2014

Big Business, Big Government and Big Legal Questions

Michael Mattioli
Indiana University Maurer School of Law, mmattiol@indiana.edu

Todd Vare
Barnes & Thornburg

Recommended Citation

Mattioli, Michael and Vare, Todd, "Big Business, Big Government and Big Legal Questions" (2014). *Articles by Maurer Faculty*. 2791.
https://www.repository.law.indiana.edu/facpub/2791

Big business, big government and big legal questions

Big data is here to stay, but what does it mean for IP? **Todd Vare** and **Michael Mattioli** look at how big data can be protected in the US, and ask whether it can thrive under the existing legal framework

According to IBM, 2.5 quintillion bytes of data is created every day, and 90% of the data in the world today has been created just in the last two years. In 2013, the amount of data generated worldwide was estimated to be 4 zettabytes.

This is big data. It is here to stay; indeed, it will only grow exponentially. IBM and countless other technology companies have recognised the commercial opportunities in big data. Some experts estimate that big data will be a $28.5 billion market in 2014, growing to $50.1 billion in 2015. According to IBM, 300 of its patents secured in 2012 related to innovation in the big data analytics field. One notable example is US patent 8,275,803, which claims a system, method, and computer program product for providing answers to questions based on any corpus of data – dubbed the Watson system, a smart question-answering computer.

Big data also has caught the eye of the government. In March 2012, the Obama Administration announced a Big Data Research and Development Initiative, with the commitment of more than $200 million from six federal government departments and agencies to fund big data research and use. The White House explained: "By improving our ability to extract knowledge and insights from large and complex collections of digital data, the initiative promises to help solve some the Nation's most pressing challenges."

Big data clearly presents enormous opportunities for businesses and government. But it also presents significant legal questions and potential liability, particularly in the area of intellectual property. In this article, we examine the nexus between big data and IP law, and consider how (or whether) IP law today can adequately protect investments in big data.

What is big data?

There is no universally agreed-upon definition of "big data". Most definitions offered by computer scientists and information experts reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data.

According to Gartner, big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making. This definition is commonly referred to as the three V's – volume, velocity, and variety. A fourth V frequently is included – veracity – reflecting the truth or accuracy of the data.

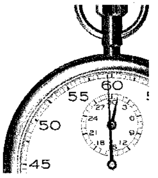
So, big data is about much more than large datasets: the term describes a technological and industrial phenomenon, spanning a variety of computing and analytical systems and processes.

But how will this big data phenomenon affect us? In May this year, the White House released a study examining "how big data will transform the way we live and work and alter the relationships between government, citizens, businesses, and consumers". This report noted:

> Aside from how we define big data as a technological phenomenon, the wide variety of potential uses for big data analytics raises crucial questions about whether our legal, ethical, and social norms are sufficient to protect privacy and other values in a big data world. Unprecedented computational power and sophistication make possible unexpected discoveries, innovations, and advancements in our quality of life. But these capabilities, most of which are not visible or available to the average consumer, also create an asymmetry of power between those who hold the data and those who intentionally or inadvertently supply it.

Ironically, the very sources of big data's grand potential also fuel the phenome-

One-minute read



Big data is here and growing. It presents great opportunities for business and government, but also poses significant legal questions, including issues to do with liability and IP protection. IP laws offer various solutions to protecting data, and its uses, though all have their drawbacks. Both patents and copyright are helpful in some respects, though trade secrets may be the most relevant tool. However, that presents a paradox as secret data and secret data processing create questions and concerns regarding the veracity of the data and its collection, processing and reuse. Ultimately, there may have to be other legal and regulatory solutions than classical IP rights to ensure that the potential of big data is fulfilled.

How big is big?

A zettabyte is 1,000,000,000,000,000,000,000 bytes, or 1 billion terabytes. One zettabyte could hold 323 trillion copies of Leo Tolstoy's 1,250-page War and Peace -"2016: The Year of the Zettabyte," Daily Infographic, March 23 2013.

Distorting data

Data - especially digital data - frequently is infused with subjective judgments of those who collect, organise and analyse it. As reported by the National Research Council: "Because digital data can be manipulated more easily than can other forms of data, digital data are particularly susceptible to distortion" (National Research Council, Ensuring the Integrity, Accessibility, and Stewardship of Data 34 (National Academies Press, 2009).

The privacy question

The privacy concern - which is socially and legally important - is beyond the scope of this article. Suffice to say that the reuse of big data has the potential - as recognised in the May 2014 White House Report - to "eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace". The critical question, of course, is whether those civil rights protections are expanded or diminished; that is, do individuals (and does society as a whole) benefit or suffer from the reuse of big data.

Meet Watson

IBM's Watson is not your run-of-the-mill computer. The size of 10 refrigerators, Watson is powered by 10 racks of IBM POWER 750 servers running Linux, and uses 15 terabytes of RAM, 2,880 processor cores and can operate at 80 teraflops - that is, 80 trillion operations per second. Totally self-contained, Watson scans the 2 million pages of content in its "brain" in less than three seconds. The question, however, is whether a patent to the super fast, efficient question-and-answer processing in Watson is still patent-eligible in view of statements from the Federal Circuit and the Supreme Court.

non's weaknesses. Big data suggests – indeed, accommodates – that huge volumes of data are created, collected, used and reused. To be effective, most big data advocates emphasise that the data must be shared, which includes the sharing of the means of creating, compiling and analysing such data. There are at least two major legal impediments to this advocated sharing, however: privacy and intellectual property. This article focuses on the nexus of intellectual property and big data.

How does intellectual property protect data generally?

To explore how intellectual property could be used to protect and enhance big data, we first examine how IP laws protect data, and associated methods of collecting, analysing, and reusing data, generally. (The unsatisfying answer: It varies.)

Although working with big data may yield patentable or copyrightable subject matter, for the most part, the law of trade secrets offers the most meaningful and robust protection from unwanted copying in this new technological domain. But therein lies the biggest impediment to truly drawing value from big data: when properly applied and complied with, trade secret protection keeps the subject matter *secret*. Yet, *secret* data and *secret* data processing inevitably create questions and concerns regarding the fourth V – the *veracity* of the data and its collection, processing, and reuse.

Patents

The patent system has historically been a poor protector of data and data processing, and recent case law suggests that it's even poorer today. Data *per se* is not patentable. Nor are collections of data. The US Patent Act states that only "new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof" qualifies for patent protection.

What about data compilations that are somehow manipulated, analysed and reused in innovative ways – the very epitome of big data? The case law suggests that merely assembling, organising or manipulating data is not itself eligible for patenting. Recently, the US Supreme Court announced the test for determining eligibility for patenting is a two-step process (*Alice Corporation Pty Ltd v CLS Bank International*). First, the court must determine whether the patent is directed to certain concepts held to be ineligible for patenting. Those are laws of nature, natural phenomena and abstract ideas. For example, mere ideas, fundamental economic practices, and mathematical formulas (algorithms) have been held to amount to no more than an abstract idea and thus not eligible for patenting.

In the second step, the court asks: "What else is there in the claims?" In other words, are there elements in the patent claim in addition to the patent ineligible concept itself that sufficiently transform the nature of the claim into patent eligible subject matter? As phrased by the Supreme Court, step two is a search for an "inventive concept".

In the decision announced in the Alice case at the Federal Circuit Court of Appeals opinion, Judge Lourie, joined by Judges Dyk, Prost, Reyna and Wallach, wrote: "At its most basic, a computer is just a calculator capable of performing mental steps faster than a human could. Unless the claims require a computer to perform operations that are not merely accelerated calculations, a computer does not itself confer patent eligibility." Thus, at least these Federal Circuit judges believe that using a computer to accelerate a process is not itself patent eligible. Nor is simply using a generic computer to perform conventional computer tasks (such as calculations on data).

Copyright

Copyright does not generally protect data itself; it may only, in certain circumstances, protect a *compilation* of data, such as in a database. The Copyright Act defines a "compilation" as:

> [A] work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship (17 USC § 101).

The Copyright Act also states that the copyright in a compilation extends only the compilation itself, and not to the underlying materials or data (17 USC § 103(b)). And, in order for a compilation of data to be protected, its selection, coordination, and arrangement must contain a modicum of originality – which was made clear in 1991 by the Supreme Court in *Feist Publications, Inc v Rural Telephone Service Company*.

As noted above, data itself is not copyrightable. Thus, even where the data compilation is copyrightable, one can extract individual datums from the compilation without violating the copyright laws. However, one cannot copy the *entire* database, since this would involve copying the entire protection expression (i.e., the compilation) provided it is selected, coordinated, and arranged in a creative, original way.

Trade secrets

All 50 US states provide some sort of trade secret protection, with the majority having adopted the Uniform Trade Secret Act (UTSA) in some way. Under the UTSA, a "trade secret" is defined as:

> information, including a formula, pattern, compilation, program device, method, technique, or process, that:
>
> i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and
>
> ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.

Trade secret law invariably prohibits the "misappropriation" of trade secrets. The UTSA defines "misappropriation" as:

> i) acquisition of a trade secret of another by a person who

Protecting a database

In *Feist Publications, Inc v Rural Telephone Service Company*, the leading case on copyright protection for data compilations, the Supreme Court announced that a compilation work such as a database must contain a minimum level of creativity to be protectable under the Copyright Act. Rural Telephone Service, a local telephone company, published telephone directories based on data from its subscribers. Feist used Rural's data to publish a "white pages" encompassing a much larger geographic area. Rural sued Feist for copyright infringement.

In its opinion, the Court rejected the "sweat of the brow" doctrine, which had provided copyright protection for compilations of data and other materials based simply on the effort used to create the compilation. The Court ruled that compilations and databases are protectable only when arranged and selected in an original manner. On the facts before it, the Court held that Rural's data compilation was not copyrightable since it did not meet the originality requirement. Specifically, Rural's selection of data (names, towns, and telephone numbers) was obvious and lacked creativity, since arranging names alphabetically in a directory was commonplace.

knows or has reason to know that the trade secret was acquired by improper means; or

ii) disclosure or use of a trade secret of another without express or implied consent by a person who

A) used improper means to acquire knowledge of the trade secret; or

B) at the time of disclosure or use knew or had reason to know that knowledge of the trade secret was

I) derived from or through a person who has utilized improper means to acquire it;

II) acquired under circumstances giving rise to a duty to maintain its secrecy or limit its use; or

III)derived from or through a person who owed a duty to the person seeking relief to maintain its secrecy or limit its use; or

C) before a material change of position, knew or had reason to know that it was a trade secret and that knowledge of it had been acquired by accident or mistake.

Under the USTA, "improper means includes theft, bribery, misrepresentation, breach or inducement of a breach of duty to maintain secrecy, or espionage through electronic or other means." In contrast, all 50 states provide that "reverse engineering" would be *proper* means.

The definition of "trade secret" would include data, data compilations and processes that compile, organise, manipulate or analyse data. The critical components to demonstrating that data, data compilations, or data processing are protectable trade secrets are that (a) there is *economic value* and (b) there are reasonable efforts to keep the subject matter secret.

Given that trade secret law expressly encompasses data compilations and processing and given the economic value that invariably derives from such compilations and processing, owners of data compilations and processing systems have generally been able to protect themselves merely by keeping it secret. Big data and big data processing likely fits well within the expansive definition of trade secret law.

Can big data thrive under today's IP legal framework?

The widespread excitement that big data has inspired should be tempered by the fact that secrecy, rather than disclosure, is the most powerful legal tool to protect many investments in this new arena. Data cannot be reused meaningfully on a large scale, after all, if there is insufficient information describing its provenance and pedigree. One of the authors of this article calls attention to this problem in a forthcoming article in the *Minnesota Law Review* ("Disclosing Big Data", volume 99, issue 2).

One might ask why private markets should not be expected to solve this problem. After all, if big data is truly the economic engine that experts say it is, then why wouldn't data producers feel compelled to disclose well-documented datasets? The answer has to do with the commercial context in which big data is developing. Smartphones, personal health devices, internet services and other primary sources of big data are provided by companies that typically aren't in the business of selling or licensing data. Rather, the data these devices produce is either an incidental byproduct, or a facet of a suite of services offered to consumers. Big data represents a secondary source of value that is external to the businesses of data producers. Because there is, as yet, no widespread market for abstract data, businesses have no economic motivation to relinquish their trade secrets.

On managingip.com

Senate trade secrets bill discussed (September 2014)

Five IP trends to track this year (blog, August 2014)

Alice influencing USPTO and Federal Circuit (July 2014)

Some might conclude that, unless our intellectual property system is somehow modified, the grand vision of big data will never be realised. A primary goal of our IP system is to incentivise technological disclosures, after all, and big data is currently channelled toward secrecy. Perhaps IP law should respond to this problem, but there is good reason to expect that any proposal to enact a new form of formal IP-like protection for data would fail politically. Since the mid-1990s, Congress has considered numerous proposals to enact *sui generis* IP protection for databases. Not one of these bills passed into law, in large part due to well-founded concerns that exclusive legal rights in data itself would frustrate scientific research and add incoherence to our legal system.

Intellectual property law isn't the only way to encourage technological disclosures, however. A piecemeal approach to the problem might ultimately be effective and helpfully tailored to different industries. The FDA might mandate rules about the data that medical device manufacturers must disclose, for instance. The FCC, meanwhile, could mandate its own data disclosure rules that pertain to communications devices and protocols. At the same time, the FTC, which is deeply interested in the implications big data holds for consumers, might seek to mandate greater disclosure from certain consumer-facing service providers.

It seems as though the future of big data is perpetually just around the corner. The public, lawmakers, and industry stakeholders would do well to explore new ways to discourage secrecy in this new domain. Until this important discussion begins, the full potential of big data may remain just out of reach.

Todd Vare

Michael Mattioli

© 2014. Todd Vare is a partner of Barnes & Thornburg in Indianapolis. Michael Mattioli is an associate professor of law at Indiana University Maurer School of Law in Bloomington