (EDM) Forum,<sup>8 9</sup> supported by Agency for Health Care Research and Quality (AHRQ) is an example of an initiative that highlights the critical need for these kinds of skills. The EDM Forum convenes investigators to advance the national dialogue and facilitate exchange and collaboration between AHRQ-funded projects on the use of electronic clinical data for the conduct of comparative effectiveness research (CER), patient-centered outcomes research (PCOR), and quality improvement (QI).

Specific topics that new graduates require (but often lack) include the following:

- The significant contribution that big data will make to improving health care and public health
- Understanding and appreciation of issues related to data quality, data representation, data timeliness, and data completeness
- Data governance and stewardship<sup>10 11</sup>
- Issues regarding the USE of data (including privacy, confidentiality and security)
- Security standards and procedures
- Privacy standards and procedures with conceptual, legal and technological implications
- Reporting strategies spanning operational to strategic levels
- Need to advance both standards and common nomenclatures
- Requirement to define comprehensive data management plan to address the dissemination of research outcomes and work products
- Prediction strategies, modeling methods and new technologies to disseminate data all enable new ways to approach scientific and medical discoveries.

---

<sup>6</sup> IOM (Institute of Medicine). Health professions education: A bridge to quality. Washington, DC: The National Academies Press; 2003.

<sup>7</sup> IOM (Institute of Medicine). The learning healthcare system: Workshop summary. Washington, DC: The National Academies Press; 2007.

<sup>8</sup> <http://www.edm-forum.org/publicgrant/About/projectprofiles/>

<sup>9</sup> <http://grants.nih.gov/grants/guide/rfa-files/RFA-HS-13-004.html>

<sup>10</sup> Safran C, Bloomrosen M, Hammond WE, Labkoff S, Markel-Fox S, Tang PC, Detmer DE, Expert Panel. Toward a national framework for the secondary use of health data: an American Medical Informatics Association White Paper. J Am Med Inform Assoc. 2007 Jan-Feb;14(1):1-9. Epub 2006 Oct 31.

<sup>11</sup> Bloomrosen M, Detmer D. Advancing the framework: use of health data--a report of a working conference of the American Medical Informatics Association. J Am Med Inform Assoc. 2008 Nov-Dec;15(6):715-22. doi: 10.1197/jamia.M2905. Epub 2008 Aug 28.

- A foundation in statistics and strong query/analytical skills
- The benefits and limitations of tools used for and with big data as well as issues regarding “real-time” data analysis

**Development of a Diverse Research Workforce.** One of our members noted the following: *Many minority students who could be candidates for training attend public state schools, which are often less well funded than private schools, and face severe budget cuts and closings. These state schools may be less connected to the large research universities and other more well known institutions using ‘big data’ analysis. Thus, many (current and potential future) students and faculty may not be aware of the types of research and applications being performed, or possible, and are unlikely to gain such exposure. Thus, these individuals (or groups) may not be seeking training in these avenues which are likely to be sources of employment and research, shaping future work and medical parameters.*

*The situation of adult learners who may be highly motivated but must either fit education around work commitments or need assurance that full-time study would ‘pay off’ in concurrent support and increased future job availability, must be addressed. The ability to see a clear and well-supported (financially and in advisement) pathway between where they are now and becoming fully trained, would be very important. In addition, because of the work commitments of adult learners they are often unable to attend conferences, meetings during the day (even if available in the same city) and so on, so have extra difficulty staying current, networking, and having a body of exposure to the problems and tools involved.*

Thus, we endorse efforts by the National Institute of Minority Health and Health Disparities and other NIH institutes to actively recruit racial and ethnic minority candidates for scholarships and fellowships, and encourage initiatives at all levels of training to help diversify the workforce.

### **Concluding Comments**

AMIA appreciates the opportunity to submit these comments. Again, we thank the NIH for issuing this request for information. Please feel free to contact me or Meryl Bloomrosen, AMIA’s Vice President for Public Policy at any time for further discussion of the issues raised here.

Sincerely,



Kevin Fickenscher, MD  
AMIA President and CEO