The expert mind in the age of junk data

Christopher L. Cosper*

Abstract: Architects bemoan the continuing decline of the authority of the profession, due to a myriad of social, economic, and technological changes. On the technological side, one of the most intriguing—and perhaps frightening—changes is the emergence of big data, or the use of massive data-sets to reveal information that could not be seen heretofore. Promising and problematic, big data has the potential to change our understanding of many fields, including architecture and urbanism. Starting with a historical look at the big data phenomenon, this paper argues that (1) much of what constitutes big data is actually junk data, (2) the expert mind of an architect is well positioned to extract the good data from the quickly expanding cosmos of junk data, and (3) big data has value for only two-thirds of the Vitruvian triad (i.e. commodity and firmness), requiring architects to guard against the misuse of big data to address issues of delight.

1. A history of big data

Although data experts Viktor Mayer-Schönberger and Kenneth Cukier argue that there “is no rigorous definition of big data,” they provide the following working definition: “[B]ig data refers to things one can do at a large scale that cannot be done at a smaller one.”

Some authors define big data in more absolute terms. In his book, focused on the type of predictive analytics used primarily for marketing and other business purposes, Steven Finlay defines big data problems as those problems which cannot be managed “on an average PC/laptop/server.”

ABOUT THE AUTHOR
Christopher L. Cosper is an assistant professor in the Architecture and Facility Management Program at Ferris State University, where he teaches environmental building systems, the history of sustainability, and design studio. Cosper holds a Bachelor of Architecture degree and a Master of Arts in English from Mississippi State University (1994 and 1996), and he is the first graduate of the Harvard University Graduate School of Design Master of Design Studies program to concentrate in Critical Conservation (2012).

PUBLIC INTEREST STATEMENT
More and more, our lives are being affected by the phenomenon of big data—or the use of massive data-sets to reveal information heretofore hidden from view. Much of the time, big data is innocuous, or even helpful, such as Amazon’s massive database providing useful reading recommendations. At other times, big data has a more sinister, more Big Brother quality, such as insurance companies using predictive analytics to determine health and automobile insurance rates, and police departments using similar information to predict areas of high crime (or, potentially, who might be a criminal).

Over the last few years, the power of big data has entered the world of architecture and urban design. Thus, architects and urban designers must learn how to use this new, powerful tool properly, extracting its promises while avoiding its potential dangers.
Finlay goes further by defining big data in absolute terms, saying that big data sources contain at least a terabyte (1,000 gigabytes) of data. Finlay also provides a more open-ended definition of big data, which this paper will address later in a subsequent section.

Big data is typically associated with computers, but that association is not required (Mayer-Schonberger & Cukier, 2013, pp. 14–15). In fact, two of the most intriguing uses of big data predate the computer.

In the middle decades of the nineteenth century, Matthew Fontaine Maury solved one of the world’s first big data problems by developing an understanding of prevailing winds and currents across the world’s oceans. His magnum opus, *The Physical Geography of the Sea*, utilized 1.2 million unique pieces of data and was intended

... to collect the experience of every navigator as to the winds and currents of the ocean, to discuss his observations upon them, and then to present the world with the results on charts for the improvement of commerce and navigation. (Maury, 1858)

The word every is critical, as it implies the $N =$ all quality of big data, where the goal is to collect every scrap of data available (Mayer-Schonberger & Cukier, 2013, p. 26). Despite Maury’s monumental achievement, he is an obscure figure.

A better known example of early big data is Dr. John Snow’s map of London cholera victims, chronicled in Steven Johnson’s bestselling book *The Ghost Map* (Figure 1). By drawing dashes on a map to indicate each death attributed to cholera, Snow was able to graphically demonstrate that the outbreak centered on a single water pump, solving the mystery of the disease by correlation during a historical period that predated our common understanding of the cause-and-effect-dependent germ theory of disease. Johnson argued that Snow’s map, however evocative, was secondary to the data—in this case, $N =$ all big data. Johnson wrote, “The real innovation lay in the data that generated that diagram, and in the investigation that compiled the data in the first place” (Johnson, 2006, p. 197) (Figure 2).

One of the clearest articulations of the potential power of big data was written by the science fiction author Isaac Asimov starting in the 1940s (Figure 3). Initially a series of short stories, Asimov’s big data thought experiment eventually became the *Foundation Series*, the first volume of which was published in 1951. Set in a distant future in which humanity has colonized the Milky Way, the *Foundation Series* is the story of a scientist, Hari Seldon, who creates “psychohistory—the science of human behavior reduced to mathematical equations” (Asimov, 1982, p. xi). Asimov wrote:

The individual human being is unpredictable, but the reactions of human mobs ... could be treated statistically. The larger the mob, the greater the accuracy that could be achieved. And the size of the human masses that Seldon worked with was no less than the population of all the inhabited millions of worlds of the Galaxy.

In Asimov’s tale, Seldon used really big data to predict the imminent fall of the Galactic Empire and an ensuing 30,000 years of chaos, so he created the titular Foundation to intervene and shorten the “Interregnum” to just 1,000 years—predictive analytics for the common good of the galaxy.

In 1944, librarian Fremont Rider looked into the future and saw chaos of a different sort—an ever-growing and soon-to-be unmanageable research library, which he explored in *The Scholar and the Future of the Research Library*. Much to the chagrin of shushing librarians everywhere, Rider’s work “exploded like an atomic bomb in the library profession” (Henkle & Lubetzky, 1946, pp. 280–284). Rider’s research showed that the collections in academic research libraries in the United States were doubling on an average of every 16 years, an exponential and unsustainable rate of growth (Rider, 1944, p. 3). About this increase in information, Rider wrote:
To many librarians it is already disconcerting, for they are watching a veritable tidal wave of printed materials yearly, monthly, daily, hourly, mount higher and higher.

Nor must it ever be forgotten that this is far more than a library problem ... it is a problem ... of civilization itself. (Rider, 1944, p. 3)

Rider proposed solving his big data problem with a series of microfilm cards, a system that relied on data compression.

In 1945, in the famous article “As We May Think,” engineer and scientist Vannevar Bush argued that data retrieval, not data storage, was the most pressing problem of the age. To address the
retrieval problem, Bush envisioned the creation of the “memex,” a machine which could both store and link data. Frustrated with the limitations of classification-based filing systems, Bush wrote:

The human mind does not work that way [i.e. by classification]. It operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain.

(Bush, 1945)

Bush is often credited with first imagining the structure of hypertext, an organizational principle essential for today’s World Wide Web.

The advent of commercially viable computers in the 1950s and 1960s changed the data storage landscape, providing both the compression of data Rider sought and the indexing ability Bush sought, but challenges remained. A 1967 report from the Rand Corporation and the University of California at Berkeley noted the limitations of data retrieval, which the authors “solved” by developing an intricate method of storing data (Levien & Maron, 1967, pp. 715–721). In retrospect, one can see that the discussed challenges were more the result of limited hardware than poor data coding. Humorously, from a twenty-first century perspective, the authors discussed “extremely large” data-sets that range from $10^4$ to $10^6$ bytes (or 100 kilobytes–1.0 megabytes) of information (Levien & Maron, 1967, p. 716)—data-sets which are one to 10 million times smaller than Finlay’s definition of a “big data” set.
Figure 3. Isaac Asimov’s Foundation Edge—predictive analytics to save the galaxy.

Source: Author.
By the end of the twentieth century, computer speed and storage capacity had improved to the point that the previously discussed problems had been largely solved, and computers invisibly ran many of the day-to-day transactions of modern life. Civilization was swimming in data, a situation which prompted researchers to ask new questions about the implications of all that data, including just how much data existed. Working in 1999 to establish the limits of data, Professors Peter Lyman and Hal R. Varian of the University of California at Berkeley estimated that between 0.7 and 2.1 exabytes of original content were created yearly (Lyman & Varian, 2000).

How big is an exabyte? It’s big (Table 1).

Today, when people commonly carry multiple-gigabyte “thumb drives” in their pockets, a person can easily forget just how much data a gigabyte represents. To visualize that amount of data, imagine that the volume of a Rubik’s Cube represents one byte of data. In such a scenario, the volume of the Empire State Building would represent approximately 5.6 gigabytes of data (Figure 4).

An exabyte is a billion gigabytes. Approximately two exabytes of original data are produced each year. And what will researchers do with all this data? Although today’s proponents of big data do not promise to save the galaxy à la Hari Seldon, they do promise big things. In a series of advertisements (Figure 5), IBM claims that predictive analytics can help

- Cities reduce crime by up to 30%.
- Universities reduce drop-out rates by 25%.
- Businesses make “smarter decisions.” (IBM, n.d.)

Big data has received big hype and “acquired the hyperbole that ‘big science’ did 50 years ago,” gracing the covers of numerous magazines and otherwise being covered in major media outlets (Borgman, 2015). The growth in big data is driven primarily by the availability of larger and larger data-sets, about which Finlay wrote:

> In recent years, there have been some improvements in the algorithms that generate predictive models, but these improvements are relatively small compared to the benefits of having more data, better quality data and analyzing this data more effectively. (Finlay, 2014, p. 7)

2. A treasure trove of big data or minefield of junk data?
Big data is predicated on more data being better data, but not all data are created equal. Much of what constitutes big data is actually junk data.

Suppose for a moment that a shepherd lives next to a minefield. What data about the minefield would be important to the shepherd?

If the shepherd wanted to avoid being killed by a mine, then the shepherd would only need to know the boundary of the minefield, which could be established by fencing and/or signage (Figure 6). This is a limited but extremely valuable set of data.

Presume, however, that the minefield is large and separates the shepherd from the local village, requiring the shepherd to cross the minefield to reach the village in a timely manner. At that point, the shepherd would need to know either a path through the minefield or the location of the individual mines (Figure 7).
Suppose one of the shepherd’s sheep wanders into the minefield and the shepherd wants to safely retrieve the sheep. The shepherd in this scenario would need to know the location of the mines or, perhaps, the triggering mechanism for the mines.

Perhaps an army engineering unit arrives with the mission of removing the mines. The sappers would need to know the boundary of the minefield to know the limit of their work, the location of the individual mines, the triggering mechanism of the mines, and the methods required to safely disable or detonate the mines.

Others may be interested in the minefield. An NGO may wish to investigate the minefield as an abuse of human rights, or social scientists may want to examine the impact of the minefield on the mental health of the shepherd and other villagers. These researchers and others could collect an almost unlimited amount of data concerning the minefield (Figure 8).

Back to the shepherd. If all the shepherd wants to do is avoid the minefield, all he needs to know is the boundary of the minefield—every other piece of information in the aforementioned thought experiment is junk data. The location of the individual mines, the location of a potential path, the triggering mechanism of the mines, the mechanism of mine disposal, who set the mines, the chemical and mechanical composition of the mines, the manufacturer of the mines, the impact of the mines on local villagers—all of these are junk data—to the shepherd. (That same junk data, however, might be valuable to an expert, the subject of a subsequent section.)
Figure 4. How big is a gigabyte? Compare the volume of a Rubik’s Cube to the volume of the Empire State Building.

Source: Image of Rubik’s Cube by Cubmundo (https://www.flickr.com/photos/cubmundo/6164861516) and modified by author; photo of empire state building by author.
Even big data enthusiasts recognize the predominance of junk data. Referring specifically to “forecasting consumer behavior,” Finlay acknowledged that “[a] huge proportion of the big data out there is absolutely useless” (Finlay, 2014, p. 16).

Good data can become junk data if it is unnecessary for the task at hand. However, some data is inherently junk data, with no value at all, because it has been falsified or otherwise corrupted.

A recent example of junk data can be seen in the Department of Veterans Affairs (VA) hospital scandal which surfaced in April 2014 (Bronstein & Griffin, 2014). Using a system apparently designed to induce fraud, VA administrators received bonuses if VA patients received prompt medical care. Thus, some managers and other VA employees created two wait lists—the official wait list and an unofficial wait list. Patients were kept on the unofficial wait list until they could receive care in the allotted time, making the official list look bonus-worthy. Of course, the official list was nothing but junk data.
Even when data is good, it might be overwhelming.

Dealing with big data—and wading through reams of junk data—can exhaust resources and patience. In a 1978 article about “energy education,” architectural educator Jeffrey Cook used the phrase “exploding information holocaust” to describe the “ever-expanding horizon of technical facts about all the ramifications of building” (Cook, 1978, p. 8). Since 1978, the amount of information just on Cook’s subject, designing energy efficient buildings, has multiplied beyond measure, and the challenge of effectively using that information has grown as well.

Furthermore, the problem with Cook’s “information holocaust” is twofold: not just the junk data that must be edited out, but the critical information that can be easily missed.
Six weeks after Hurricane Katrina, a series of design charrettes were held in Biloxi, Mississippi, to plan the reconstruction of the Mississippi Gulf Coast. Those charrettes, called the Mississippi Renewal Forum, involved more than 200 design professionals who essentially re-planned Coast in an intense, six-day session.

During the Mississippi Renewal Forum, planners gathered gigabytes if not terabytes of information: digital photos of before and after conditions, transcripts of interviews, hand sketches, copies of ordinances, and so forth. Despite the huge amount of data gathered, the team working on the Ocean Springs plan omitted this detail, from Section 22¾ Paragraph 22 of the municipal code:

Except as provided and permitted herein, it shall be unlawful to cut down, remove, deface, burn, poison, injure, mutilate, disfigure or substantially trim any tree identified herein as a **protected tree** … which has a trunk circumference of at least eighteen (18) inches when measured at a point five (5) feet above ground level … (The City of Ocean Springs, 2011)

Thus, Renewal Forum documents showing large developments in heavily wooded sections of Ocean Springs were an “utter fantasy” according to Eric Meyer, who became the city planner shortly after the completion of the forum (Meyer, 2011). One piece of data missed or ignored, and a public, published scheme is gutted.

Data become junk data when one of the following is true:

1. The data are unnecessary to the task at hand—e.g. manufacturing details of the mines for the shepherd who simply wants to avoid the minefield.
2. The data are inherently corrupt—e.g. Veteran’s Administration waitlists.
3. The data overwhelm the user’s ability to analyze them—e.g. Cook’s “information holocaust”.
4. The data gathered exclude critical information—e.g. the referenced tree ordinance.

Although big data is potentially problematic, many professionals are finding it too powerful to ignore.

### 3. Architecture is now a big data problem

Once, not so long ago, architecture was a small data problem. To design a building, an architect would develop at least a rudimentary program, perhaps by speaking with a client, studying precedents, and using rules of thumb. In the past, even relatively sophisticated programs could be crunched with a pencil and paper in a process that privileged intuition over information. Today, however, big data promises big changes. Mayer-Schönberger and Cukier wrote, “In the future—and sooner than we may think—many aspects of our world [that today are the sole purview of human judgment] will be augmented or replaced by computer systems” (2013, p. 12). For example, an energy model is inherently a big data exercise—hidden, yes, by the power of today’s desktop computers, but a big data exercise nonetheless.

Two major trends will likely ensure that small data architecture is a thing of the past: global climate change and increased client expectations of building performance.

The construction and operation of buildings are responsible for a large portion of global fossil fuel consumption. As noted many times, many ways, scientific consensus says that global climate change is real and is most likely the result of human activity, specifically the burning of fossil fuels. Many architects believe that climate change is the single most important issue facing the world today.

However, climate change is not the only issue facing architects, and a single-minded myopia will not serve the profession well. In this age of iPhones and BMWs, clients and the public at large have
the expectation that stuff will not just work, but work elegantly. Compared to the work of product and automotive designers, the work of architects often seems clumsy and unrefined. A long and continuing series of books, including *Form Follows Fiasco*, *From Bauhaus to Our House*, *How Buildings Learn*, and *Architecture of the Absurd*, pound architects for their incompetent handling of commodity and firmness (and delight, too, but that charge is harder to prove). Likewise, the trend toward design-build and other alternate owner–architect relationships can be read as a confirmation of a general dissatisfaction with the performance of architects.

The needs of today’s society require an expansion of the Vitruvian concept of firmness to include being both sustainable and technically competent: merely resisting gravity is not enough. Arguing in 2007 for a greater “research ethic” among architects, Stephen Kieran said “[t]o move the art of architecture forward ... we need to supplement intuition with science” (Kieran, 2007, p. 31).

Big data can be part of a more scientific approach to architecture, and if architects will not step up to the challenge, others will. IBM researcher Young M. Lee and his team noted that building management systems, sensors, and meters collect a huge amount of data, most of which is never read nor even kept. These researchers used data collected on a five-story office building to improve their predictive model of energy consumption (Lee et al., 2013, pp. 18-25).

Big data may be more of an attitude than a strict measurement of a data-set. Finlay said the following in his book on analytics:

Rather than getting hung up on a precise definition of big data, an alternate perspective is to view big data as a *philosophy about how to deal with data*, rather than how much data is available or what it contains. The four tenets of this philosophy are

(1) Seek.
(2) Store.
(3) Analyze.
(4) Act (Finlay, 2014, p. 7).7

This parallels Kieren’s call for architects to be engaged in a research loop of designing, building, monitoring, and learning from their projects—and then repeating the loop in a more informed state of mind (Kieran, 2007). The table below reconciles Finlay’s attitude toward data with Kieren’s philosophy of design (Table 2).

The opportunities for big data insights are multiplied on the scale of the urban, perhaps more so than one would imagine. For example, Dr. Steven Koonin, a theoretical physicist by training, was named the director of New York University’s Center for Urban Science and Progress, which is known as CUSP. Interviewed in the peer-reviewed journal *Physics Today*, Koonin said:

*Physics is an attitude as well as a subject. The kind of skills physicists bring to thinking through complicated situations, data driven and so on, are not all that common in urban science and technology at this point. Physicists have a lot to bring to the table here.* (Kramer, 2013)

An attitude, indeed. Of the 45 professionals who work for CUSP, only one has a design background (urban planning)—the rest are mathematicians, physicists, computer scientists, economists, and so forth (CUSP, 2015).

New York’s former mayor, Michael R. Bloomberg, is a data advocate (Bloomberg, 2015). CUSP is his creation, and it is designed to replicate previous successes with big data. During Bloomberg’s administration, for example, a city hall data team examined modified apartment buildings known as “illegal conversions.” Before big data, inspectors would find “high-risk conditions” only 13% of the
time. After big data sifted through a diverse set of inputs, inspectors found such conditions 70% of the time, making the inspections more than five times more effective (Lohr, 2013) (Figure 9).

Those results are impressive, but also disconcerting. In the spirit of Jane Jacobs, one wonders about the implications of this new generation of data-armed power brokers.

4. Is correlation enough?
Mayer-Schonberger and Cukier argue that in order to solve many problems, understanding causality is unimportant, while establishing correlation is critical (2013, pp. 50–72). Unlike the scientific method, which is based on theorizing if A, then B, big data shuffles through A, B, C ... Z to 10th power, etc., and sees what falls out.

If “correlation is enough” is the mantra of big data, one has to ask, “Correlation to what?” Debuting in 2008, the much ballyhooed Google Flu Trends used Google searches to predict flu patterns in the United States nearly instantaneously—or nearly two weeks faster than the Centers for Disease Control and Prevention could distribute their data. Regions with high rates of searches for “flu,” or “flu symptoms,” or so forth had highly correlated rates of flu. This big data success was published in a commonly cited 2009 article in *Nature* (Parry, 2014, p. A10).

Starting in 2011, however, Google Flu Trends went HAL 9000 and started to make mistakes, generally overestimating the rate of flu by a significant margin (Parry, 2014, p. A10). As it turns out, the correlation between Google searches for information on the flu and the flu itself was not as stable as the Google team thought.

The problem this represents for architects and urbanists is this: buildings and cities last longer than three years (the amount of time that Google Flu Trends worked before errors appeared). Google’s big data experiment can simply be tweaked before next year’s flu season; however, a building or urban design project built on irrelevant correlations remains built on irrelevant correlations.

For architects and urbanists, correlation may be more of a starting point than a goal. For example, access to daylight is correlated to higher test scores, but before flooding every elementary school with daylight, architects might want to understand the exact cause-and-effect relationship (Heschong, 2002). Similarly, William Whyte famously studied urban plazas and found a correlation between lunch-time use and linear feet of seating. However, when architects work on projects with social policy implications, correlation should be approached with extreme caution (Uprichard, 2014, pp. B14–B15). Causality is not going away. In an editorial that generally argued that “[c]orrelation is

| Step | Finlay—big data | Kieren—design process | Reconciled |
|------|-----------------|------------------------|------------|
| 1    | Seek the data   | A. Design the project  | Design based on data and with data collection in mind |
|      |                 | B. Build the project   |            |
| 2    | Store the data  | Monitor the project    | Monitor the project and store the data |
| 3    | Analyze the data| Analyze the data       | Analyze the data |
| 4    | Act on the data | Learn from the project | Act on the data to improve the current project and the next project |

Table 2. Finlay’s attitude toward data with Kieren’s philosophy of design

http://dx.doi.org/10.1080/23311886.2016.1198218
enough,” former Wired editor Chris Anderson acknowledged that “[d]ata without a model is just noise” (Anderson, 2008, p. 108).

5. The expert mind
Sifting through the noise of excessive or corrupted data is the purview of the expert mind.

What is an expert?

Writing in Scientific American, Philip Ross defines an expert as one who has “demonstrably superiority in skill over the novice” (Ross, 2006, p. 66).

Examining the work of many researchers, Ross discussed the game of chess as a way of exploring expertise. From a research standpoint, chess offers many advantages: chess is a game of pure logic with no component of luck, chess games result in a clear outcome (win/lose/draw), chess players are ranked statistically, and many of the most important games are recorded in full.

While intriguing, chess is an imperfect (or at least incomplete) platform for exploring expertise. A chess grandmaster is clearly an expert, but so is an architect, and an architect’s expertise includes realms not addressed in the game of chess. The possibilities in chess are infinite, but the game is structured, leading to common, named scenarios. Architecture has no parallel set of limitations. The win/lose/draw outcome of chess is clear, whereas the win/lose/draw outcome of architecture is highly subjective, arguably dependent on the vagaries of culture, taste, style, and zeitgeist.
Given the amount of research on expertise, surprisingly little research exists on the connection between expertise and creativity, and much of what does exist supports contradictory positions (Weisberg, 2006). This article subscribes to cognitive scientist Robert Weisberg’s arguments that A) a connection between creativity and expertise does exist and B) the development of expertise requires about a decade of self-conscious and/or directed work, a process which is often called the “Ten-Year Rule” (Weisberg, 2006).

Weisberg supports his thesis with a diverse series of case studies, looking at the creative output of Mozart, Thomas Edison, the Wright Brothers, Picasso, Alexander Calder, Jackson Pollock, Watson and Crick (who envisioned the double helix of DNA), and The Beatles. Weisberg wrote:

Creators—artists, for example—are usually not competing in a quantitative sense, as athletes [or chess players] are ... A better analogy might be to think of artists as explorers, each of whom takes a different path through heretofore unknown territory. (2006)

In a separate article, Weisberg examined Frank Lloyd Wright’s creative process for Fallingwater, concluding that the design of the iconic house was the result of Wright’s vast knowledge base and experience, not a flash of sudden inspiration (Weisberg, 2011, pp. 296-312). Weisberg concludes that experts are made over time by the purposeful exercise of their talents.

Reinforcing Weisberg’s argument that expertise is a necessary ingredient for creativity, Peter Eisenman said, “There can be no originality without a strong disciplinary authority to stand against it” (Eisenman, 2015). As described by Eisenman, the disciplinary authority is the master architect, the expert.

As an expert, an architect must reconcile numerous inputs and expectations. The idea of “balancing” a design is as old as the Vitruvian triad of commodity, firmness, and delight, which is the essence of the discipline. A developer can address commodity, an engineer can provide firmness, and a dilettante can create delight, but only an architect is trained to synthesize these often competing directives.

An architect is a special kind of expert. A good designer develops what G. Goetz Schierle calls “informed intuition,” which is akin to the knowledge that an apprentice builder would have encountered on an ancient construction site, but which can be simulated with physical models, digital tools, and other forms of representation (Schierle, 1997, p. 82).

An architect, like many other professionals, develops what Donald Schön calls “knowing-in-action,” “reflection-in-action,” and “reflecting on reflection-in-action,” abilities that allow the architect to make adjustments as he or she works, particularly when the process is disrupted by an unusual condition (Schön, 1995). Unlike decisions made from data collected in sterilized, laboratory conditions, decisions made by practitioners occur in the much messier milieu of practice. This is true for the so-called “major professions” including medicine, law, and business (Schön, 1987, p. 4).

Examining the difficulty of teaching professional “artistry,” Schön argues that architectural education—with its project-based learning, studio environment, and emphasis on intuition and creativity—is the paragon of professional education (Schön, 1987, p. 18). While many other academic disciplines have focused on “technical rationality” and the scientific method, architecture “crystallized as a profession before the rise of technical rationality and carries the seeds of an earlier view of professional knowledge” (Schön, 1987, p. 43).

Data alone cannot make a decision. In her examination of the scholarly implications of data, computer science and communications professor Christine Borgman wrote, “Data have no value or meaning in isolation” (2015, p. 4). How, then, do data move from isolation to relevance? By the
application of the expert mind—and the expert mind of an architect is well positioned to extract the good data from the quickly expanding cosmos of junk data.

Big data is a powerful tool, but as such, it can have unintended consequences, such as inhibiting the design process. Douglas Bowman, a designer who worked for Google, wrote the following about his resignation from Google:

Without a person at (or near) the helm who thoroughly understands the principles and elements of Design, a company eventually runs out of reasons for design decisions. With every new design decision, critics cry foul. Without conviction, doubt creeps in. Instincts fail. “Is this the right move?” When a company is filled with engineers, it turns to engineering to solve problems. Reduce each decision to a simple logic problem. Remove all subjectivity and just look at the data. Data in your favor? Ok, launch it. Data shows negative effects? Back to the drawing board. And that data eventually becomes a crutch for every decision, paralyzing the company and preventing it from making any daring design decisions.

... I had a recent debate over whether a border should be 3, 4 or 5 pixels wide, and was asked to prove my case. I can’t operate in an environment like that. (Bowman, 2009)

Design professors Jonathan Bean and Daniela Rosner identify three potential ways that big data can be misused as part of the design process:

(1) Crowdsourcing design decisions.
(2) Layperson design tools.
(3) Open-ended design “competitions” that obtain design services without a fair fee structure (Bean & Rosner, 2014).

Referring to product design specifically, Bean and Rosner wrote, “[T]he increasing use of crowdsourcing in design points to a dwindling role for the studio-trained designer as a creator of things” (Bean & Rosner, 2014, p. 19). Douglas Bowman’s post-Google screed provides anecdotal confirmation of this point: “Yes, it’s true that a team at Google couldn’t decide between two blues, so they’re testing 41 shades between each blue to see which one performs better” (Bowman, 2009).

Referring to design tools created for novices, Bean and Rosner wrote, “Such approaches to crowdsourcing paint the value of design as finite and knowable, and treat design as an element of a multivariate equation to be optimized” (Bean & Rosner, 2014, p. 18). A great example of this is BIM software that allows a novice to create a “design” with the initial appearance of sophistication and thought, an illusion, at least in student work, that is often burst by the inclusion of a terribly inappropriate default door or window.

Concluding their thoughts on crowdsourcing, Bean and Rosner said, “[In] the age of big data, the designer might become something else entirely. The manager of large systems? The interpreter of these data-generating events?” (Bean & Rosner, 2014, p. 18). In the context of the referenced article, this statement is pessimistic, as if the designer’s only role will be manager and interpreter. However, this could be turned into a positive, if the management of large systems and the interpretation of big data become part of a larger skill set, the skill set that designers already possess.

The success of architects with big data may be answered by who is leading whom, or perhaps who is being led by what. The key is returning the architect to a position of authority. As an anonymous architectural educator said, “Architects are most successful, in my mind, when they are able to view themselves as servants without relinquishing their responsibility for leadership” (Boyer & Mitgang, 1996, p. 141).
6. Big Data in the Curriculum

If big data is too important to ignore, then where should students learn it in architecture school? Although it might be tempting to place big data in a silo course, that temptation should be ignored, for two reasons.

First, material relegated to silo-based support courses appears unimportant to students, something other than “design.” In one educator’s view, “When technical, energy, and environmental issues are not deliberately brought into the studio course by faculty, the student’s model of a dualistic world of architecture is further reinforced” (DeKay, 1996). Placing big data content in studio courses will be challenging, of course, as the design studio sequence is already stuffed to the rafters with important content.

Second, the very nature of big data suggests integration. Looking for correlations requires an expansive approach.

The key to teaching students to work with big data is teaching them to understand the limits—not of the data, but of their understanding of the data. Expecting architecture students to develop some basic fluency, if not mastery, of the subject is reasonable. For example, architecture faculty do not attempt to create structural engineers in the architecture design studio, but they do expect students to understand column grids and the transfer of loads. Educators also hope to impart a basic structural vocabulary that will allow graduates to work productively with structural engineers.

What is commonly called research in the design studio is often a mad dash of architectural programming, an attempt to drink from a firehose of data in two weeks or so, and pretend to have established something substantive before returning to the warmer and fuzzier world of conceptual design. What is needed is more rigor in relationship to data, as Kieran advocated in 2007.

Fortunately, design studios need not work with big data per se to help students understand the implications of big data. That is true because

... big data is not necessarily better data. The farther the observer is from the point of origin, the more difficult it can be to determine what those observations mean—how they were collected; how they were handled, reduced, and transformed; and with what assumptions and what purposes in mind. (Borgman, 2015, p. xvii)

Just as most architects do not need to understand the mathematics of a moment diagram, most will not need to understand the mathematics behind a big data algorithm. What is important in both cases is being able to have a conversation with the people who do have the specialized expertise, communication skills again being fundamental for an architect, who is a generalist. At the end of the day, what the data mean is the critical point, and who is better positioned than an architect to see the big picture?

Successful architecture students inherently have many of the necessary qualities to address big data. In an article for IT professionals looking to work with big data, the following skill sets were emphasized: “a curious mind, the ability to communicate with nontechnical people, a persistent—even stubborn—character and a strong creative bent” (Harbert, 2013, p. 24). Sandeep Sacheti, a big data expert with a PhD in agricultural and resource economics, said that big data specialists must be “willing to do iterative design and [employ] agile thinking” (Harbert, 2013, p. 24). Many architecture students have the aforementioned skill sets, but those architecture students who want to go deeper into big data will need advanced mathematics, statistics, and perhaps computer programming skills.

Architecture schools are dabbling in big data—symposia, exhibits, and so forth. Given the increasing power and presence of big data, a more aggressive approach is required.
7. Big data, serendipity, and delight

Google Maps is a powerful tool—but it also is an engine designed to destroy serendipity. What happened to wandering through a city, being actually surprised by what lurked around the corner? Now, with Google Maps, a person can find where she wants to go—or thinks she wants to go—and Google’s algorithms will send her efficiently on her way, straight from A to B.

Likewise, Amazon’s website is a powerful tool. This author has certainly taken advantage of Amazon’s treasure trove of big data, which readily and sometimes accurately suggests that a person who enjoyed such-and-such book would also enjoy this-and-so. However, walking through a bookstore, particularly a disorganized second-hand book store, seems more organic, more connected, somehow more real. Amazon can help a person find books quickly and efficiently, but such purchases never feel quite as fulfilling as the more serendipitous variety, such as when this author spied the spiral galaxy cover of Asimov’s *Foundation Edge* on a bookshelf in a drugstore in Plymouth, Massachusetts, during an eighth-grade field trip—a chance encounter with big data some decades before his writing this paper.

One reason not to cede the design process to big data is the joy of the design process—a joy that comes from winnowing all the infinite possibilities of a project into the singular thing it becomes. During that process, some difficult, technical decisions have to be made, such as how much insulation to install, how much heat gain to accept from the west-facing windows, and where is the best location for the transit hub. For these and many other questions, big data can help. But as Douglas Bowman’s experience at Google vividly highlights, not being able to pick a shade of blue or the width of a border is not just discouraging but dehumanizing, leaving one to wonder, as Mayer-Schönberger and Cukier do, “What role is left for intuition, faith, uncertainty, acting in contradiction of the evidence, and learning by experience?” (Mayer-Schönberger & Cukier, 2013, p. 18).

Big data has value for only two-thirds of the Vitruvian triad (i.e. commodity and firmness); architects must guard against the misuse of big data to address issues of delight.

8. Conclusion

The N=all quality of big data has existed since the advent of data, but the power of big data was not immediately apparent. Nineteenth-century pioneers of big data, including Matthew Fontaine Maury and John Snow, used their data to solve seemingly intractable problems. With advancements in communication and data collection methods, larger data-sets became possible, but managing data rapidly became problematic. Mid-twentieth-century big data thinkers, including Isaac Asimov, Fremont Rider, and Vannevar Bush, imagined uses for data beyond the capabilities of the primitive computers of their era. Later, in the 1950s, 1960s, and 1970s, advances in computer hardware and software made the management and analysis of extremely large data-sets possible. By the end of the twentieth century, civilization was awash in data, with unique data being generated at the rate of two exabytes of data per year.

Not all data are equal, however. Given the enormous amount of data generated, even big data proponents concede that much of today’s big data could easily be reclassified as junk data. Sorting through the cosmos of junk data requires the skills of the expert mind. Architects and urbanists, experts in holistic thinking, are well positioned to extract good data from junk data, particularly in the context of the built environment.

For myriad reasons, architecture is now a big data problem, a condition acknowledged by the profession and academia. Receiving the Topaz Medallion in 2015 at the 103rd Annual Meeting of the Association of Collegiate Schools of Architecture, architect and educator Peter Eisenman said:

While there is a perceived lack of authorial presence in something called big data (which also threatens to be in itself an authorial presence), the digital proposes a panoply of originality rather than authority, which has little value in the critical other than for originality itself. But
in the end, there is always an authority figure who needs to choose. (Eisenman, 2015)\(^\text{10}\)

Amazon’s algorithms make some interesting recommendations, but a person—an expert in his or her own preferences—makes the final purchase decision.

When the decision involves the design of the built environment, would it not make sense for the decider—“the authority figure who needs to choose”—to be an expert in that environment, an architect or urbanist, perhaps?

In the 1940s, Isaac Asimov imagined the ultimate in big data, an algorithm that could predict the future of civilization. But Asimov also noted that “the individual human being is unpredictable.”

Today’s buildings and cities require certain elements that are predictable—namely those related to energy consumption and the other, scientifically measureable aspects of sustainability and performance. At the same time, they should be unpredictable: new, unique, differentiated because of site, social factors, and even whim.

And this is the entry point that big data provides architects—an opportunity to regain or relinquish authority, to fight against or succumb to junk data, to embrace design or surrender to those who would crowdsourse delight.

Funding
The author received no direct funding for this research.

Author details
Christopher L. Cosper\(^1\)
E-mails: cosperc@ferris.edu, ChristopherCosper@ferris.edu
\(^1\) Architecture & Facility Management, Ferris State University, 915 Campus Drive, SWAN 314, Big Rapids, MI, USA.

Citation information
Cite this article as: The expert mind in the age of junk data, Christopher L. Cosper, Cogent Social Sciences (2016), 2: 1198218.

Notes
1. Emphasis added.
2. Mayer-Schonberger and Cukier attribute Maury’s obversity to his serving as a Confederate spy during the American Civil War.
3. Curiously, Asimov’s prescient stories are noted by many big data bloggers but typically overlooked by big data scholars.
4. I was led to this article and the Lyman and Vairian article by Gil Press’s online article “A Very Short History Of Big Data” http://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/ (accessed July 15, 2015).
5. Emphasis added.
6. Order of sentence changed for clarity.
7. Capitalization altered and emphasis added.
8. The nature-versus-nurture debate raised in Ross’s article is beyond the scope of this paper.
9. Emphasis in original.
10. Emphasis added.

Correction
A small number of minor typographical errors have been corrected, but the meaning of the text has not been altered.

References
Anderson, C. (2008, July). The end of theory. Wired, 16, 108–109.

Asimov, I. (1982). Foundation’s edge. New York, NY: Del Rey.
Bean, J., & Rosner, D. (2014). Big data, diminished design? Interactions, 21, 18–19.
Bloomberg, M. R. (2015, April 20). Why I’m betting on cities and data. Retrieved from www.huffingtonpost.com; http://www.huffingtonpost.com/michael-bloomberg/bloomberg-what-works-cities-data-b_7097304.html
Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. Cambridge, MA: MIT Press.
Bowman, D. (2009, March 20). Stopdesign. Retrieved July 19, 2015, from http://stopdesign.com/archive/2009/03/20/goodbye-google.html
Boyer, E. L., & Mitgang, L. D. (1996). Building community: A new future for architecture education and practice. Princeton, NJ: The Carnegie Foundation.
Bronstein, S., & Griffin, D. (2014, April 23). CNN.com. Retrieved July 25, 2015, from http://www.cnn.com/2014/04/23/health/veterans-dying-health-care-delays/index.html
Bush, V. (1945, July). As we may think. The Atlantic. Retrieved July 27, 2015, from http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/
Cook, J. (1978). Thinking about energy education. Journal of Architectural Education, 31, 8–10. http://dx.doi.org/10.1080/10464883.1978.10758134
CUSP. (2015, July 28). Retrieved from cusp.nyu.edu/http://cusp.nyu.edu/

DeKay, M. (1996). Possible evolution of technology education in architecture: The case of energy and environmental issues. Constructions for Tectonics for the Postindustrial World (pp. 172–175). Washington, DC: Association of Collegiate Schools of Architecture.
Eisenman, P. (2015, March 21). Originality and authority. Retrieved from https://www.youtube.com/watch?v=FNmVIY3HYY
Finlay, S. (2014). Predictive analytics, data mining and big data. New York, NY: Palgrave Macmillan.
http://dx.doi.org/10.1057/9781137379283
Harbert, T. (2013, January 14). Big data, big jobs? Computerworld, 47, 23–27.
Henkle, H. H., & Lubetzky, S. (1946). The scholar and the future of the research library (Book Review). College and Research Libraries, 280–284. http://dx.doi.org/10.5860/cri_07_03_280
Heschong, L. (2002). Daylighting and human performance. ASHRAE Journal, 65–67.
IBM. (n.d.). Retrieved July 28, 2015, from https://www.youtube.com/user/IBM/featured

Johnson, S. (2006). The ghost map: The story of London’s most terrifying epidemic—and how it changed science, cities, and the modern world. New York, NY: Riverhead Books.

Kieran, S. (2007). Research in design. *Journal of Architectural Education*, 61, 27–31. http://dx.doi.org/10.1111/j.1531-314X.2007.00125.x

Kramer, D. (2013). Smart cities will need big data. *Physics Today*, 66, 19–20.

Lee, Y. M., An, L., Liu, F., Horesh, R., Chae, Y. T., & Zhang, R. (2013). Applying science and mathematics to big data for smarter buildings. *Annals of the New York Academy of Sciences*, 1295, 18–25. http://dx.doi.org/10.1111/nyas.2013.1295.issue-1

Levien, R., & Maron, M. E. (1967). A computer system for inference execution and data retrieval. *Communications of the ACM*, 10, 715–721. http://dx.doi.org/10.1145/363790.363817

Lohr, S. (2013, February 23). SimCity, for real: Measuring an untidy metropolis. *The New York Times*.

Lyman, P., & Varian, H. R. (2000). How much information. *Journal of Electronic Publishing*. Retrieved July 20, 2015. doi:10.3998/3336451.0006.204

Maury, M. F. (1858). *The physical geography of the sea*. New York, NY: Harper & Brothers. Retrieved July 26, 2015, from http://name.umdl.umich.edu/AJA7085.0001.001

Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A Revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.

Mayer-Schonberger, V., & Cukier, K. (2013). Big data: A Revolution that will transform how we live, work, and think. New York, NY: Houghton Mifflin Harcourt.

Meyer, E. (2011, July 14). (Author, Interviewer).

Parry, M. (2014, April 11). Recent big-data struggles are ‘birthing pains,’ researchers say. *The Chronicle of Higher Education*, p. A10.

Rider, F. (1964). The Scholar and the future of the research library. New York, NY: Hadham Press. Retrieved July 25, 2015, from http://hdl.handle.net/2027/mdp.3901503591101

Ross, P. E. (2006). The expert mind. *Scientific American*, 295, 64–71. http://dx.doi.org/10.1086/scientificamerican0806-64

Schieler, G. G. (1997). The pedagogy of architectural technology. *Journal of Architectural Education*, 51, 82–83.

Schön, D. A. (1987). *Educating the reflective practitioner*. San Francisco, CA: Jossey-Bass.

Schön, D. A. (1995, November) The new scholarship requires a new epistemology. *Change*, (pp. 26–34).

The City of Ocean Springs. (2011, August 16). Ocean Springs Municipal Code. Retrieved July 27, 2015, from Municode: https://www.municode.com/library/ms/ocean_springs/codes/code_of_ordinances?nodeId=MUCO_CH22_3-4TRSH

Uprichard, E. (2014, October 17). Big-data doubts. *The Chronicle of Higher Education*, 61, B14–B15.

Weisberg, R. W. (2006). Modes of expertise in creative thinking: evidence from case studies. In K. A. Ericsson, N. Charness, & P. J. Feltovich (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 761-788). Cambridge, MA: Cambridge University Press.

Weisberg, R. W. (2011). Frank Lloyd Wright’s Fallingwater: A case study in inside-the-box creativity. *Creativity Research Journal*, 23, 296–312. http://dx.doi.org/10.1080/10400419.2011.621814