



Top Influence

David Alan Grier, Djaghe, LLC

By IEEE measures, a 2013 article on big data is the most influential article ever published in Computer. Does that surprise you?

How do you write a highly influential article for *Computer*? It is not that difficult, if you look at the example of the most influential article that the magazine has ever published (see “Article Facts”). First, organize a special issue on a topic that is about to become prominent in both public discussion and the work of our members. Second, identify five authorial teams that address the key issues of the topic, ranging from the current state of research to the way that topic will influence other parts of computing, and end with a discussion of emerging applications. Finally, write a clear and succinct summary of the issue that explains all of the controversies that will surround this new topic for the next five years. Then get everything to the managing editor on deadline, sign the copyright transfer forms, and you will have the most cited, viewed, and downloaded article in *Computer*’s back catalog.

It is, of course, easy to dismiss this strategy as relying a little too much on happenstance and confusing popularity with influence. Both claims are serious and deserve serious attention. However, before we consider both of them, let us identify the article. The article with the most down-

loads and citations is the editor’s introduction to the June 2013 special issue on big data, “Big Data: New Opportunities and New Challenges,” by Katina Michael and Keith Miller.¹

In 2013, big data was a relatively new topic that was discussed primarily by a small group of insiders. For the years

ARTICLE FACTS

- » Article: “Big Data: New Opportunities and New Challenges”
- » Authors: K. Michael and K. Miller
- » Citation: *Computer*, vol. 46, no. 6, pp. 22–24, June 2013
- » *Computer* influence rank: #1 with 18,486 downloads and 108 citations

Digital Object Identifier 10.1109/MC.2020.2964890
Date of current version: 12 March 2020



prior to that date, IEEE *Xplore* lists only 73 articles with the term *big data* in its title and shows that the vast majority of those come from specialized conferences about databases, sensors, and knowledge engineering. The figures for the Association for Computing Machinery library are similar.

A couple of other dates help put this issue in context. The Hadoop File System from Apache, which had become important to a number of big data applications, was first released (in version 0.1.0) in April 2006. In 2011 and 2012, the IEEE Computer Society (CS) was developing a new community around the technology of cloud computing. During those discussions, several members noted that cloud computing would be an important technology for the emerging problems of big data.

From the perspective of the authors, the timing of the special issue required little special foresight. Michael, who was both coauthor of the article and coeditor of the special issue, worked within the IEEE Society on Social Implications of Technology and was aware of the growing importance of big data. “Everyone was talking big data in 2012,” she recalled, “but just like any other new buzz term, there was not much actually written on it.” At the time, Michael was editor-in-chief of *IEEE Technology and Society Magazine* and “deeply concerned about the privacy and security challenges” that big data posed on ordinary citizens. “I could see the benefits,” she remembered, “and the costs.” Initially she and Miller, her coauthor, wanted to call the issue “The Social Implications of Big Data: Pros and Cons,” but ultimately agreed to the name “Big Data: New Opportunities and New Challenges.”

If the article and the corresponding special issue grew out of a specialized community, it gained its influence by building on the strength of that community. To understand how

that influence developed and how it differed from simple popularity, we need to understand how professional technical articles are used in the work of our members. Broadly speaking, a technical article can do one of two things. First, it could be part of a conversation that is trying to find the best way to solve a technical problem. Second, the article can present a technology as a developed solution for other problems.

This is a gross simplification of the problem, and we will develop it in more detail as this column progresses.

If the article and the corresponding special issue grew out of a specialized community, it gained its influence by building on the strength of that community.

However, it is a well-recognized phenomenon in the technical literature that is known as *black-boxing*. While a technology is new and being developed, it is open for discussion and debate by the community working on it. At some point, that community concludes that it has found the best way of handling that technology, and it ceases to discuss different possible solutions. At that point, the technology is said to have been black-boxed. The community then goes looking for ways to apply the technology.

In the CS literature, the debate over the best possible technical solutions tend to take place in conference proceedings and transactions. The presentation of black-boxed technologies tends to take place in magazines and books. Again, this is a gross simplification of how the literature works. Yet, we can see some evidence of these two kinds of technical articles in the citation lists of the periodicals. Articles in transactions and proceedings tend

to cite articles from the same transactions or proceedings series. This is one way that we can see the discussion over the best technical solution to a problem.

In contrast, the articles in *Computer* and some of the other CS magazines rarely cite other articles from their prior issues. These articles tend to cite articles from transactions and conferences. In doing this, they are not engaging in the discussion over the best way to solve a problem. They are presenting mature ideas as possible solutions to other technical solutions.

So one way of assessing the importance of an article is to determine which kind of periodical is citing it. If it is being cited by proceedings and transactions, it is part of the discussion over the results of research; it is being cited in the debate about how to best solve a technical problem. If it is being cited by magazines and books, it is likely part of the presentation of a finished technology.

Again, while this characterization of articles is somewhat crude, it reveals an interesting fact about Michael and Miller’s introduction from the 2013 special issue on big data. It is being used in both kinds of technical discussions. Within the IEEE literature, it was cited by 32 conference proceedings and 11 magazines. Outside the IEEE, the split is about equal. Of the 93 citations it has garnered, 49 are from conferences or transactions. The rest are from magazines or books. Taken as a whole, it is evidence that the article was widely read and

used in a variety of situations. It was cited as part of the debate over big data technology. It was also used to present the ideas of big data to both the broad computer community and a bigger technical community. The response to the special issue, noted Michael, “has really gone far beyond our wildest expectations.”

Of course, one of the reasons that this article has been so widely circulated has been the fact that it was part of a well-researched and well-written special issue. “We had a very long list of people who we wanted to approach,” Michael explained. “People from academia of course, but also people working with industry, people from big corporations, and even government.” The five other articles are a good representation of these institutions, and they, too, have become highly visible in their own right. All of them

are in the list of top articles downloaded from IEEE Xplore.

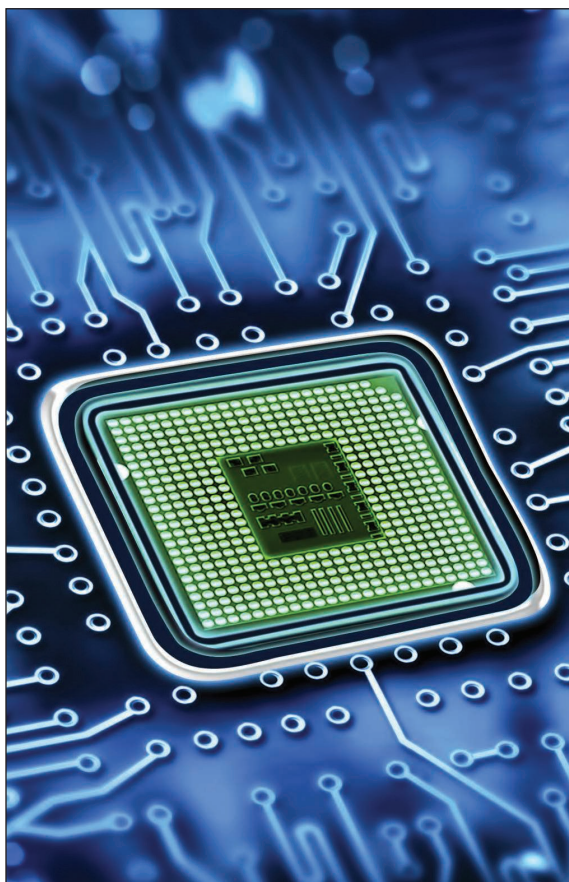
Yet, perhaps the most convincing reason that “Big Data: New Opportunities and New Challenges” has become an important article is the language of the text itself. At a distance of seven years, it describes the opportunities of this technology as well as its challenges in sentences that are direct and clear and reflect the nature of big data as we now understand it. “While big data can yield extremely useful information,” the article explains, “it also presents new challenges with respect to how much data to store, how much this will cost, whether the data will be secure, and how long it must be maintained.” If you are seeking to write an influential article for *Computer*, you will need this third step as well as the other two. After you have identified

a newly important topic for a special issue and found good authors to complete your special issue, you will do well to express your ideas in a clear, open form that will be quoted by current researchers as well as those who are attempting to understand the new technologies. **C**

REFERENCE

1. K. Michael and K. Miller, “Big data: New opportunities and new challenges,” *Computer*, vol. 46, no. 6, pp. 22–24, June 2013. doi: 10.1109/MC.2013.196.

DAVID ALAN GRIER is a principal with Djaghe, LLC. He is a Fellow of the IEEE. Contact him at grier@gwu.edu.



IEEE TRANSACTIONS ON

COMPUTERS

Call for Papers: *IEEE Transactions on Computers*

Publish your work in the IEEE Computer Society's flagship journal, *IEEE Transactions on Computers*. The journal seeks papers on everything from computer architecture and software systems to machine learning and quantum computing.

Learn about calls for papers
and submission details at
www.computer.org/tc.

Digital Object Identifier 10.1109/MC.2020.2974645

