World Congress on Water, Climate and Energy
The Convention Centre
Dublin, Ireland
May 13-18, 2012

*Plenary Keynote Address*

# Big Data Analytics: Disruptive Technology in the Water Industry?

**Paul R. Brown, AICP**
**Executive Vice President**
**CDM Smith**

Over the last several years, I have been introduced to a realm of science and technology that many of us in the water industry don't know much about – and really should pay closer attention to. In fact, it is a field of new science that barely has a name. Let's just call it "big data analytics;" and let's focus on how it can make our cities "smarter."

## New Science without a Name

It's interesting that when The Data Warehousing Institute (TDWI) surveyed data management professionals on this topic in May 2011, 7% of respondents hadn't "seen or heard of anything resembling big data analytics" – they knew nothing about it. Most respondents (65%) didn't have a name for it but generally understood its meaning. The remaining 28% (roughly a quarter) both understood the concept and had named it – most calling it "big data analytics" but including names as well, like "advanced" analytics, "discovery" analytics, or "exploratory" analytics.

And while twenty years ago, I/T professionals used to struggle with the cost of data storage and management, today the storage of massive amounts of data is virtually free, providing for increasingly sophisticated approaches to mining it, analyzing it, discovering relationships within it, and ultimately utilizing it to predict the behavior of the complex systems (and systems of systems) it represents. This is an emerging technology which will no doubt touch every aspect of our lives (and already has if you shop with a credit card or over the internet). But for our purposes it's the management of urban services and utilities we're interested in.

## What We're Already Doing

Let me pause and make a distinction here between (1) what we are already doing with data collection, and (2) the new science of big data analytics that I am introducing today. In the water industry, like all other utilities, we monitor and evaluate all kinds of data regarding the operations and performance of our physical infrastructure and mechanical systems. Supervisory control and data acquisition (SCADA) systems continuously report-out data and inform our operators regarding what's happening and where it's happening throughout these process facilities. They are generally designed by instrumentation and control engineers; and meet objectives around operations and reporting.

In a related context, we have also invested in a variety of asset management tools that store large amounts of data to assist with the maintenance, repair and replacement of system components and equipment.

And on the customer side, we generally (not always) meter water used, primarily for billing purposes. As an industry, we are making progress in deploying automatic meter reading (AMR) from vehicles rather than walking meter-to-meter; and advanced metering infrastructure (AMI), where meter readings are transmitted to an operations center at regular intervals – improvements that, frankly, electric utilities have been more effective in implementing.

With respect to direct communication with customers, our invoices are increasingly informative. However, our feedback and interaction with customers is, again, lagging behind the progress made among electric utility colleagues. It's safe to say that with all the progress we've made, we still have a long way to go on the metering side. As one executive in the water industry put it, "The average [water] meter being sold today is century-old technology wrapped in a different box."

But for our purposes these are merely sources of data that feed our emerging and insatiable universe of big data analytics. The breakthroughs that I want to discuss relate to how those data streams, combined with data from many others sources, can be structured and analyzed for entirely new insights and outcomes. In IBM's white paper "Smarter Cities on a Smarter Planet," they put it this way:

> "In the past, our data models on the future predicted long-term trends over days, weeks, and years. Now, because we can manage so much data, and we can process it so fast, we can look at traffic and predict what it will look like in an hour. We can look at weather patterns that are coming in and predict where we are likely to have fallen power lines within a square kilometer."

As an aside, of course our current SCADA systems will tell us when those lines are actually down.

By collecting data from throughout our urban systems, looking for statistical patterns in that data that suggest significant correlations and trends; forecasting probable problems and events, and ultimately exploiting that data to optimize the deployment and coordination of resources offers an entirely new framework for managing cities and their infrastructure.

At Microsoft, their researchers describe this as a whole new model "for science based on data-intensive computing." In their view, its significance is analogous to the invention of the printing press in its potential impact. A new science based on the collecting, curating, and analyzing incomprehensibly large amounts of data from all sources. What has been referred to by some as a data tsunami.

And finally, when the 2011 World Economic Forum's Council on Emerging Technologies identified trends that would have the greatest impact on the state of the world in the near future, at the top of their list they identified "informatics for adding value to information." They put it this way:

> "The quantity of information now available to individuals and organizations is unprecedented in human history, and the rate of information generation continues to grow exponentially. Yet, the sheer volume of information is in danger of creating more noise than value, and as a result limiting its effective use. Innovations in how information is organized, mined and processed hold the key to

> filtering out the noise and using the growing wealth of global information to address emerging challenges."

How will this revolution in the way we understand and manage our world affect what we do?

## Big Data Analytics is a Disruptive Technology

For many of us in the water industry this rapidly expanding field of inquiry and applications feels like what Clayton Christensen, in his groundbreaking book *The Innovator's Dilemma,* would call "disruptive technology." And while Christensen's framework was primarily product-focused and based on the market economics of competing corporations, many of his lessons apply. Christensen broadly categorizes technological innovations as either: (1) sustaining technologies, or (2) disruptive technologies.

In the first category, "sustaining technologies" are those that well-run enterprises invest in to continually "improve the performance of established products, along the dimensions of performance that mainstream customers have historically valued."

On the other hand, "disruptive technologies" are those that:

> "Bring to a market a very different value proposition than had been available previously. Generally, disruptive technologies underperform established products in mainstream markets. But they have other features that a few fringe (and generally new) customers value. Products based on disruptive technologies are typically cheaper, simpler, smaller, and, frequently, more convenient to use."

Fundamentally, "big data analytics" is a very different value proposition. The approach to modeling that is embedded in most big data analytics is based on the mathematics of statistics – not the first principles of physics. It's meteorology not mechanical engineering. It is driven by Moore's Law of exponentially increasing computer processing power, not the theoretical science of Newton's Laws of Motion or Maxwell's equations. It is a practical, cheaper, and in many respects simpler way of understanding and managing highly complex systems of systems. In fact, it may be the only way to do so.

In know from discussions with my colleagues, that for many engineers and scientists, it appears to "underperform" when compared to traditional engineering analysis and modeling best practices. It clearly over-performs if rapid adaption to real-time events affecting many city functions is the imperative.

## Thousands of Years of Sustaining Innovations

Our success as an industry has been attributable largely to our slow-and-steady progress in bringing to market the next generation of sustaining technologies, and that linear path from pre-Roman times has been remarkably consistent its trajectory to the present day. In fact, it has only been in the last several decades that we have reexamined the fundamental purposes of our water infrastructure and attempted to drive sustaining technology towards what my colleague Vladimir Novotny describes as the Fifth Paradigm of water management – a transition from rapid conveyance linear systems for water, wastewater, and drainage to more integrated, closed-loop approaches.

But now, many of us struggle with how to fully integrate our activities with the other realms of urban infrastructure and urban life in an increasingly unpredictable natural environment. At CDM Smith, we have devoted the last several years to the development of processes and tools that enable us to simulate the behavior of urban design as it impacts all aspects of water, transportation, energy, buildings, materials, and the natural environment. With support from both the Singapore Economic Development Board and the Housing Development Board, we have been working at our Neysadurai Centre to introduce an entirely new "smarter" approach to the process of planning and designing smart cities.

Big data analytics offers the data management, exploratory analytics, and data visualization tools we need to discover the important behavioral characteristics of highly-complex urban infrastructure systems that we study through computer simulations. It's in the nature of complex systems to behave unexpectedly and it is only through analytics applied to output that discoveries are made.

## Our Interest in Complex Systems of Systems

This element of discovery and surprise is intrinsic to the breakthroughs we are hearing about from IBM and others who are boldly confident that they can discover insights and improved performance through the collection and analysis of terabytes of real-time data received from sensors and sources previously unmanageable and/or unavailable. They don't know what they will find – but they are confident that their discoveries will radically improve our ability to adapt in a world where we are confronted with a combination of population growth, economic development, and resource consumption that has many brilliant thinkers declaring the world is full and we're done for.

As the breakthroughs in this new science reach into all of our organizations and enterprises, it opens up opportunities for solutions that we have not had in the past.

## The Opportunities Ahead

Let's look at what those opportunities are. The most obvious include:

***Better real-time operations decision-support.*** The most immediate and obvious use for advanced analytics is in the operations of our existing infrastructure systems. It requires us to continue to advance our asset management capabilities and increase the amount of advanced metering infrastructure being deployed in our water utilities. As this data becomes available, predictive maintenance, leak detection, energy optimization all becomes more efficient and effective.

***Improved customer relationships and communications.*** At the same time that we learn more about how and when water is used by customers we also have an opportunity to increase our communications and interaction with them. The feedback that we can provide to customers, combined with incentives for water use efficiency, offer a more dynamic and effective means of incorporating demand management into our resource mix.

In fact, as I mentioned, in these two areas electric utilities have progressed considerably further than those of us in the water industry, and we have much to learn from their experiences developing smart grids. But there are some other important areas that will benefit from this new data-intensive science. Consider:

***Real-time feedback from functional landscape designs*** (the truly green components of our integrated hybrid infrastructure solutions). As we have discussed during many of our IWA Cities of the Future conferences, integrated water management will rely more heavily on landscape that provides functional services to our communities, including less impervious surface area, more green space used as buffers and for groundwater recharge – all restoring both hydrological and ecological functions.

We have always known, for example, that trees, are major contributors to stormwater management and urban cooling – reducing energy costs and decreasing the amount of urban runoff reaching receiving waters. At the same time, tree planting is often seen as a supplemental benefit rather than a replacement for the traditional stormwater infrastructure characterized by pipes and pumps. Why? Well, do we know they are being maintained? Are they healthy? Are they even there?

Increased monitoring and sensors in the natural environment can give us the data we need to see that our natural systems are reliably doing their part in the urban environment. It would perhaps give us the confidence to substitute those green investments for more expensive expenditures on traditional hardware.

Thinking more creatively, can we engage citizens to serve as our eyes and ears, acting as volunteer caretakers for the green solutions we have provided in our cities? Since so many of the improvements are micro-scale changes in the design and functionality of system components, maintenance and repair will remain a challenge. Can we develop smart phone apps that allow volunteers to identify, photograph, and report instances of green features in need of repair? And finally:

***Better design of urban hybrid systems.*** As planners, designers, and engineers that combine small-scale distributed technologies, functional green landscape, high-performance building technology, and our backbone centralized utility grids we can benefit immensely from the analytics applied to urban systems data. Through simulation and analysis, like we are currently undertaking in Singapore, we can give planners and designers insights into their design solutions during the development process.

Throughout all of this, big data analytics will drive more institutional integration than we have ever imagined. Just look at the intelligent operations center that Rio de Janeiro has implemented in collaboration with IBM for example. If we don't integrate all of the data derived from our management of urban watersheds including, both the man-made and the natural infrastructure that supports us, it will no doubt be integrated by others.

I welcome you to join us on this journey of discovery. We have great examples of early explorations into the complex systems comprising our urban infrastructure. It becomes especially complex when we go beyond the physics of hydraulics and electricity to include the natural systems of climate and landscape, and finally when we embrace the human behavior and ingenuity of the people we serve as utilities. We can work with our customers to better understand the integrated nature of our urban systems and enlist them as our allies as we adapt in the face of new challenges. It is important because it offers a path to greater resilience in the context of an increasingly unpredictable climate and economies.

As we evolve from highly instrumented infrastructure to greater degrees of integration and data-intensive analytics, we can envision a time where the speed of the feedback, the reliability of the near-term predictions, and the range of operational flexibility allows our complex urban infrastructure to function in truly smart ways. That is the message we are hearing from IBM and others here and now.

So, as unsettling as it may be (and in many ways it is unsettling), I welcome the possibilities created by the disruptive innovations associated with big data and advanced analytics. I want to be a citizen in smart cities that we will design and create far more intelligently. And I hope you do as well.