



## Outside Counsel

## Expert Analysis

# Intellectual Property Implications And Opportunities in Big Data

**B**ig data has become a major driving force with broad applicability and implications. From air transport to retail to government to entertainment, everyone has access to enormous quantities of information that can be characterized as “big data.” It seems almost every organization today is either currently using big data or looking for ways to use it. Those who are doing neither proceed at their considerable peril.

This relatively new and expanding opportunity adds import to the role of forward-looking intellectual property systems. Capturing and building upon intra-organizational innovation, surveying patterns of innovation in the surrounding field, and understanding the changing IP legal landscape will put an organization in the best position to capitalize on developments in big data.

But what questions should we be asking of big data? This article begins with a discussion of some of the overarching business questions raised by big data. It proceeds to examine two competing frameworks for utilizing big data findings and closes with prescriptive ideas for managing and developing IP in the big data space.

### Big Picture: Asking Questions

Big data brings with it many business opportunities and challenges. It may lead to the discovery of correlations potentially helpful to the business; for example, a clothing retailer might use big data analysis to discover which day of the week 23- to 29-year-old working professionals buy the most cotton-wool blend socks online. So the first question asked should be what information—or, more specifically, what correlations—can be ascertained from data that would be helpful?

From that question, an organization may ask: Are we prepared to change business direction based on what we learn from data? This question of business direction in turn raises more



By David J. Kappos

technical challenges: Where do we find useful data? How do we detect and correct inaccurate records within the data (also known as “data cleansing”)? How do we gain search access to unstructured and heterogeneous data? How do we parse and analyze the data in order to gain insight from it?

Throughout this process, questions must be asked as to how data was obtained—and whether it is reliable. How confident can we be in a particular insight—and why? How do we seek out those questions we don’t know to ask of the data (the “stuff we don’t know we don’t know”)? And from the business and technical direction we encounter intellectual property challenges: Are new approaches needed to identify and capture inventions around big data? How do we optimally describe big data inventions in patent applications? How should patent claims be framed to protect big data inventions in ways that will withstand the test of shifting legal standards? How do we avoid—or embrace, if necessary—multijurisdictional infringement issues, divided/joint infringement issues, induced infringement issues, and extra-territorial enforcement issues? Perhaps most importantly, how do we ensure others don’t get patents that impede or prevent our own use of big data to benefit our customers and obtain maximum competitive advantage?

In addition to the first- and second-order questions listed above, there are even larger, strategic questions that will drive the development of a tailored approach to intellectual property management and development. These questions relate to how big data fits within an organization’s business, how it will be used, and who within

the organization possesses the skills required to harness the promise of big data while recognizing and avoiding traps inherent in the reliance on it. At the source of these strategic questions are the promises and pitfalls of two approaches: an approach based primarily on correlation versus an approach based primarily on causation.

### Correlation and Causation

Some leading thinkers posit that the value of big data lies in the promise of uncovering previously unknown correlations. Such discovery, the argument holds, allows businesses to profit from theretofore unseen connections. The guiding principle for these thinkers is the more data the better—at a certain point, the numbers speak for themselves. This line of thinking demonstrates a focus on the “what” while minimizing emphasis on the “why” of big data analysis.

In the socks example posited above, according

---

Big data brings with it many business opportunities and challenges. It may lead to the discovery of correlations potentially helpful to the business.

to this line of thinking, knowing that 23- to 29-year-old working professionals were more likely to buy cotton-wool blend socks online on Tuesdays than on Thursdays would be valuable in and of itself—even if there were no articulable theory of why Tuesday was a better day to buy socks.

Other leading thinkers concede that analysis of big data might reveal previously unknown correlations, but urge cautious interpretation of these newly revealed correlations. The guiding principle for these thinkers is that mere correlation, without meaningful exploration of causation, fails as an effective strategic guidepost. Numbers do not speak for themselves but rather are given a voice by those who gather and interpret them.

In the words of Albert Einstein (as quoted by New York Times reporter and blogger on big data issues Steve Lohr): “Not everything that

counts can be counted, and not everything that can be counted counts." Adherents to this line of thinking warn against overreliance on discovered correlations (the "what") with only limited investigation into the causation underlying those correlations (the "why"). Thus, they might discourage a clothing retailer from adopting a new sales strategy that relies solely on the Tuesday correlation to online sock purchases without first exploring why sales to a particular demographic tend to be higher on that day.

Put differently, there are two approaches to managing the correlation/causality interface. The first, "read the gauges" approach demands that the organization put primacy on the correlations discovered and for the most part suspend the search for causation. If massive amounts of reliable data reveal that socks sell better on Tuesdays than on Thursdays, then the correlation itself justifies a strategic approach that takes that correlation into account.

The second, more conservative approach recognizes that statistical inferences and correlations are stronger and more safely used when backed up by an understanding of their root causes. The fact that 23- to 29-year-old working professionals bought more socks online on Tuesdays than Thursdays could be due to any number of factors—from the day those workers deposit paychecks to merely coincidental weather patterns on Tuesdays. Adjusting strategy based on correlation alone, from this viewpoint, is a relative shot in the dark.

As a more salient example of the pitfalls inherent in this correlation/causation duality, consider the critical mistake of a leading financial forecasting firm. The firm had previously advised its clients that there was no need to understand the intricacies of the various data elements of the economy; all that was necessary was an ability to "accurately read those gauges." Yet the firm's straightforward approach of merely reading the gauges failed to accurately forecast the direction of the economy in 2011. The firm, relying on a mix of variables backed by correlations it uncovered through big data analysis, predicted a deep recession. As it turned out, the market actually took a turn for the better, with the S&P 500 gaining 21 percent in five months and GDP growing 3 percent.

This example illustrates the danger of relying too much on finding the "what" in complex correlations within vast data sets without sufficiently investigating the "why" undergirding those correlative findings. This cautionary tale not only offers guidance as to how big data might be more productively applied, but also

guidance into building the best team for seeking big data innovations. In order to avoid the pitfall of overreliance on correlations (to the exclusion of adequately investigating underlying causes) an organization's ideal team should include both talented data analysts and deep thinking actuarial scientists. The composition of the team, and its relative focus on correlation versus causality, will factor into the calibration of an intellectual property strategy best suited to bolster an organization's big data goals.

### Intellectual Property

Protecting innovations in the big data space demands a disciplined approach to intellectual property development and management. A successful strategy is one that incorporates stakeholder perspectives and accounts for IP law trajectory.

Protecting innovations in the big data space demands a disciplined approach to intellectual property development and management. A successful strategy is one that incorporates stakeholder perspectives and accounts for IP law trajectory. The four-step process outlined below—which, depending on organizational capacity, may benefit from consultation with outside IP experts—posits a framework for developing such an approach.

**1. Stakeholder Focus Groups.** An initial step toward developing an appropriate IP strategy is to facilitate meetings with program managers, product/service developers, actuarial scientists, data analysts, and software programmers. The goal of these meetings is to identify key directions, problems, drivers, and opportunities that occupy their attention in the course of business. These discussions will be informative as to where sustained attention should be focused in identifying inventions within the organization and monitoring the disclosures and patent filings of industry rivals. They will also provide baseline information that is prerequisite to facilitating invention development within the organization ahead of the field, the subject of Step 2.

**2. Invention Generation Sessions.** A workshop or series of workshops ought to be arranged that brings together the organization's brightest technologists, analysts, actuarial scientists,

program managers, and software developers for the purpose of generating invention ideas. An appropriate goal for such a workshop might be to leave with five to 10 inventions—inventions that are aimed at major high-level challenges faced in the big data space particular to the organization or the industry in which the organization operates. The key is to look for potential inventions—and identify the disclosures to be built around them—that would lead to broad future-focused claims by the organization and at the same time preempt competitors who would seek to make broad filings of their own.

**3. Addressing IP Law Issues.** Using the insights and current-project inventions flowing from Step 1, and the broad-based invention ideas flowing from Step 2, an organization should carefully consider legal risk. Shifting legal doctrines (such as divided/joint infringement, inducement, and extraterritoriality) present strategic obstacles to the development of valuable intellectual property in the big data space. A firm understanding of these shifting legal doctrines, and their impact on the value of specific IP, is crucial to a successful approach.

**4. Processes to Systematize Learning.** The final step is to develop templates and best practices that the organization can use on an ongoing basis to facilitate the preparation and prosecution of patent applications. These practices should focus on creating an efficient flow of invention gathering and evaluation for inventions created within the organization as well as evaluation of issued patents that may be valuable acquisitions in building a robust IP portfolio. Creating a systemized learning approach will provide the dual benefit of harnessing the innovation already occurring in the big data space and minimizing the administrative costs associated with funneling that innovation toward actionable intellectual property outcomes.

### Conclusion

Big data is big business. It is here to stay. As such, proactive intellectual property business leaders and attorneys are well-served to understand core causation and correlations challenges relevant to their industry, and implement IP strategies to position themselves and their businesses for long-term success.

Reprinted with permission from the February 21, 2014 edition of the NEW YORK LAW JOURNAL © 2014 ALM Media Properties, LLC. All rights reserved. Further duplication without permission is prohibited. For information, contact 877-257-3382 or reprints@alm.com. #070-02-14-26