Deloitte.
University Press

Australian edition

Tech Trends 2017
The kinetic enterprise
The Australian Cut

Dark analytics

Illuminating opportunities hidden within unstructured data

ACROSS ENTERPRISES, EVER-EXPANDING STORES OF DATA REMAIN UNSTRUCTURED and unanalysed. Few organisations have been able to explore nontraditional data sources such as image, audio, and video files; the torrent of machine and sensor information generated by the Internet of Things; and the enormous troves of raw data found in the unexplored recesses of the "deep web." However, recent advances in computer vision, pattern recognition, and cognitive analytics are making it possible for companies to shine a light on these untapped sources and derive insights that lead to better experiences and decision making across the business.

IN this age of technology-driven enlightenment, data is our competitive currency. Buried within raw information generated in mind-boggling volumes by transactional systems, social media, search engines, and countless other technologies are critical strategic, customer, and operational insights that, once illuminated by analytics, can validate or clarify assumptions, inform decision making, and help chart new paths to the future.

Until recently, taking a passive, backward-looking approach to data and analytics was standard practice. With the ultimate goal of "generating a report," organisations frequently applied analytics capabilities to limited samples of structured data siloed within a specific system or company function. Moreover, nagging quality issues with master data, lack of user sophistication, and the inability to bring together data from across enterprise systems often colluded to produce insights that were at best limited in scope and, at worst, misleading.

Today, CIOs harness distributed data architecture, in-memory processing, machine learning, visualisation, natural language processing, and cognitive analytics to answer questions and identify valuable patterns and insights that would have seemed unimaginable only a few years ago. Indeed, analytics now dominates IT agendas and spend. In Deloitte's *2016 Global CIO Survey* of 1,200 IT executives, respondents identified analytics as a top investment priority. Likewise, they identified hiring IT talent with analytics skills as their top recruiting priority for the next two years.[1]

Leveraging these advanced tools and skill sets, over the next 18 to 24 months an increasing number of CIOs, business leaders, and data scientists will begin experimenting with "dark analytics": focused explorations of the vast universe of unstructured and "dark" data with the goal of unearthing the kind of highly nuanced business, customer, and operational insights that structured data assets currently in their possession may not reveal.

In the context of business data, "dark" describes something that is hidden or undigested. Dark analytics focuses primarily on raw text-based data that has not been analysed—with an emphasis on unstructured data, which may include things such

as text messages, documents, email, video and audio files, and still images. In some cases, dark analytics explorations could also target the deep web, which comprises everything online that is not indexed by search engines, including a small subset of anonymous, inaccessible sites known as the "dark web." It is impossible to accurately calculate the deep web's size, but by some estimates it is 500 times larger than the surface web that most people search daily.[2]

In a business climate where data is competitive currency, these three largely unexplored resources could prove to be something akin to a lottery jackpot. What's more, the data and insights contained therein are multiplying at a mind-boggling rate. An estimated 90 percent of all data in existence

today was generated during the last five years.[3] The digital universe—comprising the data we create and copy annually—is doubling in size every 12 months. Indeed, it is expected to reach 44 zettabytes (that's 44 *trillion* gigabytes) in size by 2020 and will contain nearly as many digital bits as there are stars in the universe.[4]

What's more, these projections may actually be conservative. Gartner Inc. anticipates that the Internet of Things' (IoT) explosive growth will see 20.8 billion connected devices deployed by 2020.[5] As the IoT expands, so will the volumes of data the technology generates. By some estimates, the data that IoT devices will create globally in 2019—the bulk of which will be "dark"—will be 269 times greater than the amount of data being transmitted

Figure 1. The expanding digital universe, 2013–2020

In 2020, the digital universe is expected to reach 44 zettabytes. One zettabyte is equal to one billion terabytes. Data valuable for enterprises, especially unstructured data from the Internet of Things and nontraditional sources, is projected to increase in absolute and relative sizes.
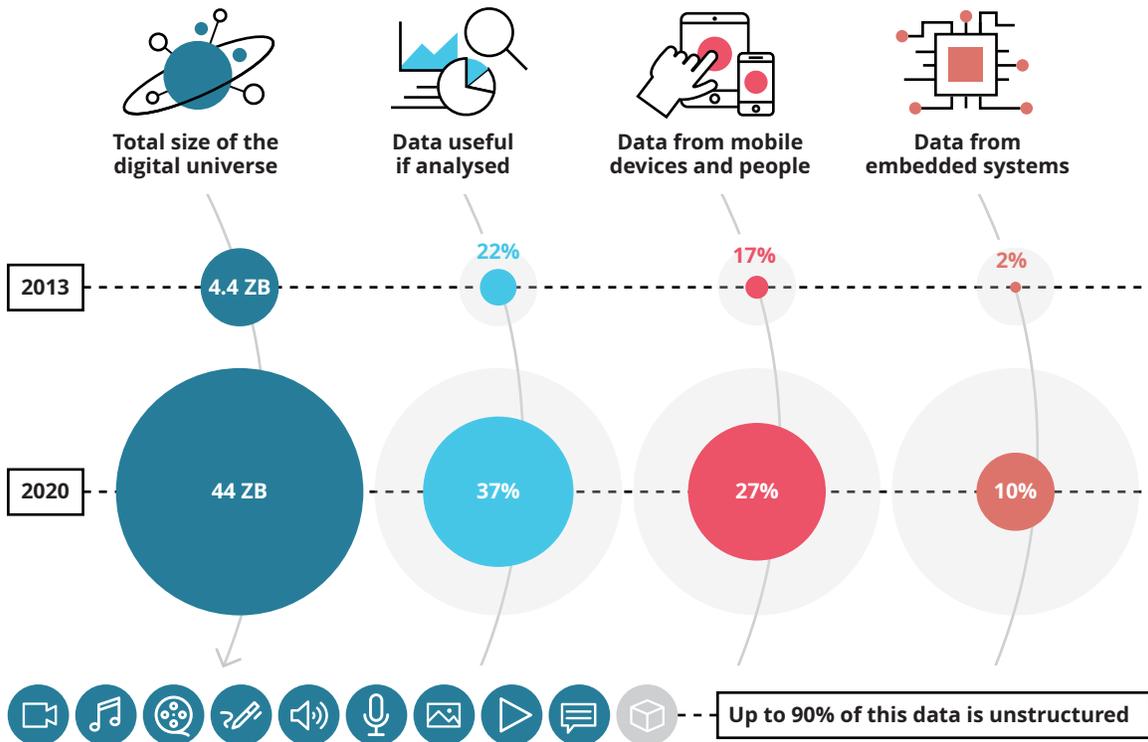
Sources: EMC Digital Universe with research and analysis by IDC, "The digital universe of opportunities: Rich data and the increasing value of the Internet of Things," April 2014; International Data Corporation, "IDC iView: Extracting value from chaos," 2011, www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf, accessed December 29, 2016.

Deloitte University Press | dupress.deloitte.com

to data centres from end-user devices and 49 times higher than total data-centre traffic.[6] Against this statistical backdrop, big data, as a business imperative, might be more accurately described as "enormous data."

To date, companies have explored only a tiny fraction of the digital universe for analytic value. IDC estimates that by 2020, as much as 37 percent of the digital universe will contain information that might be valuable if analysed.[7]

But exactly how valuable? IDC also projects that organisations that analyse all relevant data and deliver actionable information will achieve an extra $430 billion in productivity gains over their less analytically oriented peers by 2020.[8]

Let there be light

When we think about analytics' potential, the possibilities we often envision are limited to the structured data that exists within the systems around us. Dark analytics seeks to remove those limits by casting a much wider data net that can capture a corpus of currently untapped signals.

Dark analytics efforts typically focus on three dimensions:

Untapped data already in your possession: In many organisations, large collections of both structured and unstructured data sit idle. On the structured side, it's typically because connections haven't been easy to make between disparate data sets that may have meaning—especially information that lives outside of a given system, function, or business unit. For example, a large insurance company mapped its employees' home addresses and parking pass assignments with their workplace satisfaction ratings and retention data. The effort revealed that one of the biggest factors fueling voluntary turnover was commute time—the combination of distance to the office, traffic patterns based on workers' shift schedule, degree of difficulty in finding a parking spot, and length of walk from car to their workspace.

Regarding "traditional" unstructured data, think emails, notes, messages, documents, logs, and notifications (including from IoT devices). These are text-based and sit within organisational boundaries but remain largely untapped, either because they don't live in a relational database or because until relatively recently, the tools and techniques needed to leverage them efficiently did not exist. Buried within these unstructured data assets could be valuable information on pricing, customer behavior, and competitors. Particularly in multinational companies, they may also contain potentially valuable yet untranslated data assets created for or generated in non-English-speaking markets.

What percentage of the data in existence today is unstructured? No one knows for sure. The generally accepted figure has long been 80 percent—known as "the 80 percent rule"—though recent estimates put the number closer to 90 percent.[9]

Nontraditional unstructured data: The second dark analytics dimension focuses on a different category of unstructured data that cannot be mined using traditional reporting and analytics techniques—audio and video files and still images, among others. Using computer vision, advanced pattern recognition, and video and sound analytics, companies can now mine data contained in nontraditional formats to better understand their customers, employees, operations, and markets. For example, a retailer may be able to gain a more nuanced understanding of customer mood or intent by analysing video images of shoppers' posture, facial expressions, or gestures. An oil and gas company could use acoustic sensors to monitor pipelines and algorithms to provide visibility into oil flow rates and composition. An amusement park could gain greater insight into customer demographics by analysing security-camera footage to determine how many customers arrive by car, by public transportation, or by foot, and at what times during the day.

Low-cost high-fidelity surround cameras, far-field microphone areas, and high-definition cameras make it possible to monitor all business activities taking place within an enterprise. The ability to apply analytics to audio and video feeds in real time opens up profound new opportunities for signal detection and response. Such digital cues offer new ways of answering existing questions and exploring

new opportunities. Moreover, in recent years data storage costs have declined by an estimated 15 to 20 percent, making the archiving of image and audio files a more realistic option for smaller organisations.[10]

Data in the deep web: As a dimension of dark analytics, the deep web offers what may contain the largest body of untapped information—data curated by academics, consortia, government agencies, communities, and other third-party domains. But the domain's sheer size and distinct lack of structure can make it difficult to search. For now, only data mining and analytics efforts that are bounded and focused on a defined target—for instance, licensable data owned by a private association—will likely yield relevant, useful insights. Just as the intelligence community monitors the volume and context of deep web activity to identify potential threats, businesses may soon be able to curate competitive intelligence using a variety of emerging search tools designed to help users target scientific research, activist data, or even hobbyist threads found in the deep web. For example, Deep Web Technologies builds search tools for retrieving and analysing data that would be inaccessible to standard search engines.[11] Its software is currently deployed by federal scientific agencies as well as several academic and corporate organisations. Stanford University has built a prototype engine called Hidden Web Exposer that scrapes the deep web for information using a task-specific, human-assisted approach. Other publicly accessible search engines include Infoplease, PubMed, and the University of California's Infomine.[12]

Flashlight, not interplanetary star

To be clear, the purpose of dark analytics is not to catalog vast volumes of unstructured data. Casting a broader data net without a specific purpose in mind will likely lead to failure. Indeed, dark analytics efforts that are surgically precise in both intent and scope often deliver the greatest value. Like every analytics journey, successful efforts begin with a series of specific questions. What problem are you solving? What would we do differently if we could solve that problem? Finally, what data sources and analytics capabilities will help us answer the first two questions?

Answering these questions makes it possible for dark analytics initiatives to illuminate specific insights that are relevant and valuable. Remember, most of the data universe is dark, and with its sheer size and variety, it should probably stay that way.

IU Health's Rx for mining dark data

As part of a new model of care, Indiana University Health (IU Health) is exploring ways to use nontraditional and unstructured data to personalise health care for individual patients and improve overall health outcomes for the broader population.

Traditional relationships between medical care providers and patients are often transactional in nature, focusing on individual visits and specific outcomes rather than providing holistic care services on an ongoing basis. IU Health has determined that incorporating insights from additional data will help build patient loyalty and provide more useful, seamless, and cost-efficient care.

"IU Health needs a 360-degree understanding of the patients it serves in order to create the kind of care and services that will keep them in the system," says Richard Chadderton, senior vice president, engagement and strategy, IU Health. "Our organisation is exploring ways to mine and analyse data—in much the same way consumer-oriented companies are approaching customer data—to develop this deeper understanding."[13]

For example, consider the voluminous free-form notes—both written and verbal—that physicians generate during patient consultations. Deploying voice recognition, deep learning, and text analysis capabilities to these in-hand but previously underutilised sources could potentially add more depth and detail to patient medical records. These same capabilities might also be used to analyse audio recordings of patient conversations with IU Health call centres to further enhance a patient's records. Such insights could help IU Health develop a more thorough understanding of the patient's needs, and better illuminate how those patients utilise the health system's services.

Another opportunity involves using dark data to help predict need and manage care across populations. IU Health is examining how cognitive computing, external data, and patient data could help identify patterns of illness, health care access, and historical outcomes in local populations. The approaches could make it possible to incorporate socioeconomic factors that may affect patients' engagement with health care providers.

"There may be a correlation between high density per living unit and disengagement from health," says Mark Lantzy, senior vice president and chief information officer, IU Health. "It is promising that we can augment patient data with external data to determine how to better engage with people about their health. We are creating the underlying platform to uncover those correlations and are trying to create something more systemic.

The destination for our journey is an improved patient experience," he continues. "Ultimately, we want it to drive better satisfaction and engagement.

More than deliver great health care to individual patients, we want to improve population health throughout Indiana as well. To be able to impact that in some way, even incrementally, would be hugely beneficial."[14]

Retailers make it personal

Retailers almost universally recognise that digital has reshaped customer behavior and shopping. In fact, $0.56 of every dollar spent in a store is influenced by a digital interaction.[15] Yet many retailers—particularly those with brick-and-mortar operations—still struggle to deliver the digital experiences customers expect. Some focus excessively on their competitors instead of their customers and rely on the same old key performance indicators and data.[16]

In recent years, however, growing numbers of retailers have begun exploring different approaches to developing digital experiences. Some are analysing previously dark data culled from customers' digital lives and using the resulting insights to develop merchandising, marketing, customer service, and even product development strategies that offer shoppers a targeted and individualised customer experience.

Stitch Fix, for example, is an online subscription shopping service that uses images from social media and other sources to track emerging fashion trends and evolving customer preferences. Its process begins with clients answering a detailed questionnaire about their tastes in clothing. Then, with client permission, the company's team of 60 data scientists augments that information by scanning images on customers' Pinterest boards and other social media sites, analysing them, and using the resulting insights to a develop a deeper understanding of each customer's sense of style. Company stylists and artificial intelligence algorithms use these profiles to select style-appropriate items of clothing to be shipped to individual customers at regular intervals.[17]

Meanwhile, grocery supermarket chain Kroger Co. is taking a different approach that leverages Internet of Things and advanced analytics techniques. As part of a pilot program, the company is embedding a network of sensors and analytics into store shelves that can interact with the Kroger app and a digital shopping list on a customer's phone. As the customer strolls down each aisle, the system—which contains a digital history of the customer's purchases and product preferences—can spotlight specially priced products the customer may want on 4-inch displays mounted in the aisles. This pilot, which began in late 2016 with initial testing in 14 stores, is expected to expand in 2017.[18]

Expect to see more pilots and full deployments such as these in the coming months as retailers begin executing customer engagement strategies that could, if successful, transform both the shopping experience and the role that nontraditional data plays in the retail industry.

MY TAKE

**PAUL OPPENHEIMER,
CHIEF INFORMATION OFFICER**
RMIT UNIVERSITY

Illuminating the potential

When we think of data and analytics, there is such potential. When we look around an office, lecture hall or university campus, the bustling ecosystem is filled with interactions, meetings, information gathering and a myriad of data points.

Out of that comes the potential for robust information. Insights and analytics can be harnessed to paint a powerful picture of what is possible. From operational and executive dashboards to advanced analytics case studies, we need to think about what might drive analytics across the education sector and other associated industries.

Unlike a bank where a core asset is the information it holds on accounts and services, we might need to shift the mind-set a bit in education to focus on why data is fundamental to unlocking growth and opportunity.

Bringing to light the key issues

With a strong focus on learning and the delivery of content in third level education, data sometimes tends to sit in the background, but it is time for it to start moving front and centre.

As we move into a rapidly evolving digital world, digitising day-to-day elements and interactions means we can use the analysis to inform how we educate, interact and deliver content.

It's important to consider how we can create a robust ecosystem to support the university sector. Whether this is creating educational solutions with inbuilt analytics or having a diversity of channels, it would be fantastic to be able to synthesise data insights quickly and efficiently.

Purpose: what's in it for me?

If we look externally at the range of channels being used to capture data, such as social media, data in these situations indicate a propensity to buy or be interested in a certain item or product.

What if the same insights could be harnessed from data around subject choices, lecturer reviews, content feedback, course interests and more?

There is sentiment of freedom across a university campus and the role of digital or collecting insights from data might constrain this notion. So it needs to be balanced carefully – taking into account the role of regulation, policies and technology – with the end user and outcome in mind. We need to always ask ourselves: what are we trying to achieve?

In order to incorporate a move towards data being a part of university life, there would need to be education and incentives around how the use of analytics will improve and enhance a student's experience overall. A careful buy-in from the university population would ensure that people understand why data is being captured and how it is being used or interpreted.

> " ANALYTICS WILL ELEVETATE IT AS A KEY DRIVER AND COLLABORATOR "

The power of cognitive

Once you have buy-in, there is an enormous amount of potential that comes from cognitive technologies. One area to explore is capturing emotions and sentiment in the lecture sessions and classrooms to answer such questions as *"Are people engaged or disengaged?" "How are they processing the information?", "Are students interacting positively with the lecture style and format?"*

Information gleaned from this data could then be used to fuel decisions around investments in the number of classes, spatial design, collaboration areas and more.

The whole is greater than the sum of its parts – Aristotle
While in some parts of an organisation or university setting IT can be viewed as merely a supplier, it's critical that we elevate the role of IT as a key driver and collaborator on all things data.

We need to collectively start asking the right questions to elicit the data we are trying to seek. When we look at the power of data across an organisation or institution, the whole picture is more meaningful than if we only look at the individual components in isolation.

Some practical measures on what we can do right now are around leveraging analytics across the university sector and other industries includes three elements:

#1 Use what you have: there is no need at times to launch into buying and implementing new software or devices. Often the existing tools, used correctly and right now, can offer insights.

#2 Be proactive: rather than waiting for a complete strategy to be implemented, there is plenty of background research that can be done into what we can we harness right now from a data and analytics perspective.

#3 Collaboration: across a university (and other industries) there are many departments and areas that can contribute to the data conversation. Collaborate to achieve the best possible outcome.

In terms of where the sector is heading, we're only at the tip of the digital data iceberg of what's possible around *illuminating dark analytics*. What is next is a focus on asking questions and working together to explore innovative solutions and ideas.

CYBER IMPLICATIONS

Deploying analytics technologies to help illuminate actionable insights not only within raw data already in your possession but also in derived data represents a potentially powerful business opportunity. Yet dialing up your data mining and analysis efforts and importing large stores of nontraditional data from external sources can lead to questions about data veracity, integrity, legality, and appropriateness of use. These are questions that few organisations can afford to ignore today.

On the flip side, deep analysis of more data from a variety of sources may also yield signals that could potentially boost your cyber and risk management efforts. Indeed, the dark analytics trend is not just about deploying increasingly powerful analytics tools against untapped data sources. From a cyber risk perspective, this trend is also about wielding these and other tools to inspect both the data in your possession and third-party data you purchase.

As you explore the dark analytics trend, consider the following risk-related issues and opportunities:

Sourcing data: To what degree can you trust the data's integrity? If you can't confirm its accuracy, completeness, and consistency, you could be exposing your company to regulatory, financial, and even brand risks. The same goes for its authenticity. Is the source of the data who it says it is? If not, the data could be recycled from other dubious sources or, worse, stolen.

Respecting privacy: The specter of privacy law casts a long shadow over audio and video data sourced outside the enterprise. In many cases, the privacy laws that apply to audio or video clips are determined by the nationality of the individuals appearing in them. Likewise, in some countries, even recording an Internet protocol (IP) address is considered a violation of privacy. In developing a dark analytics cyber risk strategy, you should remain mindful of the vagaries of global privacy law. Finally, data that may appear to be benign could, in fact, carry potential privacy risk if it has been derived from analytics. For example, analysis of customer data might suggest a correlation between customer meal preferences and certain medical conditions or even

"THE DARK WEB REPRESENTS ONLY ONE SMALL COMPONENT PART OF THE LARGER DEEP WEB."

their religion. As you begin to curate and analyse data, how do you put proper controls in place and manage the associated privacy and legal risks? What liability could you face if you archive data containing such correlative findings?

Building predictive risk models: As you apply analytics to nontraditional data sources, there may be opportunities to create predictive risk models that are based on geography, hiring practices, loans, or various factors in the marketplace. These models could potentially help companies develop a more nuanced understanding of employee, customer, or business partner sentiment that may, in turn, make it possible to develop proactive risk mitigation strategies to address each.

Spotlighting third-party risk landscape: Global companies may depend upon hundreds or even thousands of vendors to provide data and other services. By analysing data from nontraditional sources, companies may be able to create predictive risk models that provide more detailed risk profiles of their vendors. Some of the risks identified may likely be beyond your control—a point to keep in mind as you source third-party data.

Leveraging deep web signals: The data contained in those parts of the web that are currently accessible to search engines is too vast for organisations to harness. Expanding that already-infinite universe to include nontraditional data from the deep web may present data analysis opportunities, but from a cyber risk standpoint, it may also present considerable risks. The dark web represents only one small component part of the larger deep web, but it has, time and again, been at the root of cyber challenges and issues. As such, it will likely amplify the risk challenges and complexities that companies face should they choose to explore it. Proceed with eyes wide open.

That said, by applying risk modeling to bounded data sets sourced from the larger deep web, organisations may be able to further expand their knowledge in the realms of cybersecurity, competitive intelligence, customer engagement, and other areas of strategic priority.

Where do you start?

Three years from now, your organisation may find itself overwhelmed by immeasurable volumes of unstructured data generated by Internet of Things devices. Working today to develop the discipline and tools you will need to manage and mine all of this dark data can help your organisation generate data-driven insights today while preparing for even greater opportunities in the years ahead.

This process begins with a series of practical steps:

- **Ask the right questions:** Rather than attempting to discover and inventory all of the dark data hidden within and outside your organisation, work with business teams to identify specific questions they want answered. Work to identify potential dark analytics sources and the untapped opportunities contained therein. Then focus your analytics efforts on those data streams and sources that are particularly relevant. For example, if marketing wants to boost sales of sports equipment in a certain region, analytics teams can focus their efforts on real-time sales transaction streams, inventory, and product pricing data at select stores within the target region. They could then supplement this data with historic unstructured data—in-store video analysis of customer foot traffic, social sentiment, influencer behavior, or even pictures of displays or product placement across sites—to generate more nuanced insights.

- **Look outside your organisation:** You can augment your own data with publicly available demographic, location, and statistical information. Not only can this help your analytics teams generate more expansive, detailed reports—it can put insights in a more useful context. For example, a physician makes recommendations to an asthma patient based on her known health history and a current examination. By reviewing local weather data, he can also provide short-term solutions to help her through a flare-up during pollen season. In another example, employers might analyse data from geospatial tools, traffic patterns, and employee turnover to determine the extent to which employee job satisfaction levels are being adversely impacted by commute times.

- **Augment data talent:** Data scientists are an increasingly valuable resource, especially those who can artfully combine deep modeling and statistical techniques with industry or function-specific insights and creative problem framing. Going forward, those with demonstrable expertise in a few areas will likely be in demand. For example, both machine learning and deep learning require programmatic expertise—the ability to build established patterns to determine the appropriate combination of data corpus and method to uncover reasonable, defensible insights. Likewise, visual and graphic design skills may be increasingly critical given that visually communicating results and explaining rationales are essential for broad organisational adoption. Finally, traditional skills such as master data management and data architecture will be as valuable as ever—particularly as more companies begin laying the foundations they'll need to meet the diverse, expansive, and exploding data needs of tomorrow.

- **Explore advanced visualisation tools:** Not everyone in your organisation will be able to digest a printout of advanced Bayesian statistics and apply them to business practices. Most people need to understand the "so what" and the "why" of complex analytical insights before they can turn insight into action. In many situations, information can be more easily digested when presented as an infographic, a dashboard, or another type of visual representation. Visual and design software packages can do more than generate eye-catching graphics such as bubble charts, word clouds, and heat maps—they can boost business intelligence by repackaging big data into smaller, more meaningful chunks, delivering value to users much faster. Additionally, the insights (and the tools) can be made accessible across the enterprise, beyond the IT department, and to business users at all levels, to create more agile, cross-functional teams.

- **View it as a business-driven effort:** It's time to recognise analytics as an overall business strategy rather than as an IT function. To that end, work with C-suite colleagues to garner support for your dark analytics approach. Many CEOs are making data a cornerstone of overall business strategy, which mandates more sophisticated techniques and accountability for more deliberate handling of the underlying assets. By understanding your organisation's agenda and goals, you can determine the value that must be delivered, define the questions that should be asked, and decide how to harness available data to generate answers. Data analytics then becomes an insight-driven advantage in the marketplace. The best way to help ensure buy-in is to first pilot a project that will demonstrate the tangible ROI that can be realised by the organisation with a businesswide analytics strategy.

- **Think broadly:** As you develop new capabilities and strategies, think about how you can extend them across the organisation as well as to customers, vendors, and business partners. Your new data strategy becomes part of your reference architecture that others can use.

Bottom line

With ever-growing data troves still unexplored, aggregation, analysis, and storage are no longer end goals in the agile organisation's analytics strategy. Going forward, analytics efforts will focus on illuminating powerful strategic, customer, and operational insights hidden within untraditional and dark data sources. Be excited about the potential of unstructured and external data, but stay grounded in specific business questions with bounded scope and measurable, attributable value. Use these questions to focus your dark analytics efforts on areas that matter to your business—and to avoid getting lost in the increasingly vast unknown.

CONTACTS

LEE ANDERSON

Partner, Consulting
+61 407 003 039
leeanderson@deloitte.com.au

MELISSA FERRER

Partner, Consulting
+ 61 403 349 192
meferrer@deloitte.com.au

AUTHORS

TRACIE KAMBIES

US Retail Analytics & Information Management and IoT lead, Deloitte Consulting LLP

NITIN MITTAL

US Analytics & Information Management practice leader, Deloitte Consulting LLP

PAUL ROMA

Deloitte Analytics leader, Deloitte Consulting LLP

SANDEEP KUMAR SHARMA, PH.D.

Deputy chief technology officer, Deloitte Consulting LLP

ENDNOTES

1. Khalid Kark, Mark White, Bill Briggs, and Anjali Shaikh, *Navigating legacy: Charting the course to business value*, Deloitte University Press, November 10, 2016,

2. Marc Goodman, "Most of the web is invisible to Google. Here's what it contains," *Popular Science*, April 1, 2015, www.popsci.com/dark-web-revealed.

3. Åse Dragland, "Big data, for better or worse: 90% of world's data generated over last two years," *Science Daily*, May 22, 2013, www.sciencedaily.com/releases/2013/05/130522085217.htm.

4. EMC Digital Universe with research and analysis by IDC, "The digital universe of opportunities: Rich data and the increasing value of the Internet of Things," April 2014, www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm.

5. Gartner Inc., "Gartner says 6.4 billion connected 'things' will be in use in 2016, up 30 percent from 2015," press release, November 10, 2015, www.gartner.com/newsroom/id/3165317.

6. Cisco, "Cisco Global Cloud Index: Forecast and methodology, 2014–2019," October 28, 2015, www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html.

7. Dan Vesset and David Schubmehl, "IDC FutureScape: Worldwide big data, business analytics, and cognitive software 2017 predictions," December 2016, International Data Corporation.

8. Ibid.

9. Seth Grymes, "Unstructured data and the 80 percent rule," *Breakthrough Analysis*, August 1, 2008, https://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/.

10. Cindy LaChapelle, "The cost of data management: Where is it headed in 2016?" *Datacenter Journal*, March 10, 2016, www.datacenterjournal.com/cost-data-storage-management-headed-2016/.

11. Microsoft, "Multi-lingual, federated search solution provides global access to scientific research," March 24, 2015, https://customers.microsoft.com/en-US/story/multilingual-federated-search-solution-provides-global.

12. David Barton, "Data mining in the deep web," *The Innovation Enterprise*, July 14, 2016, https://channels.theinnovationenterprise.com/articles/data-mining-in-the-deep-web.

13. Interview with Richard Chadderton, senior vice president of engagement and strategy, Indiana University Health, December 9, 2016.

14. Interview with Mark Lantzy, CIO of IU Health, November 21, 2016.

15. Jeff Simpson, Lokesh Ohri, and Kasey M. Lobaugh, *The new digital divide*, Deloitte University Press, September 12, 2016, https://dupress.deloitte.com/dup-us-en/industry/retail-distribution/digital-divide-changing-consumer-behavior.html.

16. Ibid.

17. Hilary Milnes, "How Stitch Fix's happy relationship with Pinterest helps customers," *Digiday*, March 16, 2016, http://digiday.com/brands/how-stitch-fix-depends-on-pinterest-to-make-its-customers-happy/.

18. Kim S. Nash, "Kroger tests sensors, analytics in interactive grocery shelves," *Wall Street Journal*, January 20, 2017, http://blogs.wsj.com/cio/2017/01/20/kroger-tests-sensors-analytics-in-interactive-grocery-shelves/.