Private-Sector Applications of Data Science

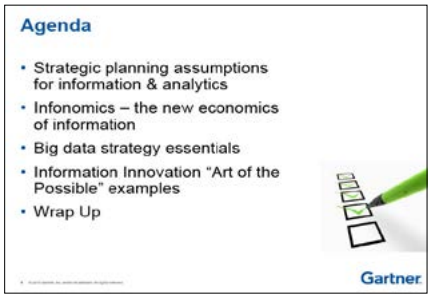
Doug Laney
Vice President and Distinguished Analyst,
Chief Data Office Research Team, Gartner, Inc.

. . . organizations are using information and analytics in innovative ways.

The following are Doug Laney's remarks to an Intelligence Community audience in 2016 on the ways that now abundantly available data are being used in the private sector.

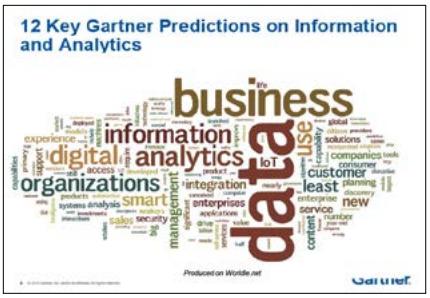
I'm part of Gartner's Chief Data Officer (CDO) research team and, as part of that team, focus on data and analytic strategy. So, while a lot of my colleagues at Gartner will track vendors and technologies in the marketplace, those of us on the CDO research team are under a mandate that, if we ever work on a magic quadrant, we'll be fired.[a] So, we tend to focus just more on the strategy aspects.

We're going to cover a few things. I know we've got about 45 minutes, and I will leave some time for questions and answers, as well. I'm going to talk about some strategic planning assumptions, or what we at Gartner call "our predictions." It wouldn't be a Gartner presentation if I didn't share some predictions with you. I'm going to share some of the thoughts on infonomics—this idea that information is an actual corporate asset, or becoming treated much more like an asset, and kind of the imperative behind that. I'll share a bit on big data strategy essentials, some of our high-level thinking on what big data means, and how to approach it. Then I have some time to share some examples; we've compiled a library of examples of how organizations are using information and analytics in innovative ways. We've been compiling this for about five years, and I want to share with you mostly what's happening in the commercial world with respect to big data and analytics.

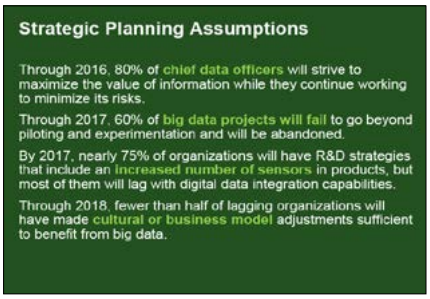


At Gartner, we issue probably 500 to 700 predictions a year. These are 24-word predictions with some detail behind them about what's happening in the world of technology over the next few years, usually about a five-year horizon.

a. The term "magic quadrants" refers to Gartner's proprietary research methodology that looks at challengers, leaders, niche players, and visionaries within a major technology market. Source: Gartner, Inc.; available online at http://www.gartner.com/technology/research/methodologies/research_mq.jsp.



So, we'll start with a look near-term. , and one of the things that we see is chief data officers—this emergence of the chief data officer. Right now, there are probably 2,000 to 3,000 chief data officers—or people with a similar kind of title or role—throughout the world. We find that, initially, they get installed to focus on compliance- and security-related issues, but very quickly they have to turn their attention to value-oriented activities.

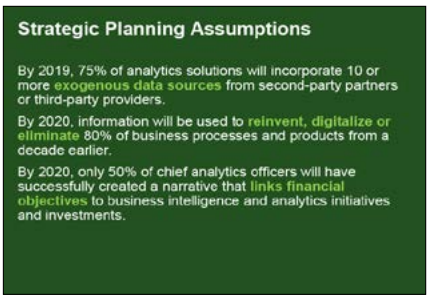


Over the next couple of years, we'll continue to see big data projects fail—and there's a lot of press around how big data projects typically fail. We think that's a good thing. Most big data projects should be experimental in nature, and the companies that are most successful with big data treat it as an ongoing type of experiment. Google, of course, is a great example of that; they run hundreds or thousands of experiments a day with information.

We think that organizations should become more interested and involved in doing R&D (research and development) around information and not just R&D around their core products and services. We're also seeing, of course, an increased interest in sensors and IoT (Internet of Things) devices and many organizations doing R&D around IoT. I take most of my calls, the last month or so, from clients about "How do we install a sensor or how do we engage somebody who wants to install a sensor in one of our products or services? Who owns the data? Who gets rights to the data? How does that impact the relationship and the economics of the product or service?" So, that's of serious concern as well.

Yet, we see most CDOs and most organizations failing to make the kinds of cultural or business model adjustments to really leverage information. It's great to come up with ideas for how to use information or analytics. But if you're not prepared as an organization to actually act on it, then really you've wasted time. We've seen too many organizations bringing on board a data scientist or a wanna-be data scientist who's developing all sorts of interesting insights, but not actually linking them to the business.

There's a client of ours with Bell Helicopter who had a data scientist developing all sorts of interesting insights, and he said to the data scientist, "You want to keep your job?" And the guy said, "Well, yeah, I want to keep my job." He said, "Well, here's what I need you to do. Next year—by next year—I need you to sell me four more helicopters." He said, "I don't care how you do it, what kind of data you analyze. Figure out how to help Sales and Marketing sell four more helicopters." It's that kind of goal-oriented approach that's going to make big data projects and data scientists much more effective.



As far as analytics go, a lot of organizations are no longer just staring at their own data navels when it comes to using information; they're increasingly looking to the outside. What kind of data can they capture from secondary or tertiary sources—from partners, from customers, from suppliers, from syndicated data sources? There are an estimated 5,000 data brokers out there, all aggregating and licensing data. There are an estimated 10 million public data sources produced mostly by government organizations worldwide—and so, commercial organizations are starting to tap into this as a differentiator. The data itself may not be a differentiator, but their ability to curate it and leverage it is a differentiator.

By 2020, we find that information is going to be used to digitalize or reinvent or eliminate most business processes from a year earlier. Think

about any kind of business process that you're involved in, maybe just as a citizen, 10 years ago. There's probably some aspect of that process or that service or that offering that has become digitalized in some way, even if it's just the tennis racket that you use. If you use one of the Babolat tennis rackets, it now has a sensor in the handle that you can track your swing. Examples abound of this and it's certainly going to continue.

By 2020, we think chief data officers are going to still continue to struggle to link what they're doing to financial objectives. I had the pleasure to speak with a number of organizations in the [Washington] DC area the last couple of days, and this continues to be a struggle. How does doing data quality or doing data governance or doing master data management or even analytics tie directly to the mission of the organization? That's going to continue to be a challenge and something that CDOs really need to pay attention to.

These are some new predictions. We have not published them yet, but I wanted to share them with you. You're among the first to actually see these and perhaps you'd like to help me vet them. By close to the end of the decade, we see organizations rejecting vendor solutions, and this is something that's of serious concern to some of our clients—that their package application vendors, especially when that application is running in the Cloud, feel that they have some dominion over that data, and that if you as an organization want to extract data from that application, then you have to have a license for anybody who's using that data. We think that's absolutely ridiculous. They refer to that as multiplexing; we think that's just absolutely nonsense. If Microsoft or SAP thinks they own the data that you're creating in their application, then we think there's a serious issue with that and that organizations are going to get a little more cognizant about doing business with those kinds of companies.

How does doing data quality or doing data governance or doing master data management or even analytics tie directly to the mission of the organization?

Next is that we see the importance of algorithms. Even though organizations can't copyright or secure the intellectual property of information—because information is not considered property or an asset—organizations can patent the ways in which they are leveraging that information. From a commercial perspective, we're seeing organizations scrambling to secure the ways they're leveraging information. I just posted a blog on this a few weeks ago.[a] I did an analysis of what kind of organizations were patenting algorithms. Any guess who's in the top 10?

Audience Member: Google?

DL: Google—no.

Audience: IBM (International Business Machines)?

DL: IBM's number 10.

Audience: Oracle [Corporation]?

DL: Oracle's not.

Audience: Car companies?

DL: No.

Audience: Uber [Technologies Inc.]?

DL: Not Uber.

Audience: Qualcomm?

DL: Qualcomm's not in the top 10, but close. Nine of the top 10 are Chinese companies or universities. Thirty-three of the top 40 organizations patenting algorithms are Chinese organizations or companies. So, if you want, you can read my blog and see why I think that's actually happening. I'm sure your suspicions are the same.

Next, we see companies participating in online marketplaces of information. These marketplaces are starting to emerge in the healthcare space and other spaces where companies can participate in making certain data available, or in licensing that data. That's increasingly one of the exogenous data sources that is available. We think that within organizations, as information starts to become recognized as an actual asset, we need a new language for businesspeople, for IT people—for information people—to be able to communicate effectively. One of my colleagues, Valerie Logan, has come up with this notion of "information as a second language," as a way to develop a vernacular that helps people talk about information and analytics in a consistent, clear way.

Finally, we're seeing a prevalence of equity analysts becoming interested in enterprises—in companies' in-

a. The blog entry to which Mr. Laney refers, "Algorithm Patents Increased 30x The Past Fifteen Years," is available online at http://blogs.gartner.com/doug-laney/patents-for-algorithms-have-increased-30x-the-past-fifteen-years.

. . . organizations are going to start to audit value their information assets as a way to prove they're more serious about information and analytics. . .

formation and analytics capabilities. What we believe is that organizations are going to start to audit and, at least internally, value their information assets as a way to prove to investors that they're serious about information and analytics.

Let's talk now about this economics of information. After 9/11, I was an analyst with Gartner in my first go-around (I'm a Gartner recidivist). In my first go-around, some clients starting calling us, lamenting, of course, not only the tragic loss of life, but also the loss of their data. And while we revel in the first responders, I actually think there were some worst responders after 9/11—and those include the accounting profession and the insurance industry.

What happened was that some companies submitted . . . claims to insurers for the value of what [they] lost. A lot of companies (remember, this is in the days before Cloud and off-site backups) actually lost their data. So they submitted claims to their insurers. The insurers said, "Now, hold on a second. We don't think that information actually constitutes property. Therefore, we're not going to cover it on your property and casualty policies."

What ensued was a number of court cases, and we've tracked dozens of court cases around this. The courts are thoroughly confused as to whether information constitutes property or not. Some of the courts have said things like, "Well, yes, information can be represented by bubbles on an optical drive; therefore, it's physical in nature and should be covered." Other courts have said ridiculous things like, "Well, electrons have negligible mass; therefore, information should not be considered property." So, that's the world that we're in right now: no major insurance companies offer information insurance. They offer business continuity insurance and cyber insurance and but do not insure the value of the information itself.

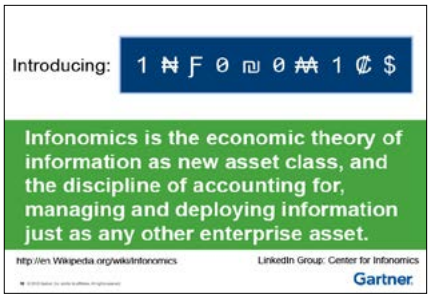
Not to be outdone, the accounting profession followed suit a few years later and updated a key accounting standard to state that, "Even if you wanted to capitalize the information that you have, even if you're a Google or Experian or Transunion and all you do is aggregate and sell data, you can't put the value of that on your balance sheet. It's not something that is capitalizable." And so, that's the world that we're in today. And that introduces, of course, some challenges for some companies but also opportunities—opportunities to leverage information in a way that's "off-book." We see some companies taking advantage of that. But we also see that there is some risk involved as well.

Okay: so accountants don't recognize information as an asset. However, they made an appeal to Congress.

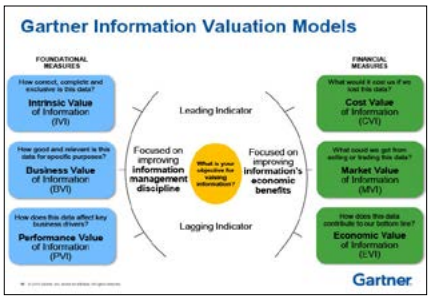
In the months before 9/11, there was a hearing on how to evolve a 1930-style accounting system into the 21st century. Of course, the appeals fell on deaf ears, and nothing ever happened. The accountants claimed, "The inability for us to account for information assets and other kinds of intangibles introduces undue volatility into the marketplace." But, again, nothing happened from that.

But we are starting to see investors take notice of the value of information, and one of the ways that they're doing so is in rewarding companies that are more information savvy. We took a look at companies that have chief data officers, that have data scientists, that have enterprise data governance functions. Those companies have a market-to-book value—more specifically, if you're an accountant, a Tobin's "q" ratio, which is a market-to-tangible asset value, so a metric devised by James Tobin, who was a Nobel economist in the 1960s, simple ratio—but anyway, companies have basically a market-to-book value that's two-to-three-times higher than the norm. There's something about these companies (and I'm not saying that investors are really paying attention to whether you're doing enterprise data governance) there's just something about these companies that investors notice, and it's really significant. We're looking to redo that study here shortly.

Infonomics is basically the idea that information is or should be a recognized asset. And even if it isn't, organizations should treat it as if were one.

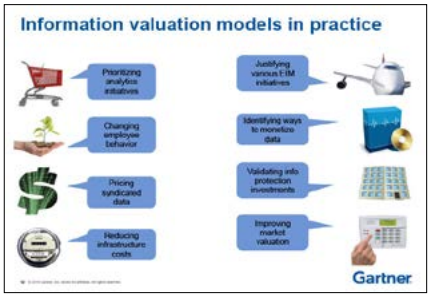


All of this gets us to the concept of infonomics. Infonomics is a term I started using somewhat casually about 15 or 16 years ago but never really put the meat on the bones. Gartner asked me to return and start to develop this idea. Infonomics is basically the idea that information is or should be a recognized asset. And even if it isn't, organizations should treat it as if were one: they should monetize it, measure it, and manage it with the same kind of discipline as they do their other assets.



We have developed models on how to quantify information's value. I'm not going to go into detail, but we have come up with two sets of models. The first is a foundational model that looks at the intrinsic value of information. The second is a financial model, which looks at information's relevancy across a range of business processes.

The performance value of information is looking at information's contribution to one or more key performance indicators—nonfinancial key performance indicators. On the financial measures side, we've borrowed from the way organizations value any kind of asset—that is, using the cost approach, the market approach, and the income approach. We've adapted some of the models to accommodate the nuances of information, like, "when you consume information, it's not depleted," and "when you sell information, you're actually licensing it—you're not transferring ownership of it," in most cases. So, we've adapted those models. Now, our clients are using them, mostly in combination and very often they adapt them to their purposes.



Just a couple of examples to show how organizations are using these models. The two at the bottom [in the slide above, "Information valuation models in practice"] are probably the most interesting. There's an energy company in Indiana, MISO (Midcontinent Independent System Operator, Inc.) Energy, that used the cost model to understand the cost of retaining information—of collecting, securing, storing, and managing certain information assets that they felt they weren't utilizing very well. They ran the economic model on them and found that certain information assets were costing them more to retain than the future probable economic benefits that they were going to achieve from them. They were able to then make a defensible disposal decision, saving over a million dollars a year just on that one kind of information asset—by disposing of it.

Another company—a security system company in commercial and residential security system—said, "Listen, we've got a lot of data that we feel is underutilized." And they ran the business value of information model just to find out which information had potential value if they were to leverage it across a range of business processes. Then they used the economic model to say, "Okay, what is this information actually generating from us in terms of economic value?" And where they found some great discrepancies, rather than disposing of the data, they said, "Let's figure out how to raise the value of that data—the realized value of that data—by leveraging it across these business processes." So, basically what they did was they innovated around underperforming information assets, which was something I hadn't ever really considered. But it turns out that this $2 billion company increased the market value of the company by $300 million by going through that effort.

The accounting perspective stuff notwithstanding, we think that if people are talking about information as an asset, then they should manage it like an asset. I was talking to the CIO (chief information officer) of the Navy yesterday and having this discussion, which is, "You guys are

The woman who used to head information innovation for the Navy and the Marines said, "You know, it's a sad state of affairs that we have a better inventory of the toilets in the Pentagon than our information assets."

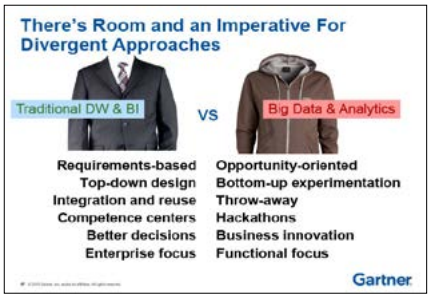
experts at managing certain kinds of physical assets and human capital. Why haven't you thought about how to apply those kinds of principles and practices to the management of information?" This is a discussion that I have with most organizations.

A former head of information innovation for the Navy and the Marines once said, "You know, it's really a sad state of affairs that we have a better inventory of the toilets in the Pentagon than our information assets. For the business that we're in, that's a really sad state of affairs." But this is endemic in organizations of any size and in any industry—that they have a better accounting of things that don't really contribute business value to the organization than their information assets. So, job one is to inventory the information. But there are other asset management principles and practices that we think apply—or should apply—to information.

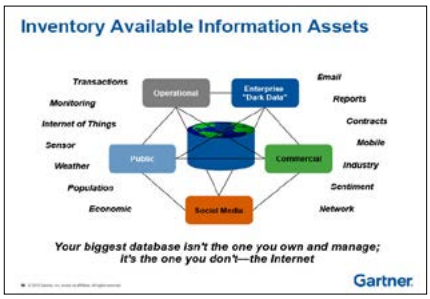
Some quick essentials on data strategies—first, "big data" is number one. "Big data" is the number one most ambiguous term, according to the global language monitor. So, a company in Texas that monitors terms on the web said, "Yeah, big data is the most ambiguous term, because it's ill-defined." It's also the number one search term on Gartner. Up until this year, it was number one search term about three years running. So, our clients were very concerned about it—what is it, what to do about it.

So, years ago, I came up with this idea of the three Vs—volume, velocity, and variety—and I was thinking in terms of Y2K (Year 2000) efforts and the emergence of e-commerce at that time. The three Vs are now applied kind of as a catch-all definition for big data and that's great, but Gartner's definition has evolved beyond that to appreciate that big data should be used for not just decisionmaking but also for generating insights for automation or optimization and, increasingly, for monetization as well. Some companies are commercializing their data as well. But we advised companies, "Hey, you're in the realm of big data when you've got to retool your processing or your architecture and introduce new forms of innovative processing." So, that's really our overall definition—a bit of a mouthful.

When it comes to leveraging big data and the difference between doing enterprise analytics and big data and advanced analytics, we refer to that as kind of the "suits versus the hoodies."

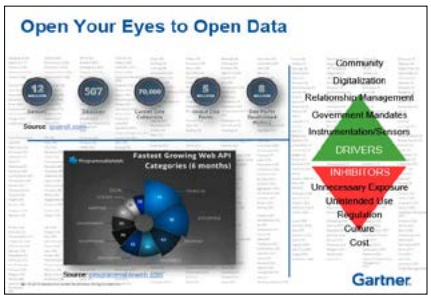


The suits are the classical enterprise data warehouse and business intelligence crowd, building solutions from the bottom up, having an enterprise focus, focusing on decisionmaking, more top-down design. The hoodies, on the other hand, are the folks who are more inventive, creative, experimental with information, even analyzing things and then throwing it away—very functionally focused. When we look at this library of examples and do a meta-analysis of them, almost all of them are on the right side, here—the hoodies. They are very functionally specific use cases, and high value; they're not "enterprise reporting" it. It's not pretty pie charts and bouncy bar charts: it's much deeper than that.



When it comes to what kind of information is available, we advise our clients to think beyond their own four walls. When it comes to

information, think about not only your transaction data but about what else you could capture as part of that transaction—either using IoT devices or some kind of observation technology. Be aware that there are these thousands of commercial data sets available and millions of public data sets available—social media as well. But more than that, in the center here, a company's biggest data base is the Internet itself. You probably know this, but a lot of commercial organizations don't realize that they could harvest content from the Internet and use that for a particular advantage.

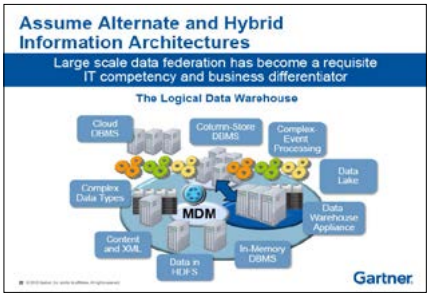


So, lots of open data out there; I'm not going to belabor the point. This is just an example of a couple of the companies that provide access to a variety of open and syndicated data.

When it comes to generating ideas for big data, a lot of clients come to us and they say, "What are others in our industry doing?" So, I really welcomed the opportunity to talk to you here today, because I don't know a lot about what others in your industry are doing, but I know a lot about what others in other industries are doing. I say to our clients, "It's good to know what others in your own industry are doing with information and analytics. But why do you want to be in second place? Why not gather ideas from other industries and think about how to adapt and adopt them to your own industry?"

When it comes to data monetization, we're not so doctrinaire as to say that data monetization involves the licensing or sale of data. Data monetization is any way that you're generating some economic benefits from the data. I know that may not be a comfortable vernacular in the public sector, but there's a range of indirect and direct ways to monetize data, ranging from improving efficiencies, reducing risks, developing new products, all the way down to actually selling or licensing the data yourself. The most common way that we see organizations generating economic value from their data is by bartering it or exchanging it.

When you go to the grocery store and you scan your loyalty card, they call that a discount—you're getting a discount on your groceries. But actually, it's a barter transaction: you're exchanging information about you and your purchase for free food. Of course, the grocers don't want to disclose it that way because that would have tax implications, but the reality is, that's what's happening. That kind of thing is becoming a lot more prevalent in the B2B (business to business) spaces, especially with telcos (telecommunications companies) and retailers.

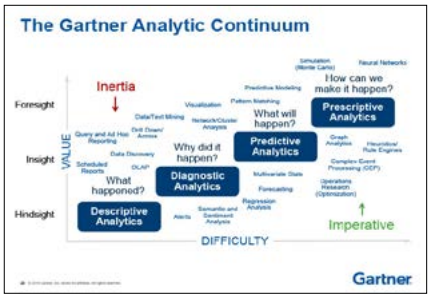


Hybrid architecture—so, I'm not going to talk much about technology other than to say most companies are thinking much more about a hybrid architecture when it comes to information, with a layer on top that makes it look like a common architecture to applications or individuals. Those application components may be data warehouse appliances; they may be a Cloud-based data, or Hadoop[a] data, or in memory databases. But generally, it's no longer a world of your father's enterprise data warehouse—the monolithic enterprise data warehouse.

When it comes to analytics, we're desperately trying to get our clients "off the schneid" when it comes to doing just descriptive analytics. There's a huge comfort level with or-

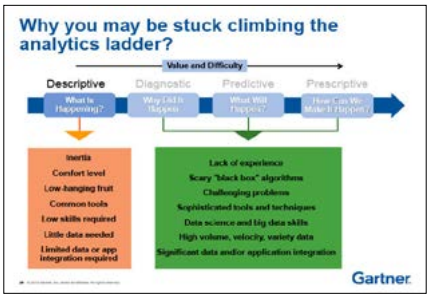
a. Hadoop is "an open-source software framework for storing data and running applications on clusters of commodity hardware. It provides massive storage for any kind of data, enormous processing power and the ability to handle virtually limitless concurrent tasks or jobs." Source: http://www.sas.com/en_us/insights/big-data/hadoop.html.

ganizations' doing descriptive analytics that answers the question, "What happened?" This is represented by traditional typical BI (business intelligence)—pretty pie charts, bouncy bar charts, Excel spreadsheets—that kind of thing. We're trying to get them to move up this maturity or this continuum into doing more diagnostic analytics that answers the question, "Why did something happen?", predictive analytics that answers the question, "What's going to happen?", and prescriptive analytics that answers the question, "How do we make something happen?" The technologies—the skills—become much more sophisticated as you move up that continuum; but we think that there's a real imperative to do so. This is the land of the data scientists, of course. Yet we see so much inertia pushing down on organizations, because there's so much comfort level with people just sitting around creating bar charts and pie charts.



Okay. I think I just covered this (referring to slide above, right, entitled, "Why you may be stuck climbing the analystics ladder?"). We'll include it in the materials, but this is basically what I was saying about the inertia and the challenges:

We also see organizations setting up or evolving their business intelligence competency centers more toward analytic centers of excellence, involving a host of analytic kinds of capabilities from project prioritization to experimentation and governance. As organizations move toward self-service analytics, on one end, and data science on the other, this provides some adult supervision over that. Sometimes these analytic centers of excellence are virtual; sometimes they're more physical organizations. We're about to publish a piece detailing about 30 or so different competencies that we think ought to be part of an analytic center of excellence.



I'm going to wrap up with some examples; I have tons of examples! I made some notes, because the slides ended up getting too big if I had actually included them. I'll make the same offer to you that I made to the Navy and the counterintelligence folks—that, while we generally don't share the entire library with our clients (we have several hundred, maybe 300 examples in the library), we do share more relevant examples with them.[a]

Let me share some examples of operational performance. One of the interesting ones I really like, because I gave them the idea, is ***Lockheed Martin***. Lockheed runs very complex, very sophisticated manufacturing projects that involve lots of people, lots of contractors, lots of different technologies and subsystems, and very protracted types of projects. Like most companies, they had been gauging the status of these projects using project managers' reports, status reports, and then rolling them up. Well, that takes some time and, as you can imagine, most project managers want to couch potential issues.

So, we were having a discussion about dark data—dark data being data that is un- or underutilized in the organization—and Lockheed said, "Well, what dark data do we have?" And I said, "You know, you could probably use your e-mails and project communications and other kinds of project documents, and run some machine learning over them to learn what are the leading indicators of issues—project issues related to scope or budget or personnel or technology or name whatever kinds of issues there are, and then throw up yellow flags when there are such issues."

This is actually the type of work they've done over the last few years. They implemented a system to identify these leading indicators by analyzing project communications, and it has led to their having what

a. Here, Mr. Laney refers to Gartner's "Information Innovation Library," a proprietary database.

they claim is three times greater foresight into emerging project issues than the previous, more manual methods—and it's saving them hundreds of millions of dollars a year in cost overruns. So, that's one great example.

Another one is a company called ***PASSUR Aerospace, Inc***. I didn't know this, but airline pilots actually are the ones who issue the ETAs (estimated time of arrivals). It's not issued via any really formal process. Very often, they're off by five or 10 minutes, and that affects the scheduling of gate and ground personnel at airports. What PASSUR has done is to introduce passive radar at airports and also crunch all sorts of historical airline and weather and other kinds of data, and they create these various "sky scenarios." When the plane is flying, they say, "Ah, this is just like one of these sky scenarios that we've seen before. Let's adjust the ETA appropriately." At airports where this system is being used, they're saving millions of dollars a year in more efficient personnel—ground and gate personnel—scheduling. United Airlines has claimed that, at airports where this system is being used, they're avoiding two to three diversions per week.

The refinery of the Mexican oil company ***PEMEX*** had a terrible reputation for shutting down due to the failure of one or more components. So, somebody asked the engineers, "Well, how do you know when a certain component is about to fail?" And the engineer said, "Oh, hace ruido"—"It makes noise." Of course! So, they said, "Well, what creates noise is vibration . . . let's put vibration sensors on these components and take baseline readings, and then identify when they're about to fail and do more proactive maintenance than reactive maintenance, keeping the refinery up and running." They've added thousands of hours of more capacity to the refineries where they've implemented this.

They implemented a system to identify these leading indicators by analyzing project communications, and it has led to their having what they claim is three times greater foresight into emerging project issues . . .

There are a lot of examples of organizations doing the same kind of thing. ***Trulia*** and ***Zillow***—they're now one company—the real estate information aggregator. If you've ever bought or sold a house or are looking for a house, you probably landed on their site. Ninety percent of their web traffic were people looking at photos—photos of houses. Yet Trulia realized, "We have no idea what's in those photos. They've never been tagged." When you upload a photo of your house, you don't typically tag it as, "here's the living room, here's the dining room," etc. People just kind of scroll through it.

But Trulia wanted to understand what people are looking at in these houses and how that corresponds to their actually scheduling visits or purchasing houses, or how it relates to pricing. They actually want to improve the way they're pricing houses. So they built a billion-node neural network to understand what's in those photos. And now you can go to Trulia and say, "Show me houses in the Hamptons [New York] with wine cellars in the basement," and it could actually pull those up for you automatically. I know Google's got a similar feature, but this is really one of the first and most interesting uses of machine learning that I've seen.

Another example is ***Coca-Cola***. This is probably one of my favorite examples. Coca-Cola owns the Minute Maid brand for orange juice and they've owned that brand since the 1960s. The CIO I spoke to there said, "Yeah, we had this issue one time where Chick-fil-A was going to cancel their contract with us, because the quality of our orange juice was inconsistent from month to month. Well, of course, the supply of oranges is inconsistent: sometimes they're coming from Florida, sometimes they're from California or Mexico or Brazil or Israel or wherever, and based on disruptions in the supply chain, we never know what kind of oranges we're going to get. We just kind of blend them as best as we can."

What they decided to do was, first, understand the 600 flavors that comprise an orange. They had flavorologists and scientists to analyze this. Then they started capturing all sorts of external data—weather data, crop yield data, multilevel insight into the supply chain and the suppliers—and developed a precise, dynamic formula on how to blend orange juice. They say that they take into consideration a quintillion data points to do this! (I've never done the math to see if that's reasonable; that's their claim.) So, now if there's a hurricane or a freeze or a frost, Minute Maid can replan the entire business in five to 10 minutes and achieve this orange juice quality.

The ***Smithsonian Conservation Institute*** tracks the migration of mostly large fauna across the United

States and it does this using tags—radio transmitters they tag the animals and then they pick up the signals by satellite. The problem was they wanted to start tracking smaller fauna, like fireflies and butterflies and grasshoppers. They said, "Well, the tags, the trackers, are too big. They're too heavy to be able to be picked up by satellites." They developed smaller transmitters they put on insects, and they outfitted United Airlines planes with detectors, so every United Airlines plane criss-crossing the United States has one of these detectors that can track and pick up signals from the small fauna. I thought that was a really interesting example of a private-public partnership.

A great example of leveraging an algorithm from a different market has to do with the ***Los Angeles***, where somebody came up with the idea that crimes tend to follow a pattern that looks like seismic aftershocks—earthquake aftershocks. They experimented with this and found that by applying seismic algorithms, they can help to predict where crimes are going to occur in the city. By deploying resources accordingly, they've been able to reduce violent crimes in the parts of town where they've deployed this by 30 percent.

I'll share two more. At the ***University of Rochester***, some students and researchers started tracking people who tweeted that they had attended a particular restaurant. Then they tracked them for a 72-hour period to see if they subsequently tweeted about nausea or stomach pains or something like that. What they found was several hundred instances of this, that when they looked at the public database of New York City restaurant health scores correlated precisely to those restaurants. So, that was just a little experiment. But the New York City Department of Health is now looking at the solution as something that they might want to use more proactively.

And then finally there's a fun story of a really interesting use of information. In ***Iceland***, everybody's related to everybody else, right, which makes for really fun family reunions but difficult to find a mate who is not your cousin. So, some enterprising young folks took the Icelandic genealogy database, which is publicly available, and baked it into a mobile app. So, now if you're out carousing and you're out at a pub—and this is their tagline—this is not my tagline: "you can bump your phones before you bump in bed." Apparently it's the most popular app in Iceland, and everybody uses it just for fun, I guess, to keep cousins from kissing, or worse!

Let's move to some of the financial examples, then I'll take some questions. A lot of companies are concerned about knowing who their customers are. ***Westpac*** is one of the larger banks in Australia (but keep in mind Australia's about as big as Missouri, in terms of population). They had this problem where they just didn't know their customers very well. They weren't able to provide them targeted offerings.

So they created this "know me" program, to better understand their customers—track them on all the touch points. Within nine months, they were able to target 25 percent of their customers with offerings that were meaningful to them and added about $25 million of additional revenue through this program. A classic "customer 360" example.

There is a guy in the Chicago area I know, who did something really interesting with respect to monetizing information. He built a private residence at the ***University of Illinois***. His company is called HERELife. The private residence tracks all of the students—everything they're doing: when their doors are opening, when they're using the microwave. They have special iPads to use; they have cars they can share; the equipment in the gymnasium is tracked—they track everything about what these students are doing. He aggregates the data and then makes it available to major brands like Coke and Pepsi and Walmart and so forth. The kids are very happy to be part of this grand experiment, and they get some discounts on the housing as well. But it's a really interesting way of setting up an IoT-laden residence for students and then monetizing that data.

We talk about this world of exogenous data and all the millions of data sets that are out there. One company in particular is taking advantage of this. It's a software company called ***Prevedere***, which has pre-curated two million external data sets from some syndicated sources and some open data sources. They can take a company's forecast and tune that forecast, based on leading indicators. For example, they worked with Anheuser-Busch to improve its forecast for beer sales in China. Their beer sales forecast to China was about 75-percent accurate. By identifying what kinds of weather or economic indicators or other kinds of indicators were leading indicators of Anheuser-Busch's beer sales, they were able to improve its forecast to 95-percent

accuracy. I don't know how that kind of solution might be leveraged in your world, but it's interesting to know that there's someone out there in the commercial space who is aggregating millions of data sets to improve forecasts.

Now let's talk about Walmart. Walmart had a search engine, of course, so you'd go to Walmart online and type what you're searching for, and it would take you where you wanted to go—a great search engine. Walmart relies on about 45 million or so searches per month and uses that data very well to help you find what you're looking for; however, early on, it didn't take into consideration what was happening in the world at the time.

They tell this story about how people were searching for the word "house," and how the search engine was taking them to housewares. So they realized they needed to be more cognizant of what's happening in the world to get people to what they want more quickly on the website. They introduced something they call the "Project Polaris" search engine, which now considers and rescores based on what's trending on social media, and it has added 10 to 15 percent more purchases—reduced shopping cart abandonment—by 10 to 15 percent. In Walmart terms, that's a billion dollars a year.

In traditional bank lending decisions, banks consider about 150 or so data points, including credit scores and whatnot. But 75 percent of individuals don't have sufficient credit data to get loans. A company in Germany called ***Kreditech*** has solved this in an interesting way: they say they've been able to understand somebody's credit worthiness by analyzing their browser history and their telephone records. If you're desperate enough to want to get a loan, and you're willing to share your browser history and your telephone records with them, they can analyze that and come up with a credit score that's even better than many traditional credit scores. They're also able to analyze the data so quickly that they take into consideration about 15,000 data points per individual and they can process or determine loan worthiness in about 35 seconds, and process the loan in about 15 minutes.

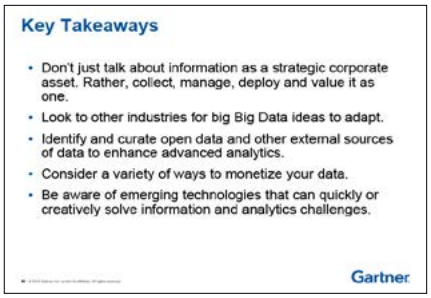
There's a company in Chicago called ***Food Genius***. Often restaurateurs and chefs open restaurants on a whim; they tend to be very creative. What Food Genius said is, "Listen, why don't you analyze what kinds of menu items are trending—what kinds of ingredients are trending—and use that to determine what kind of restaurant to open and where?" They started scouring the interwebs, harvesting all sorts of data on millions of menu items and 20,000 to 30,000 different types of food ingredients. They then make that data available to restaurateurs and chefs. They ended up getting acquired by US Foods, a food distributor, because US Foods found this data was invaluable to help their salespeople in targeting certain restaurants for certain ingredients, like, "When is white truffle oil trending in a certain part of town? Maybe this is something you want to add to your menu."

Another dark data example: ***Infinity Insurance*** is a small insurance company out of Alabama, and they realized that they were sitting on 10 years of archived claims reports. The claims adjusters will process a particular claim; they'll write up a report. The claim will either be paid or denied, and then that report—that adjuster report—gets archived. What they realized is that they could mine these adjuster reports for indicators of fraud. In doing so, they were able to identify $12 million of previously paid fraudulent claims (this is, again, a really small insurer). They were able to subrogate that money and then bake those algorithms back into their sales and marketing systems to prevent them from doing any more than that.

One more example: at the ***University of Michigan***, I judged a business information competition some years ago. Some researchers at the university found they were able to identify companies that were going to suffer subsequent years' performance woes by analyzing the complexity of the language in their annual reports. Using a fog index, they were able to say, "Yeah. These companies are trying to hide. They're trying to obfuscate bad news," and it actually correlated with subsequent years' financial woes.

Just some key takeaways from all of this: a lot of organizations just talk about information as a strategic asset—as a corporate asset. We're trying to help them go beyond that, to actually manage it and monetize it, to measure it as an asset. We implore organizations to look to other industries and gain ideas from those industries and think about how to adapt and adopt them to their own. Very important to look at this world of exogenous data that's out there—public data, syndicated data, harvesting data from the web, entering into agreements with partners and suppliers to do so.

There are certainly emerging technologies out there to creatively solve certain kinds of problems, but it isn't really all about technology. The biggest issue for organizations of the three Vs—the volume, velocity, and variety—is the variety of data. It's the biggest challenge for organizations. Volume and velocity can be dealt with by scaling infrastructure and architecture, but for the most part, it's that variety of data that becomes really challenging.



I'm going to wrap up here. Thank you for having me in today. I really appreciate the opportunity, and I don't often get to say I appreciate the work that you do. Every time I take off and land safely and travel safely, I think of you all—so thank you very much.

❖ ❖ ❖

Agenda

- Strategic planning assumptions for information & analytics
- Infonomics – the new economics of information
- Big data strategy essentials
- Information Innovation "Art of the Possible" examples
- Wrap Up

Gartner.

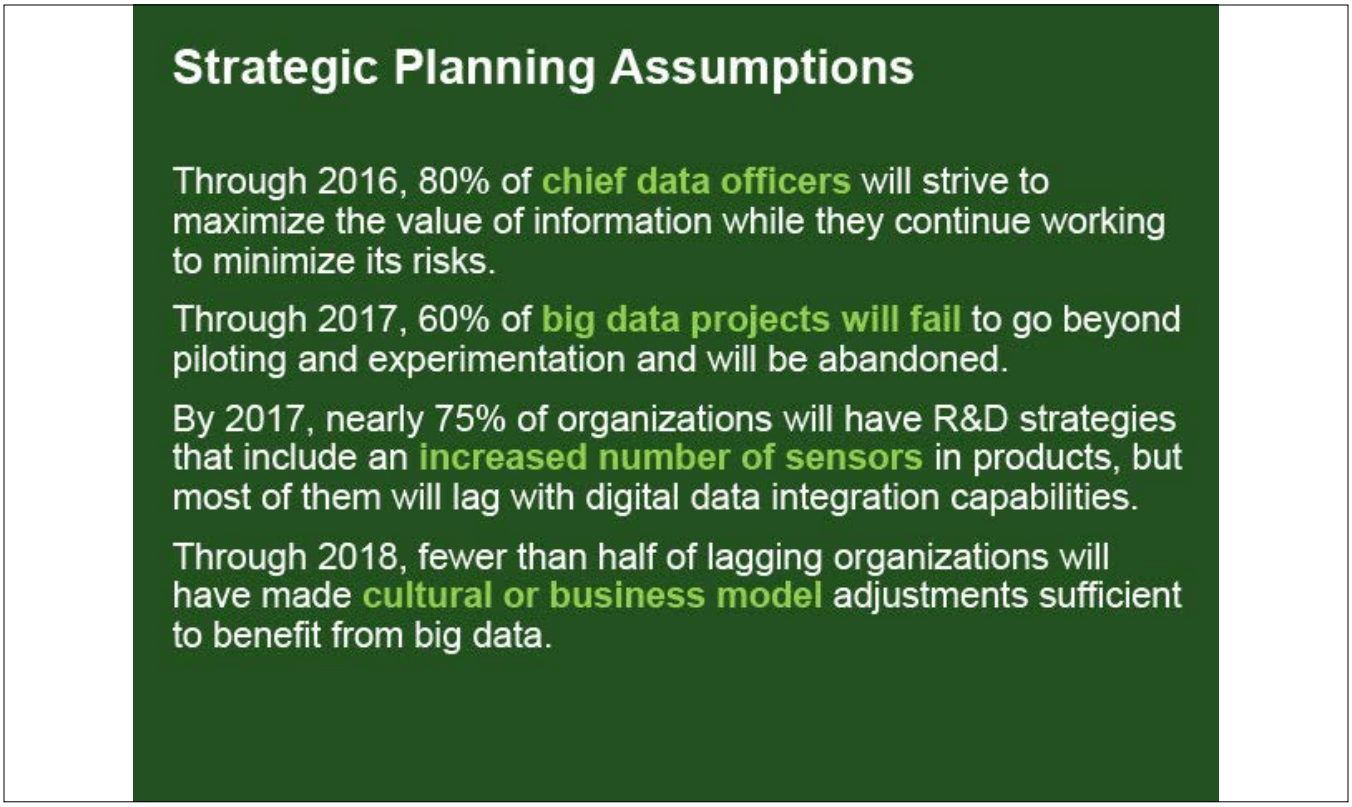
Agenda

12 Key Gartner Predictions on Information and Analytics

Produced on Worldle.net

Gartner.

Predictions



Assumptions-1

Assumptions-2

Strategic Planning Assumptions (Proposed for 2017 – D&A Strategy)

By 2019, 90% of organizations will reject vendor solutions that contractually **inhibit their ability to extract their own data**, or that monitor data.

By 2019, the importance of **algorithms** results in organizations scrambling to secure business method patents.

By 2020, 20% of companies will be either sellers or buyers of data via formal **online data marketplaces**.

By 2020, 80% of organizations will acknowledge an extreme deficiency in **information literacy** with the initiation of deliberate competency development.

By 2021, the prevalence of equity analysts valuing organizations' information portfolios in valuing businesses themselves, will lead to formal **internal information valuation and auditing practices**.

Assumptions-3

From First Responders to Worst Responders

Insurance Industry
Modified *Commercial General Liability Policy* Standard

Accounting Profession
Modified *International Accounting Standard* (IAS) 38

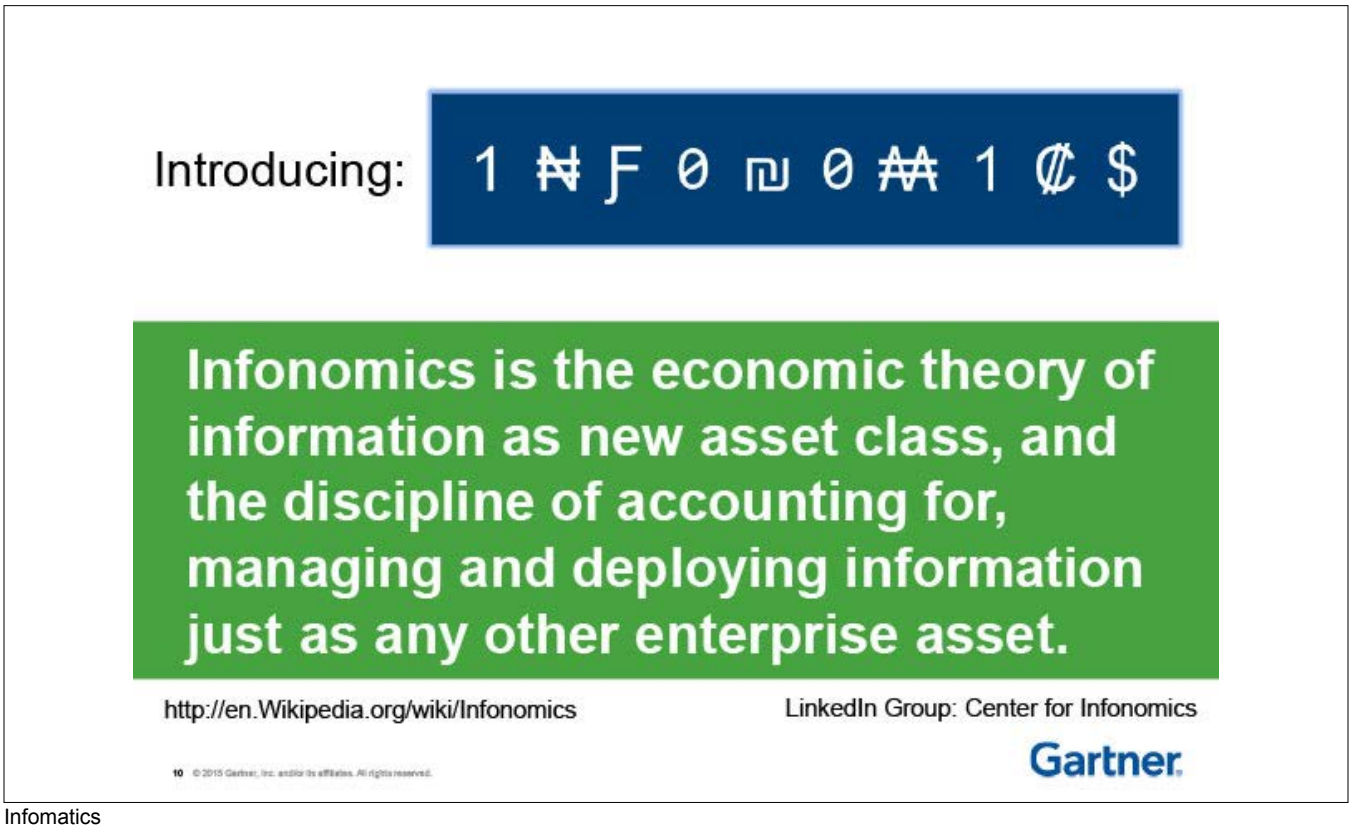
Legal System
Conflicting Rulings Worldwide

Recent Policies, Practices and Rulings Challenge Changing Perceptions of Information's Role

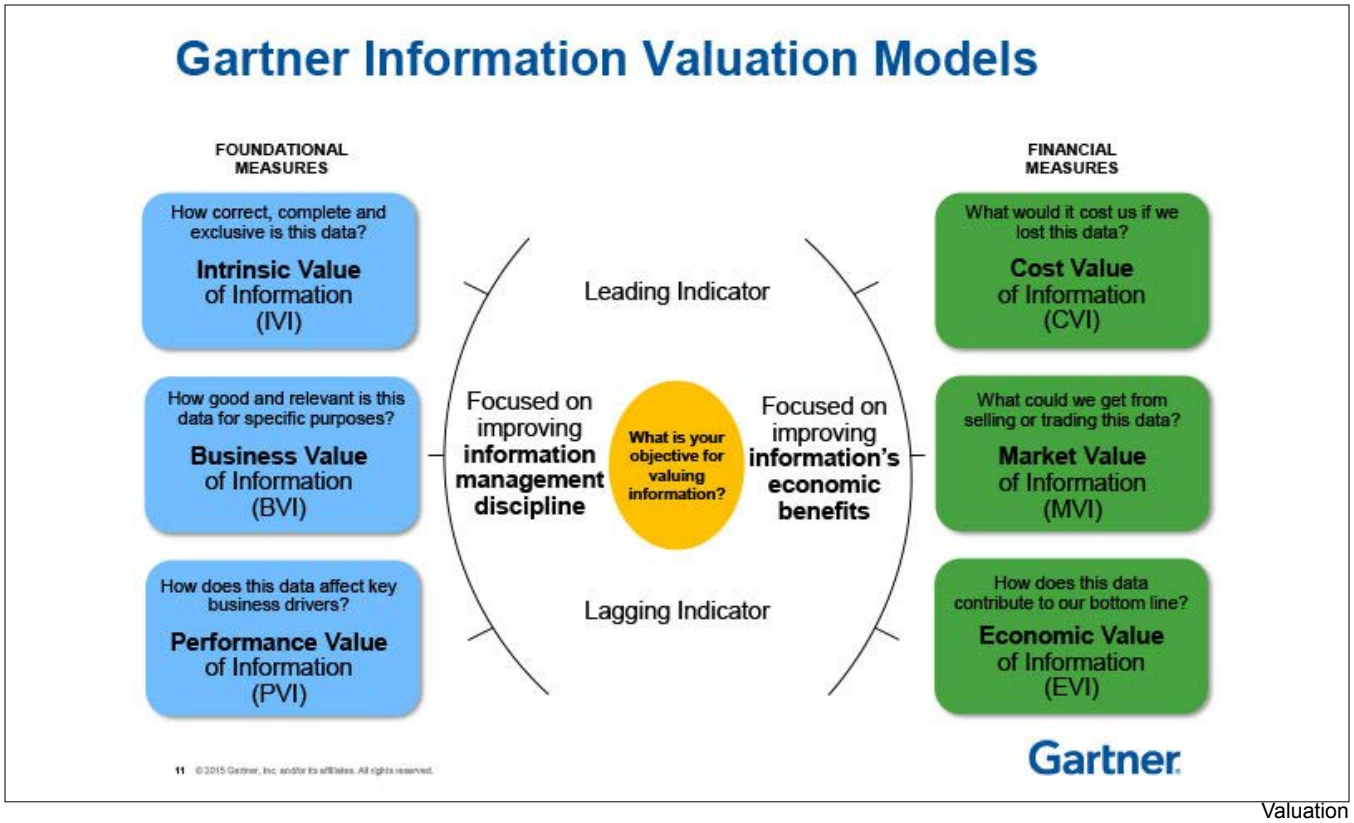
Gartner

Responders

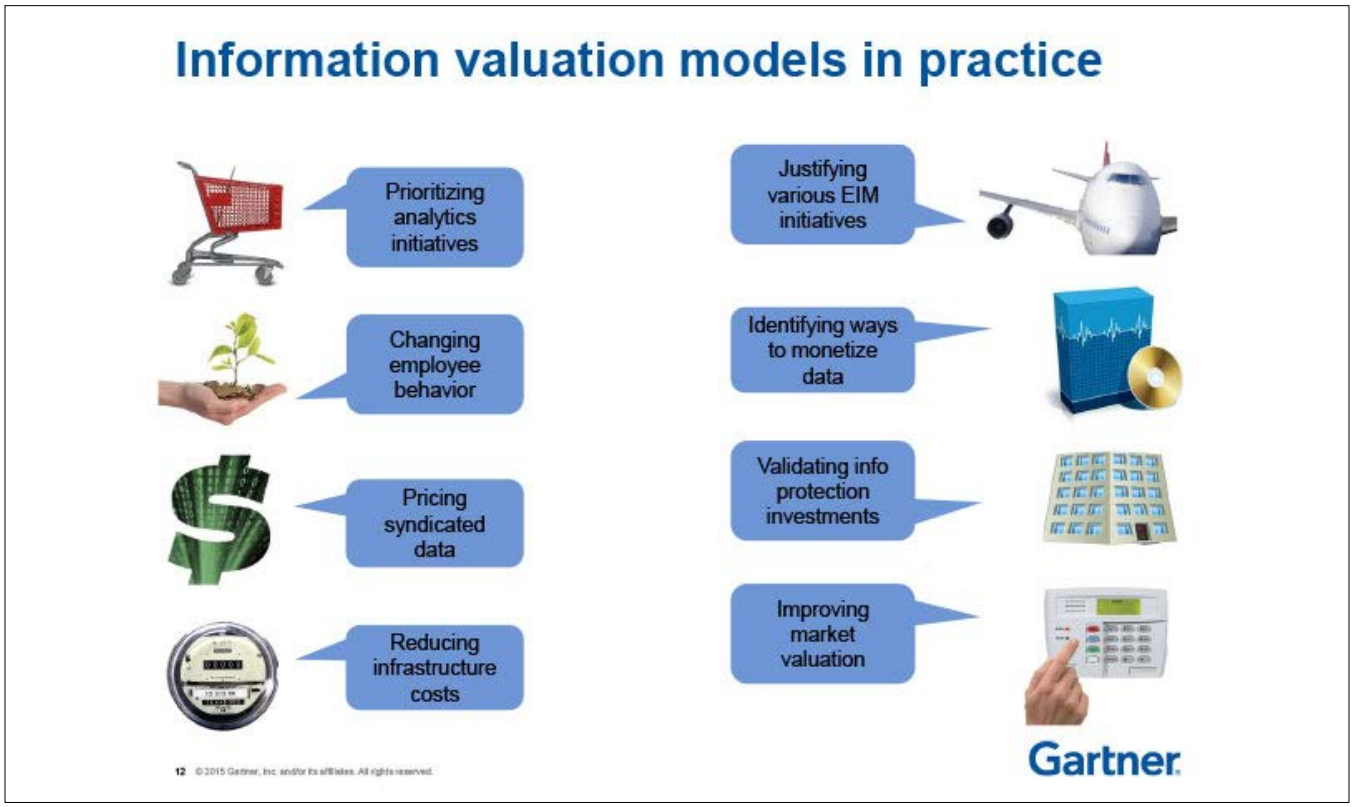
Wall Street



Infomatics

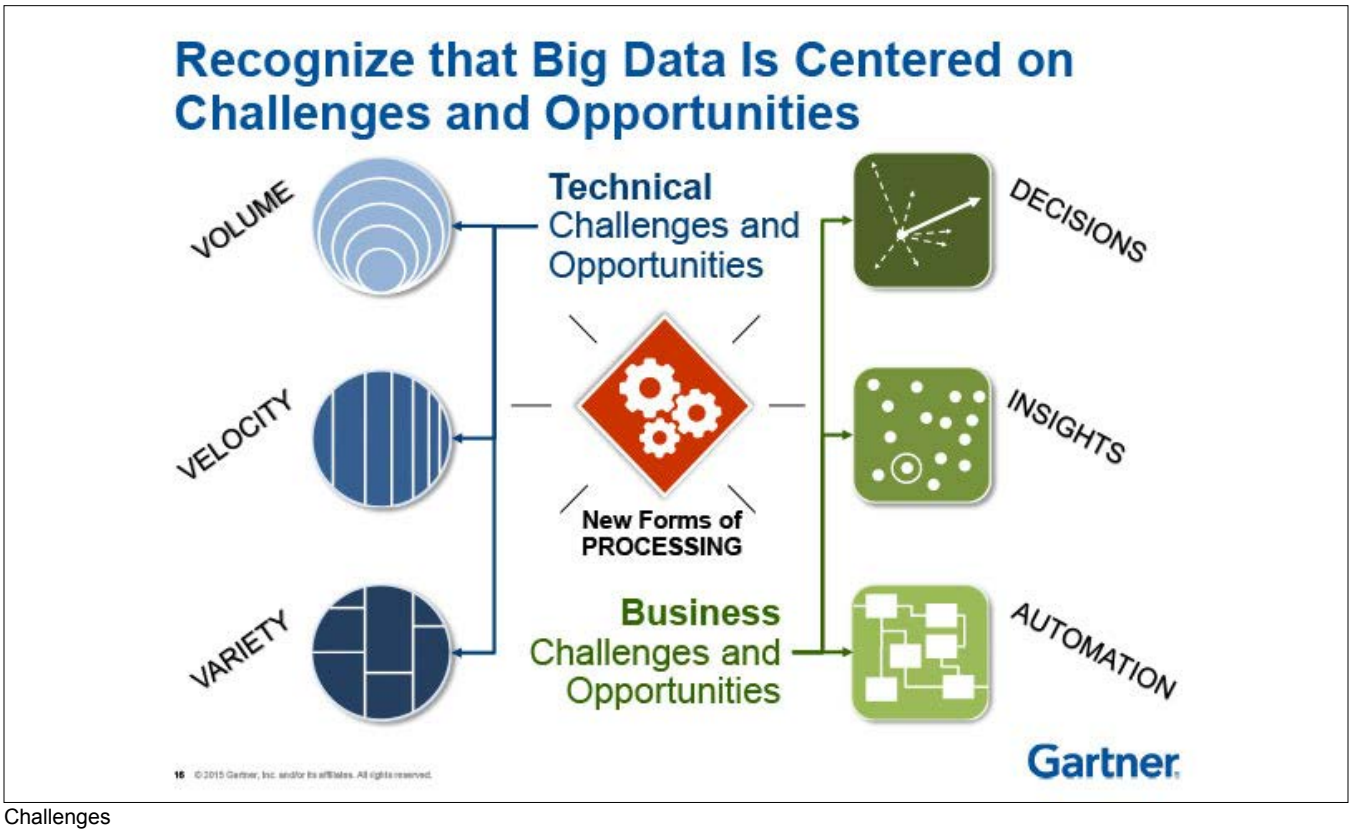


Valuation

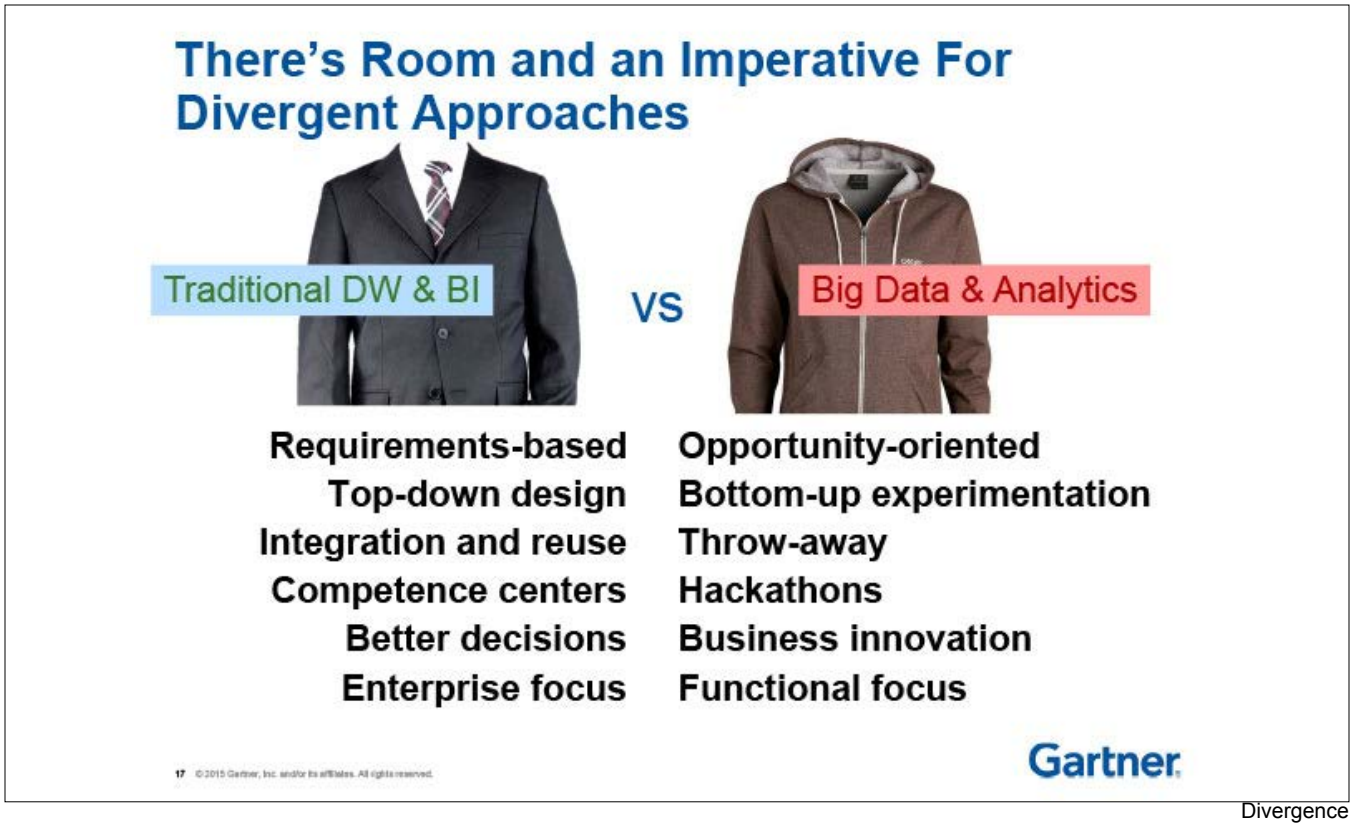


Validation

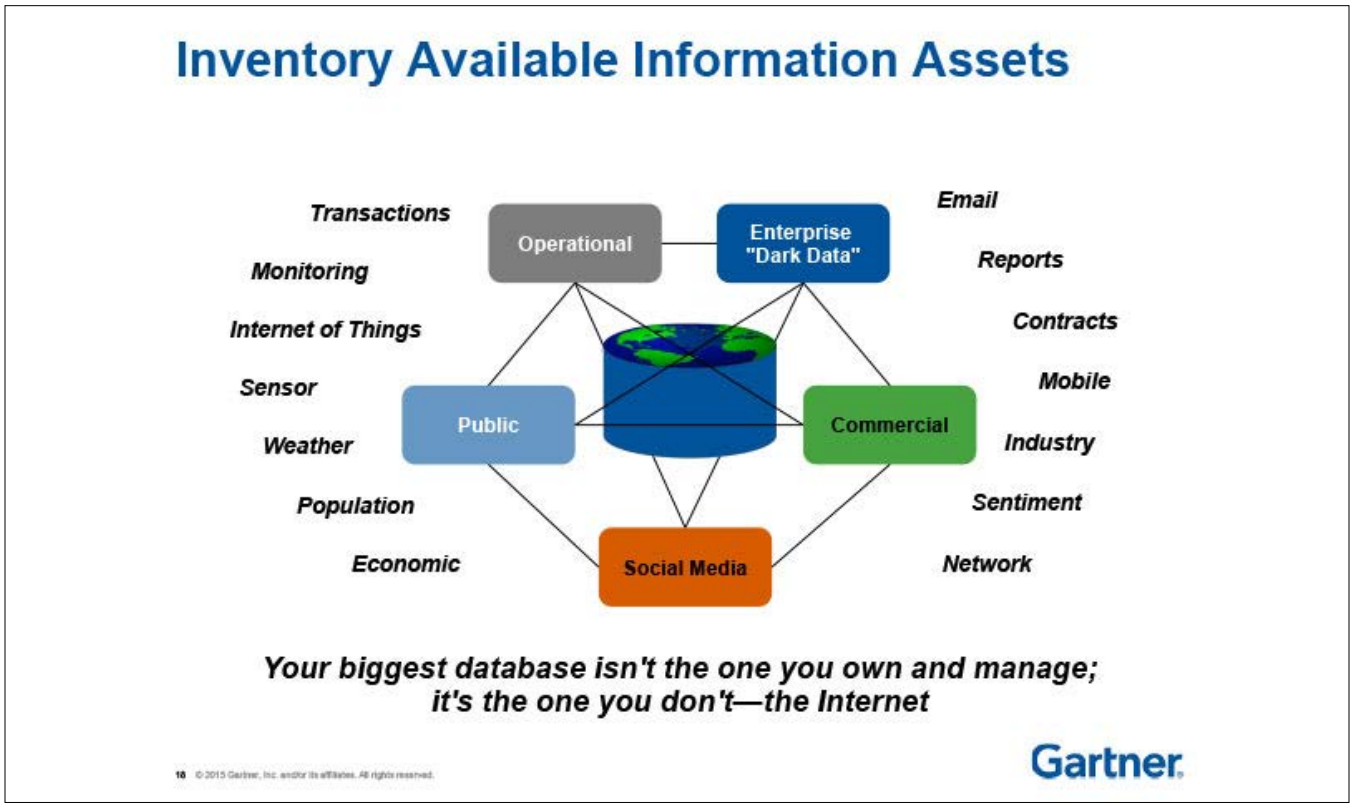
Asset Mngmnt



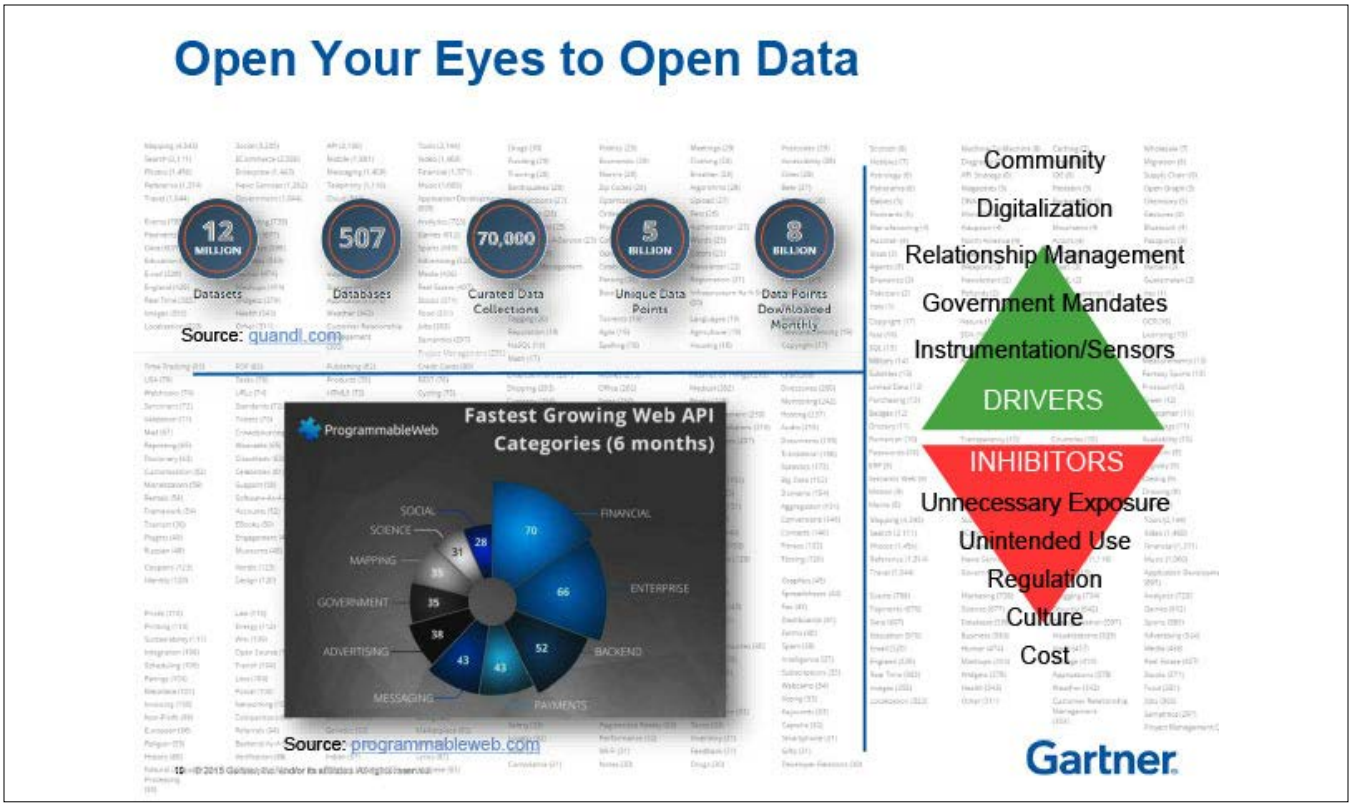
Challenges



Divergence



Inventory



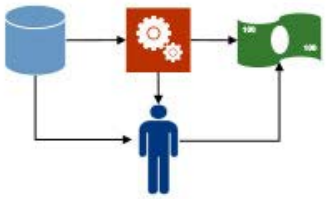
Open Eyes

Generate

Identify Ways to Generate Direct and Indirect Revenue Streams from Data

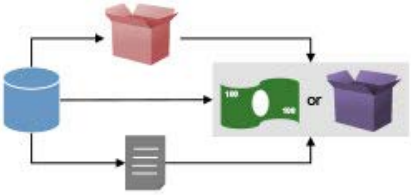
Indirect Data Monetization

- Using data to improve efficiencies
- Using data to measurably reduce risks
- Using data to develop new products, markets
- Using data to build and solidify partner relationships
- Branded indices



Direct Monetization

- Bartering or trading with information
- Information-enhanced products or services
- Selling raw data through brokers
- Offering data/report subscriptions



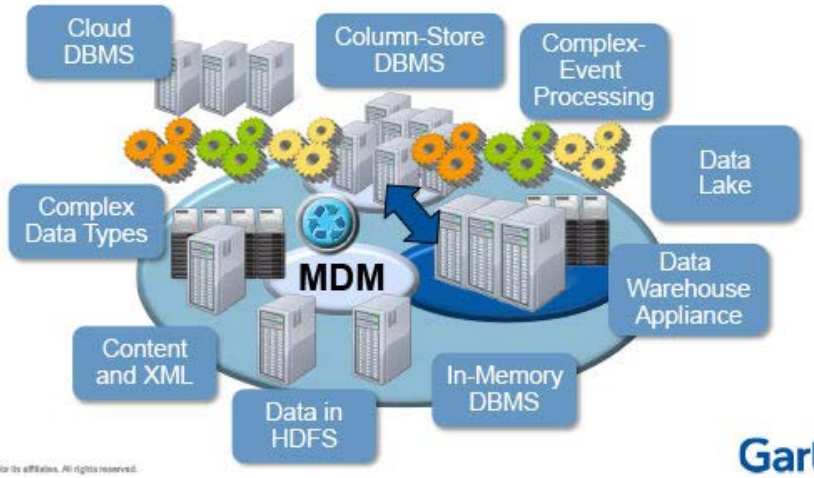
Gartner

Identify

Assume Alternate and Hybrid Information Architectures

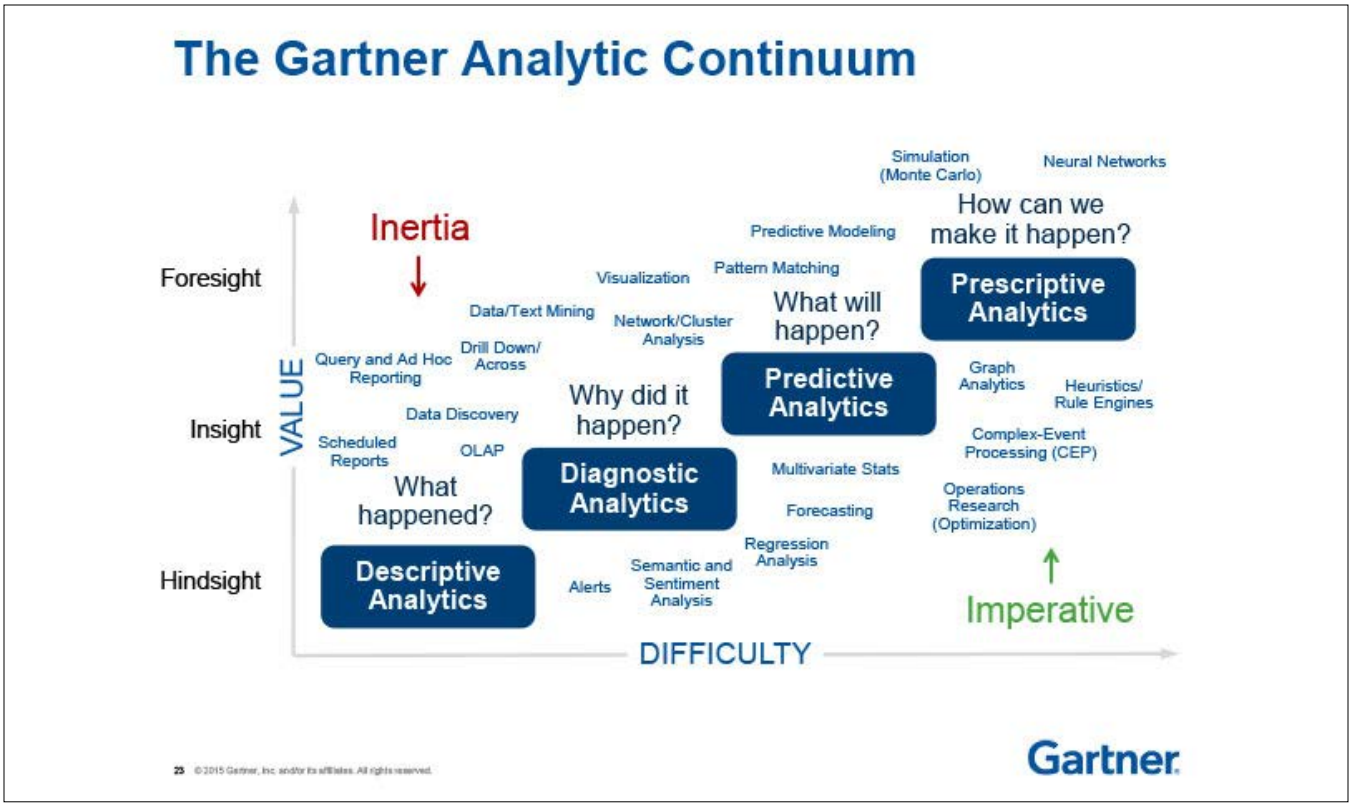
Large scale data federation has become a requisite IT competency and business differentiator

The Logical Data Warehouse

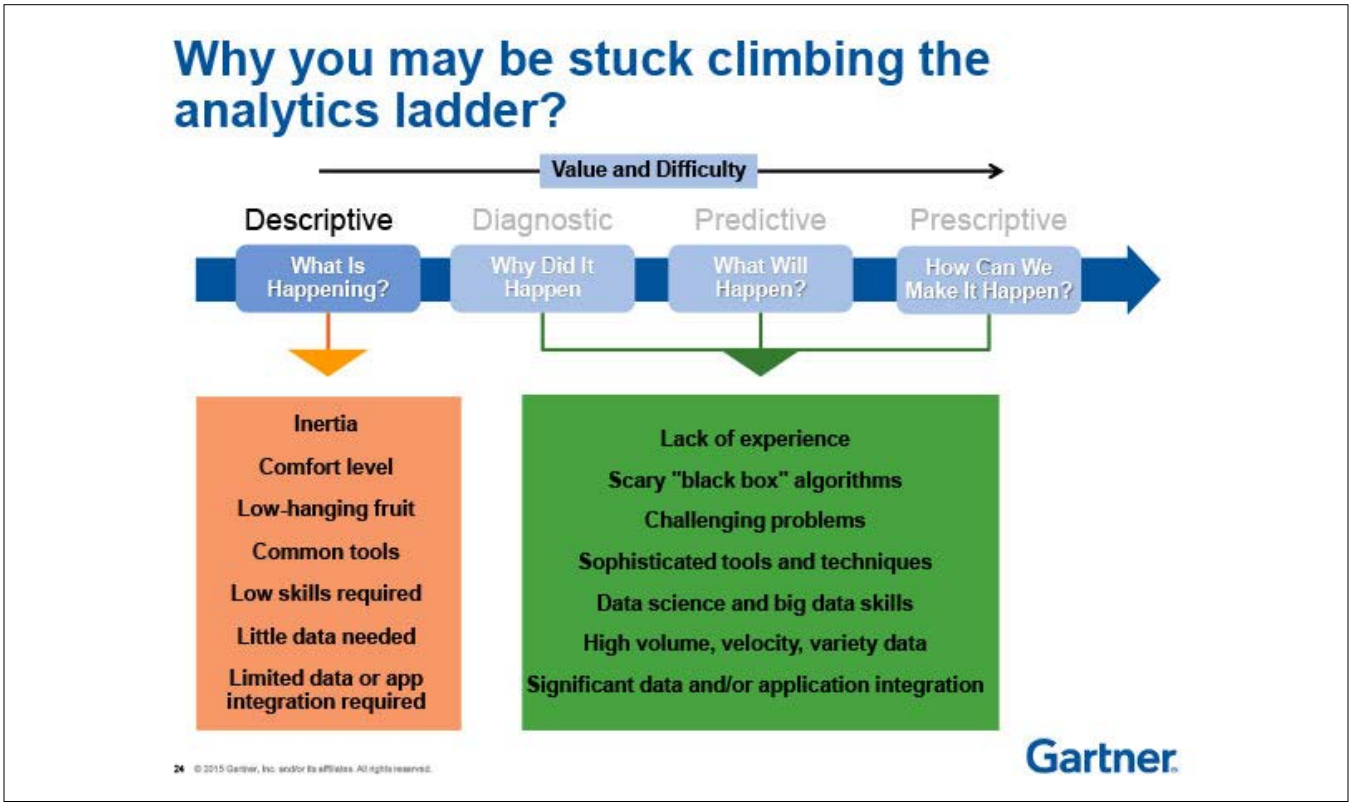


Gartner

Architectures

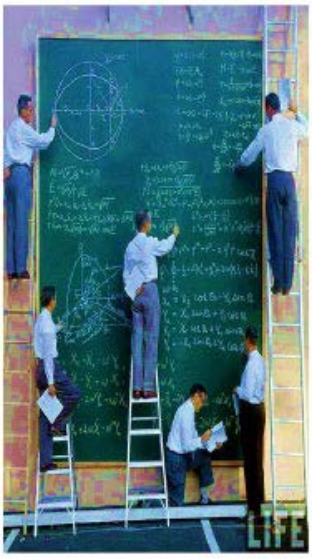


Continuum



Ladder

What Your Analytics Center of Excellence May Be Missing



- Tool Standards
- Reference Architecture
- User Engagement Procedures
- Measuring Success
- Roles and Responsibilities
- Project Collaboration
- Expert Roster
- Project Prioritization
- PMO Involvement
- Monitoring Non-ACE Projects
- Experimentation & Innovation
- Governance
- Technology Trends
- Budgeting & Financials
- Function Knowledgebase
- Information Asset Directory
- Infrastructure
- Innovation
- Communicating Capabilities and Results
- Steering Committee
- Change management
- Enterprise portal for analytics requests

Gartner

Missing

Key Takeaways

- Don't just talk about information as a strategic corporate asset. Rather, collect, manage, deploy and value it as one.
- Look to other industries for big Big Data ideas to adapt.
- Identify and curate open data and other external sources of data to enhance advanced analytics.
- Consider a variety of ways to monetize your data.
- Be aware of emerging technologies that can quickly or creatively solve information and analytics challenges.

Gartner

Takeaways