by terence k. huwe
Director of Library and Information Resources
Institute for Research on Labor and Employment
University of California–Berkeley

Big Data, Big Future

WHAT ABOUT THE INFORMATION PROFESSIONS? BIG DATA PRESENTS A VASTLY IMPORTANT NEW OPPORTUNITY FOR US.

The winter and early spring of 2012 have seen a whole new level of excitement on the topic of Big Data. The idea of Big Data—certainly the meme du jour these days—was already sweeping through society and the academy at a fast clip. The tenor of speculation about Big Data took a more serious (and interesting) turn in the wake of the World Economic Forum in Davos, Switzerland, where its potential to remake the marketplace, and perhaps even to drive more effective economic development across the globe, was a major topic. *The New York Times*, ever a worthy commentator on the topic of information and society, published a very entertaining column titled "The Age of Big Data" on Feb. 12, 2012. Forrester Research and Bloomberg have paid close attention to the topic too. *The Times* column cited a conference held by the National Bureau of Economic Research (NBER) titled Opportunities in Big Data—as in, career opportunities. And as if all this weren't enough, here comes the initial public offering of Facebook, itself a major Big Data producer as well as the subject of Big Data analysis. Indeed, the reach of Big Data even informs the uniquely robust real estate market in San Francisco, which is seeing a dramatic uptick in home sales, involving multiple offers and exceeding the asking price. Reason: Qualified buyers are "rushing in" to beat the highly anticipated wave of Facebook millionaires who prefer San Francisco to Silicon Valley, driving prices higher. Big Data is everywhere, and it's gaining the attention of everyone.

The commentators I mention here address the potential of Big Data for society, but what about the information professions? In many ways, Big Data presents a vastly important new opportunity for us; the NBER was prescient to hold a conference about opportunities to come, and the same applies to the information professions. However, to get into the Big Data game, we may need to create our own opportunities. To gain the best strategic sense of what we can offer (as well as create), we will need to focus on two core activities. First, we will need to reach beyond our native sources of wisdom and practice—we need to think like economists or others who already handle Big Data. Second, we will need to collaborate with these like-minded data jugglers, quite possibly on their turf instead of our own.

Events are moving quickly now, so speculating on what is to come between the point of a columnist's deadline and the issue going to press is getting a bit tricky. Nonetheless, I will take a stab at Big Data, crossover thinking, and strategic preparation, by assessing some of the forces driving Big Data application. However, I have a feeling that the data manipulators among us, once they get started on their own, may trump what's being said now in short order. Certainly, I hope that is the case.

Data-Driven Management

Working closely with organizational behavior experts as I do, I've come to respect the tradition of trial and error in the real

world of management. Certainly there are "fads" galore in the field of business administration, but it's worth noting that many of these fads (think Total Quality Management, Re-engineering, and so on) have their roots in the practice of observation, experimentation, and not least, intuition. Big Data brings nonnegotiable facts into the mix, enabling managers to base vital decisions on solid information, accumulated from a rich variety of sources and delivered in real time. The American workplace has embraced data-driven management, using web search strings, what's trending on Twitter and Facebook, real estate dynamics (such as my pre-Facebook IPO example from San Francisco), and much more. The term "data-driven" has made its way into the pantheon of library managers' feedback loops in the past few years too.

Data-driven management puts a more solid floor under daring business plans and has delivered results that can be quantified. For libraries, the challenge will be to move beyond our own data streams and include related data. We can start now through collaboration close to home; most college campuses and even municipalities are studying data trends. They might welcome the addition of ours and share their own, to a mutual advantage.

Data Modeling as Metaphor

The field of economics depends upon modeling to control for quirks in data and false discoveries, as well as to neutralize biases and beliefs that may skew objective study. In this respect, data modeling operates much like metaphor does in language—a fact also well stated by Steve Lohr in *The Times*' column I cited. He also points out that models and metaphors can mislead, depending on what is being assessed. This matters a lot when it comes to individual privacy, banking and credit data, and other sensitive areas. For Big Data to thrive, strong safeguards are needed to protect people. At the same time, vast streams of data need not be so personal to be useful. I can foresee that maximizing our use of data such as admissions trends, evolution in curricula, syllabi, and other

easy-to-obtain data streams will inform superior, targeted outreach services. Once again, a strong start in this area depends upon collaboration—so we can influence the metaphors we will live by.

The Power of Prediction

Big Data allows metaphor to become more specific and fine-grained, revealing more insight and wisdom. In actual practice, this means that analyzing online search requests can show "hot spots" that precede events, such as the common flu coming on strong, market trends, and real estate spikes (OK, I admit it. I wish I had bought in Noe Valley when I could have). In our own areas of expertise, a better sense of predictive awareness could be a boon for reference providers, information literacy instructors, and collection developers. In particular, I think that a strong awareness of educational trends, including student life, the classroom, and administration, could strengthen the role of the informational professional as a stakeholder and contributor to developing the universities of the 21st century. We do not come to the table empty-handed, but rather with data in hand. If we can craft our analyses more finely, we may gain attention as potential collaborators and developers ourselves.

Zero Degrees of Separation

The concept of "six degrees of separation" was established in the 1960s and has had some staying power, but social networks have evolved since then. This was already underway during the internet's infancy, when email took the place of the telephone cold call as a way to open communication. Nowadays, we can know lots more about our close and not-so-close friends; LinkedIn harnesses this as the "new normal" and seeks to transform it into a networking (and job-hunting) vehicle. The same associative power informs web search analysis to uncover trends. It is also being explored in the next wave of integrated library systems, as a means to recommend related content on the basis of search history. Once again, this kind of service must preserve privacy. But it can also transform the way research, teaching, and scholarship operates.

THE BENEFITS OF STAKING A CLAIM IN THE BIG DATA WORLD FLOW NOT ONLY TO OUR FIELD BUT TO SOCIETY AT LARGE.

One example may be seen in ResearchGate, the Berlin-based scientific network that launched in 2008. Taking a lead from the Public Library of Science (PLoS), ResearchGate seeks to combine the useful characteristics of social networking, Facebook-style, with the rigorous culture of peer review. ResearchGate might be focusing on scholarly publishing and the workflows that precede publication; because of this, its own data stream, conducted in a climate of "open science," will itself yield fascinating insights about how scholars will conduct their work in the near and long terms. It has also attracted venture capital, which lends some staying power (see www.researchgate.net).

Big Data, Our Future

The scale and scope of Big Data cry out for serious research and development in our world as well as in the world of "big science." Unfortunately, "big funding" is harder to come by these days, even at the largest institutions. Nonetheless, there is a growing consensus, within and beyond the profession, that Big Data is here to stay, we need to get used to it, and we would be well-advised to harness its potential. Our recent history with web development and content creation suggests that we have what it takes to join the ranks of Big Data analysts. Many library mashups are pretty good and capture the best of that crucial balance between static and dynamic information. The availability of Big Data invites experimentation, as well as collaboration, and our finest experimenters can apply what they have learned using the more familiar world of mashups in posing questions we might ask the data stream.

If we are to ask any questions at all, the best place to start is in the context of community. Asking our questions out loud and in the company of our academic colleagues is a very public way of joining the dialogue on Big Data. Furthermore, the ability to traverse intellectual boundaries is crucial, and everyone following Big Data finds themselves dabbling across the spectrum of knowledge—which in turn strengthens collaboration. Unless I'm sorely mistaken, we are as good as or better at forming collaborative alliances than many of our allied colleagues. The benefits of staking a claim in the Big Data world flow not only to our field but to society at large too. After all, ResearchGate defines its mission as follows: "ResearchGate was built for scientists, by scientists, with the idea that science can do more when it's driven by collaboration." This sentiment applies to our world as well, and if ever there were a cohort of collaborators, it is we. ■

Terence K. Huwe is director of library and information resources at the Institute for Research on Labor and Employment at the University of California–Berkeley. His responsibilities include library administration, reference, and overseeing web services. His email address is thuwe@library.Berkeley.edu.