

# Transformative Changes to Embrace, Manage, and Exploit “Big Data”

*Authors’ Commentary*

## CHAPTER ABSTRACT:

Technological advances in research and electronic health records have shifted the emphasis at academic health centers from a focus on data generation to one of data management and analysis.

Transformation of the academic health center environment, workflow, and workforce is needed to fully exploit the potential of “big data” in accelerating the discovery and application of new biomedical knowledge. This chapter provides a detailed framework for how academic health centers can establish their own big data resource initiatives to exploit large national databases and merge data from throughout their research and clinical enterprises.

While it is easy to talk about “big data,” it is much more complicated to put an infrastructure in place that effectively addresses the realities of exploiting big data in meaningful ways. Part of the complexity stems from the fact that there are various sorts of big data. For example, some big data exist in forms that are relatively easy to compute, such as claims data from insurance. By contrast, though, the electronic health record, with its abundance of free text, produces data that are more daunting to process.

We were well aware of those issues before we undertook this chapter, but writing our thoughts in this form confirmed and underscored that, when it comes to making sense of big data, the vision and the reality remain far apart because of how challenging a problem it is. Individual academic health centers have the opportunity to establish big data resource initiatives that can exploit large national databases, merge data from throughout their research and clinical enterprises, and accelerate the discovery and application of new biomedical knowledge. But doing so effectively requires that each academic health center carefully plan the right big data environment, workflow, and workforce.

Based on our experiences at Pitt, we found that it is important to adequately plan a computing infrastructure. Big data, both from clinical care and research, are only valuable if they are recorded with care, using a nomenclature that permits subsequent merging and sharing for integrated analyses. It is imperative to plan an infrastructure that anticipates growth in data volume and expansion of data types as well as to invest in developing an institutional culture that fundamentally understands the importance of big data. Healthcare personnel at every level, from clinic intake staff to physicians, must work with a mindset that ensures the entry of complete, accurate, and uniform data—good data “hygiene.”

In terms of infrastructure, rather than reinvent the wheel, academic health centers should explore partnerships that can help them gain the expertise they need—such as university partners and other collaborators who can help design and implement an appropriate infrastructure—and help them avoid mistakes, including overspending, that result too often when an institution tries to go it alone. At Pitt, we took advantage of existing expertise in the schools of health sciences, other parts of the university, and in Pittsburgh at large. We describe some of these partnerships in the chapter, in the hope that they might be illustrative for other academic health centers.

In fact, we recently announced a new and very substantial agreement related to healthcare and big data. This agreement engages our

**Arthur S. Levine, MD**

*Senior Vice Chancellor for the Health Sciences  
John and Gertrude Petersen  
Dean, School of Medicine  
Professor of Medicine and  
Molecular Genetics  
University of Pittsburgh*

**Michelle L. Kienholz**

*Science Communications,  
Institute for  
Personalized Medicine  
University of Pittsburgh*

**Rebecca S. Crowley,  
MD, MS**

*Professor of  
Biomedical Informatics  
Department of  
Biomedical Informatics  
Chief Information Officer,  
Institute for  
Personalized Medicine  
Director, Biomedical  
Informatics Graduate  
Training Program  
University of Pittsburgh*

**Jeremy M. Berg, PhD**

*Associate Senior Vice  
Chancellor for Science  
Strategy and Planning  
Health Sciences  
Pittsburgh Foundation Chair  
and Director, Institute for  
Personalized Medicine  
Professor of Computational  
and Systems Biology  
University of Pittsburgh*

University, Carnegie Mellon University (CMU), and the University of Pittsburgh Medical Center (UPMC). UPMC holds one of the largest patient data-bases in the United States; CMU has a top-tier computer science/machine learning program; and, our medical school has leading departments of biomedical informatics and computational biology. The goal of this agreement is to identify aspects of this new tri-partite platform that can be commercialized. UPMC will provide abundant capital in support of commercialization.

As with so many innovations, it is important for academic health centers to do their homework before plunging into the vital but challenging world of big data.

