> **In few years our production data has grown exponentially.** We need tools to analyze and process the relevant information: *Big Data.*
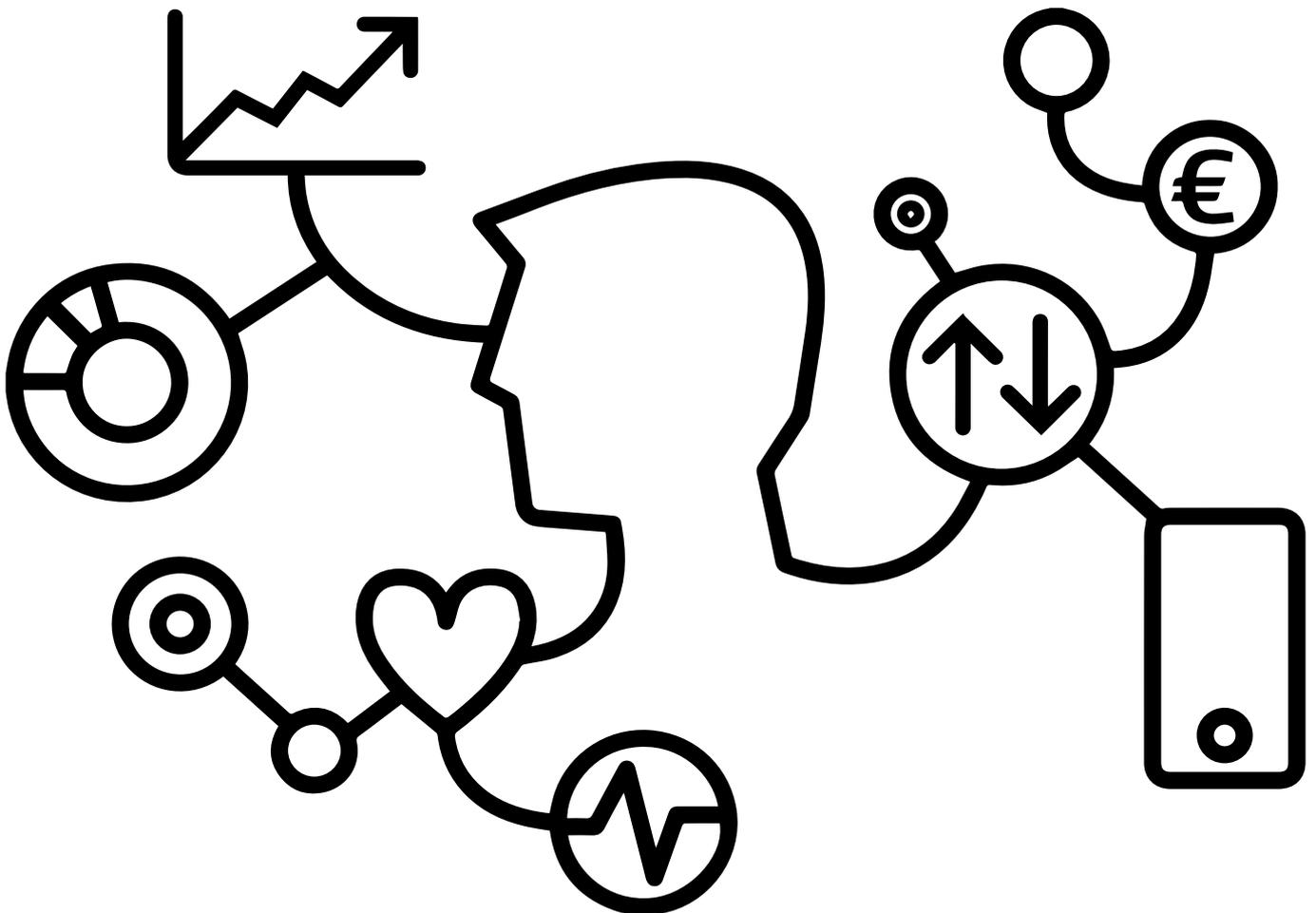
> **How are the implications of *Big Data* for us as individuals and relationships with our environment.**

> **How to improve our human performance though large volume of data process.**

Big data
The power of data

"The Big Data technology already is here, now to learn how to use it."

Big data
The power of data

Acknowledgements

Trends Forum meeting

Thank you so much
Fundación Innovación Bankinter

Fundación
Innovación
Bankinter

Index

Big data
in context

Chris Meyer. Founder Monitor Talent & Bankinter Foundation Trustee

➲ **In the 1990s,** when computerization of individual business functions was giving way to enterprise software suites, information professionals spoke of "data, information, knowledge, wisdom" as a hierarchy of usefulness.

Data was a raw material to be aggregated into information; patterns would be discerned that led to knowledge, which in turn would support insights that would make us wiser. Enterprise software, contributing data from all areas of an organization to "data warehouses," was to be analyzed from time to time to reveal insights into business performance.

But the journey took a different path: the Internet. The World Wide Web leapfrogged enterprise data and offered a shortcut to knowledge. In the past two decades much of world's accumulated knowledge

has become accessible. No sooner is new knowledge created—medical papers, travel guides, pop culture criticism—than it is publicly posted. As the value of access to the web soared, so did the incentive to obtain access, and mobile smart phones are becoming universal, delivering this knowledge almost everywhere—in the metro tunnel, and in the bush.

Consequently, the development of the information economy is returning to the earlier route. Every connected device is also a sensor, able to gather data on people's locations, machines' performance,

factories' pollutants, fields' moisture, patients' health, and individuals' sentiments, posted on social media or observed from their behavior. It is this proliferation of sensors that has spawned the spew of Big Data—vast numbers of readings, many in real time, outstripping for the moment the ability to aggregate them to the level of knowledge. But the industry is racing to catch up.

In the following publication you'll read about the explosion of Big Data, a term comprising:

> **Growing awareness** by managers, policy makers, and individuals that data is newly available to help guide decision-making. Consider an automobile navigation system, proposing an alternate route to you based on real time traffic information, generated from the rate of travel of thousands of GPS-equipped phones. Or sensors on electric meters, helping power dispatchers manage peak loads.

> **An exploding industry** offering to harness Big Data through technical infrastructure that gathers, stores, and provides access to the data itself. Continual access has replaced the batch oriented data warehouse model—urban transit systems give real-time updates on bus and train arrivals. The accumulated data readily becomes information about the performance of the system.

> And at the frontier, data science techniques create knowledge and even insights from these data. For example, a medical researcher gathered a large database of health records and loaded them into a machine learning algorithm without any particular objective; the program found that Type II Diabetes is actually not one but four distinct diseases.

And a speculation: the neurologist Antonio Damasio--a participant at the Future Trends Forum--has written that consciousness developed in humans as a way to integrate the data about the outside world brought to us by our senses with our awareness of our internal state. Big Data is adding external

awareness to our information systems, originally built to describe the internal state of our organizations.

When our smart phones, unbidden, tell us it's time to leave because the traffic between us and the event we bought tickets for last week, on-line, has gotten worse, will we call them conscious?

Big Data occupies center stage right now because it offers large, immediate rewards, driving real time decisions like the price of a hotel room for a given customer, the decision to restock inventory, or the targeting of marketing resources. But as with each advance in the information economy, the development of the human and technical resources to extract value from Big Data will in turn lead to the next stage, in which knowledge and insight are routinely generated by our information systems. The report that follows offers descriptions of the current state of the art, challenges for moving forward, and a sketch of what may be possible when we do.

Fundación
Innovación
Bankinter

Introduction

1

Our ability to produce information has evolved at the speed of light over the last few years. Because of the massive amount of available data, new tools to analyze and process data and identify relevant information have emerged. We have just started to seize the opportunities created by current technologies.

Esther Paniagua
Journalist and Author
of FTF report

◂

In the last five years, more scientific data have been generated than in the entire history of mankind.

"If you stacked the new books being published next to each other you would have to move at 90 miles an hour just to keep up with the end of the line", wrote Stephen Hawking in The Universe in a Nutshell thirteen years ago already. "If the exponential growth continued, there would be ten papers a second in my kind of theoretical physics, and no time to read them", said the acclaimed physicist.

In fact, that is exactly what has happened. According to the professor of bioinformatics at the Harvard School of Public Health, Winston Hide, **in the last five years, more scientific data have been generated than in the entire history of mankind.** The sheer volume of existing data is so that if they were stored physically, they would take up more space than a galaxy, assures Hide.

This new, accelerated speed at which we generate contents has contributed to the emergence of a new science: big data. The name is self-explanatory. However, this concept in and of itself does not suffice to explain the science, which goes beyond the sheer concept of quantity. This is more than just massive data; it involves value creation through processing and analyzing data—that is **truly revolutionary of big data** and the reason why thirty experts from around the world met at the XIII Future Trends Forum organized by the Bankinter Foundation of Innovation, the inspiration for this publication.

Fundación
Innovación
Bankinter

In the process leading up to this point, numerous devices and applications that measure, structure, process and analyze huge amounts of data have been developed thanks to new technological infrastructure designed to support petabytes of information. There are new, more polished technologies emerging constantly to improve organizational efficiency, accurately identify customers' tastes and needs, connect with governments and citizens, reinforce security and smart management in cities, carry out all sorts of forecasts and form useful **behavior patterns** for companies to gain deeper knowledge of their target public and raise awareness among individuals regarding their lifestyle, even recommending certain changes to habits in order to improve some aspect of their life & health.

This is just the beginning, we've barely started to harness the potential unleashed by big data analysis. Big data has opened the door to a

better informed, more efficient society, capable of achievements that not so long ago belonged to the realm of science fiction.

Hawking may not have time to read all scientific papers written, but he will not need to. Some big data technology will do it for him, and he will just interpret the results. In fact, he will not have to wait much: thirteen years have gone by since his forecast, and this is part of the present. Data-driven **informed decision making** is a reality today, and ignoring it equals lagging behind.

The challenge now is to find the needle in the haystack, identify what is **relevant** and apply it while engaging all stakeholders (companies, organizations, governments, professionals and citizens), and overcoming legal, structural and human barriers along the way. No-one said it would be easy, but it sure is worth trying.

Big data has
opened the door to
a better informed,
more efficient
society, capable of
achievements that
not so long ago
belonged to the realm
of science fiction.

Show me
the money

2

Turning
data into gold

Lita Sands. Global Head of Digital Transformation at Novartis

⊙ Pandora's box has undoubtedly been opened and as digital technologies continue to transform just about every aspect of human life, it is the rare corporation that can afford to sit on the sidelines. We are experiencing a revolution in how we engage with the world around us and the ensuing data is rapidly transforming staid business models.

"90% of all data generated by devices such as smartphones, tablets, connected vehicles and appliances is never analyzed...60% of this data begins to lose value within milliseconds..." **IBM** [1]

"Every day, we create 2.5 quintillion bytes of data...90% of the data in the world today has been created in the last two years alone." **SINTEF** [2]

"Data, data everywhere but not a thought to think."
Theodore Roszak

Consider the market for life insurance. John Hancock Financial recently announced they would offer discounts for policyholders who wear Internet-connected fitness trackers. With every new life policy, a Fitbit is issued to track your activity. The more you exercise, the more you save (up to 15%). They've just created a competitive advantage virtually overnight by attracting and then managing a healthier pool of customers.

Consider the pursuit of drug discovery in the pharmaceutical industry. IBM's Watson, in partnership with Baylor College of Medicine, discovered six proteins that modify p53, an important protein in many cancers, and therefore novel drug discovery, in a matter of weeks. This is notable considering that over the last 30 years, scientists averaged **one** similar target protein discovery per year.

Every aspect of our life leaves a digital footprint that can be analyzed to provide insight. Microsoft was able to mine search data to identify early clues for adverse drug events that had never been detected.

The transparency that this introduces can be both enlightening and frightening to customer and employees. By replacing paper based detail aids with iPads, Pharmaceutical companies can use closed-loop marketing data to create early leading indicators for sales down to a prescriber basis if they wish. Compare this to what existed before – either absolutely no clue because there was zero access to that data, or the purchase of very expensive and time-lagged data that could not be attributed to any individual's performance.

Imagine the reaction of the sales representative who can now be held accountable for the direct impact of their presentations. Or the brand marketer who was being measured on activity but is now being measured on results.

This shock is happening to employees across most major industries. With data come insights, transparency and accountability. With digital comes the possibility of a connected, collaborative workforce working faster, better, smarter, while scattered around the globe. But few companies are prepared for what is essentially the rebirth and re-imagining of what it means to be a corporation in the 21st century.

This next chapter will explore these themes, and how we need to take into account both the what and the how of transformation to effectively compete in this brave new world.

1: ibm.biz/BdXRSt

2: SINTEF. "Big Data, for better or worse: 90% of world's data generated over last two years." ScienceDaily. ScienceDaily, 22 May 2013. www.sciencedaily.com/releases/2013/05/130522085217.htm.

3: http://www.johnhancock.com/about/news_details.php?fn=apr0815-text&yr=2015#.VTPAjpftvsU.twitter

4: ibm.biz/BdFYLE

5: http://research.microsoft.com/en-us/um/people/horvitz/pharmocovigilance-signals%20from%20the%20crowd.pdf

Fundación Innovación Bankinter

Introduction

⊗ Competitive advantage.
Two magic words invoked by big data.

Their pledge: to create new business opportunities based on high added value data. Their means: the newest platforms and solutions to record and process business interactions and activities. And their power: their ability to dig out fresh information about organizational and industrial value chains and tackle hitherto unsolved business problems.

The likes of Google or Amazon saw it crystal clear from the start. It's not for nothing that their business models include exploiting the precious data placed on their platforms by millions of users. But they are not alone. Companies the likes of Acxiom make a business of gathering data generated on the Internet by private individuals, processing and selling them to third parties. Based on their over $1 billion in revenues per year, they are doing pretty well. Software companies are also riding the wave—Splunk, for one, made more than $300 million in 2014 with its data analysis and visual reports business.

"Telecommunication companies are also selling data, other companies are trading in Twitter data, and then there are people selling information", said the big data expert and researcher **Esteban Moro,** from the Department of Mathematics of the

Esteban Moro
Researcher at Universidad
Carlos III de Madrid

Carlos III University of Madrid. There are a myriad options and the undeniable truth is that there is a market for data, and we're just starting to become aware of the value. We are referring to the potential of big data to overhaul an industry—as opposed to simply increase efficiency—and the opportunities available at all levels of the decision-making process: organizational, process management, customer acquisition, redefinition of current research models, product development and commercialization... This is a cross-sectional change in strategic areas and how it is used to quantify, predict and change, optimize and innovate.

Quantify

2.1

In our clothes, our phone, our watch, the traffic lights and even trash containers. They are everywhere, and we don't even detect their presence. Sensors pervade our daily lives, disguised in any type of object, device and accessory. Their function is to gather information, and they are a key element of big data. In many cases, they are the first step in the quantification process. And not a petty one, by the way.

That is what **Joseph Kvedar,** the founder and director of the Center for Connected Health of Harvard Medical School, calls "the power of feedback loops". He's referring to applications that measure daily activity—such as Moves—downloadable to our smart phones, or the smart bracelets and watches—such as Fitbit or UP by Jawbone—increasingly present on our wrists and "providing objective, powerful data".

And these data can be compiled to obtain more data. For example, about how an earthquake had an impact on people's sleep in the area where it took place. "Thanks to these apps I can learn that raking leaves in my garden takes up almost double

Joseph Kvedar ▲
Director of Center for Connected Health

Maarten Den Braber ▲
Co-Founder of SHIFT

the calories than biking", says Kvedar. This is what he calls the **"power of the Internet of (healthy) things".** "They increase awareness and provide an opportunity to learn about ourselves", he adds.

Ask **Maarten Den Braber,** co-founder of SHIFT and a collaborator with Singularity University. Den Braber is a founder of the European network of the Quantified Self community, created to keep abreast of the latest self-measurement tools and apps and help people understand them. He is a self-confessed addict to **self quantification.** Besides him, many others share their measurements, sometimes as odd as a man's heart rate while proposing.

When the concept of quantification is applied to business, there is a world of data to be quantified out there—no sensors or modern measurement devices needed. Google and Amazon are not the exclusive holders of useful information about their users. However, most organizations that collect data do not intend to analyze them; but rather, they do so because their protocols require that they have detailed records and multiple data of their
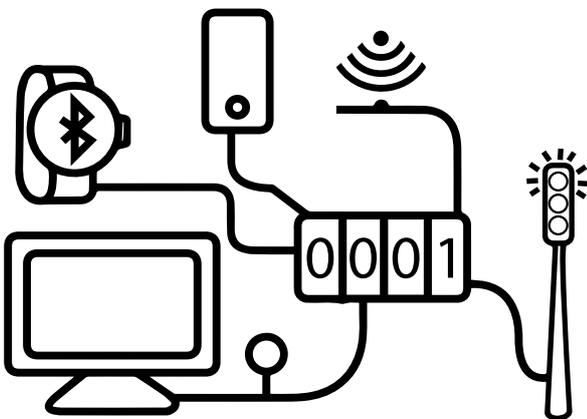
Sensors pervade our daily lives, disguised in any type of object, device and accessory.

stakeholders. Even though these data are being gathered and stored, that's the end of it most of the times: they are not processed, analyzed or used for goals beyond the sake of complying with procedures, and of course, no profit is made.

That is precisely the shift introduced by big data: the need to capitalize on the massive amount of information owned by organizations, together with new enabling tools. The banking industry and the debit & credit card sector use them already. Companies in these sectors own very valuable data about the behaviors of their users. "Payment cards provide a wide range of data about consumers and retailers", said the relevant experts at the FTF.

According to these experts, 60 million transactions on average are processed every hour, originating from close to 2,000 million cards in 220 countries and territories, through more than 40 million stores. "Such a rich, broad data set enables the detection of changes in the economy with incredible accuracy and depth", they explained. Therefore, data are also useful to present new ideas to retailers, banks, governments and other players, in order to help them make better decisions about how to deliver to their target audiences.

These experts summed up the essentials of big data in **four Vs:**

> 1. **Volume.** That is the size of the data set generated today. As a token: the data currently produced by the world in two days equal all data generated before 2003.

> 2. **Velocity** (and frequency): Data can quickly become outdated and worthless. Especially since data are created every second, including transactions, pictures and videos.

> 3. **Variety:** They may be sourced from heat sensors, cameras, smart phones, payments, cars, navigation and GPS systems, social networks, travels, postcards and others.

> 4. **Veracity:** Most data are raw and include missing or incorrect fields. The complexity increases when providers use different formats and data come from different countries, since they may be dramatically different based on local customs and habits. "Cleaning data might be the most challenging creative activity in order to create value", they assured. It is an endless task.

> **According to these experts,
> 60 million** transactions on average
> are processed every hour, originating
> from close to **2,000 million** cards
> in **220 countries** and territories,
> through more than **40 million** stores.

Predict and change

Predict and change behaviors, which makes big data a very persuasive method.

And now... What do we want all these data for? What is the point of quantifying? It is essential to know why we are doing it. There are multiple opportunities; and some of them are shooting quite high. A very powerful objective is to predict and change behaviors, which makes big data a very persuasive method. The technique must be refined. Fine processing and analysis of quantified data is required.

As illustrated by attendants to the FTF, a good job here enables capturing such "useful and interesting" data as learning that a considerable amount of consumers changed their daily shopping habits over a certain period of time. Not only that, but who they were, what their features were, where and how they did it. According to them, the usefulness of big data resides in finding out "the two to three things you had not noticed and become apparent thanks to the data". "It is a great opportunity to obtain new knowledge about consumer behavior", they say, and to predict their behavior, because **this new knowledge enables making predictions as well.**

Data may supply enough information to forecast certain aspects of customers, stakeholders, the organization itself or the market. "15 years ago, an American chain of seafood restaurants found out that satellites in the Gulf of Mexico were

relaying data about the temperature of the ocean", said Chris Meyer. "This was also an indicator of shrimp fertility, which breeds in proportion to the temperature of the water where they live", he continued. So the owners of the chain of restaurants used the data collected from the satellite to forecast the future price of shrimp, and then buy at the best price.

"15 years later, we can monitor the behavior of shellfish, fish and other animals more directly and accurately than through the temperature", said Meyer, referring to useful, current sensors. Moving on to the human kingdom, there are examples of
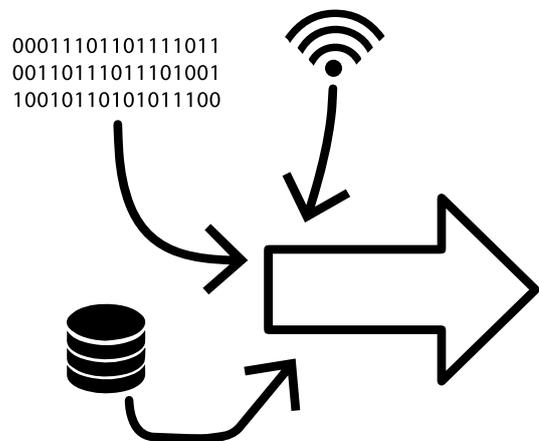
how aspects of your personality that are not directly revealed can be predicted or inferred , such as your psychological profile or the people who matter to you, even to the point of suggesting activities related to them, such as texting or phoning a friend before the soccer match every week so that they will not forget.

These applications make their predictions based on big data and using available information, but we can make predictions without the data, by generating situations that provide us with data. **Behavioral experiments** may be a good option to obtain them, according to the expert in economic analysis of UC Berkeley, **Gregory La Blanc.**

La Blanc explains that there are many possible techniques based on the needs, and he offers some examples of experiments. A web of downloads saw how the green button had a conversion rate 34% higher than the red button. A South African credit company realized that adding a female face to their advertisements increased conversion more than reducing interest rates by 2%.

Is decision-making transparent or options are manipulated? What happens when participants know it is an experiment and answer strategically? Where lies the ethical responsibility?

These experiments are obviously useful to change behaviors (green button to get more downloads) and make the highest profit on a product (increasing profits by 2% just by introducing a woman on the advertisement). This is called data-driven decision making. **"Every company in America and around the world is using or will use experiments to measure their customers' behaviors",** estimated La Blanc.

We're talking about predicting and changing consumption behaviors, but big data can be applied to many other areas. Politics, for one. Combining big data analysis, behavioral psychology and the experimental method was instrumental in the campaign that led Barack Obama to the White House in 2008 and 2012.

The political analyst Sasha Issenberg explains in his book The Victory Lab that democrats "replaced the instinct of wise advisors by a new strategic paradigm based on data". Through call center interviews and data extracted from voter registration records, consumer data bases and contacts from the last campaign, they compiled thousands of variables about their voters, such as their interests, concerns, friends or contacts. **Using statistical models, they derived electoral data down to the individual level and also predicted their vote.**

Of course, there are **ethical implications** to all this. According to the book's author, "they carried out random experiments, treating voters as unwitting

Chris Meyer ▲
Founder Monitor Talent & Bankinter Foundation Trustee

Gregory La Blanc ▲
Lecturer Haas Economic Analysis at Berkeley University

Antonio Damasio
David Dornsife Professor of Neuroscience, and Director of Brain and Creativity Institute at University of Southern California

guinea pigs". Attendants at the FTF raised several questions regarding this type of experiments: is decision-making transparent or options are manipulated? What happens when participants know it is an experiment and answer strategically? Where lies the ethical responsibility? La Blanc replied that this "game" already existed before big data, "except that these campaigns can be much more effective now".

Publicity is another industry that benefits tremendously from this science; it has given way to what is often called **neuromarketing.** It is a mix of neuroscience and data mining, involving computational learning, human intuition and big data. This enables the direct measurement of consumers' thoughts about an advertisement in order to find out its persuasiveness, taking into account that the more emotionally aligned with the viewer, the more persuasive the ad. One just needs to observe if the brain areas linked to attention, emotion, memory or personal meaning are activated by these impacts.

Establishing—even in real time—the affective response to stimuli from marketing actions is a fact. However, how comparable results are remains unclear. **Antonio Damasio,** founder of the Brain and Creativity Institute of the University of Southern California, remains cautious about this. He says it is a "brand new science" and that "we are over-confident in these systems".

Damasio explained that "it doesn't matter that we know how to provoke a reaction and the steps to follow, it would only work 100% when applied to

a machine". "Human beings are biological beings. Every cell in our brain is alive and independent: they have their own lifecycle—which we cannot control—and the brain filters everything following its own judgment—the algorithms in these systems do not know or take into account this judgment," explained the neuroscientist.

So when we talk about people, it is not as easy as saying or doing H or B to obtain an X or Y reaction. "We will not be able to convince people of doing the right thing just by informing them of the advisable option: we have extremely powerful motivations, desires and appetites to activate our reward system that counter objective information", said Damasio. This is how the neurologist illustrates the complexity of changing behaviors, and leaves the solution in the hands of his admired philosopher Baruch Spinoza: "The only really effective way of fighting a negative feeling is a more intense positive feeling".

Neuromarketing.
It is a mix of neuroscience and data mining, involving computational learning, human intuition and big data. This enables the direct measurement of consumers' thoughts about an advertisement in order to find out its persuasiveness, taking into account that the more emotionally aligned with the viewer, the more persuasive the ad.

Optimize

Move beyond performance indicators (KPIs) and control panels and embrace performance analysis and data engineering.
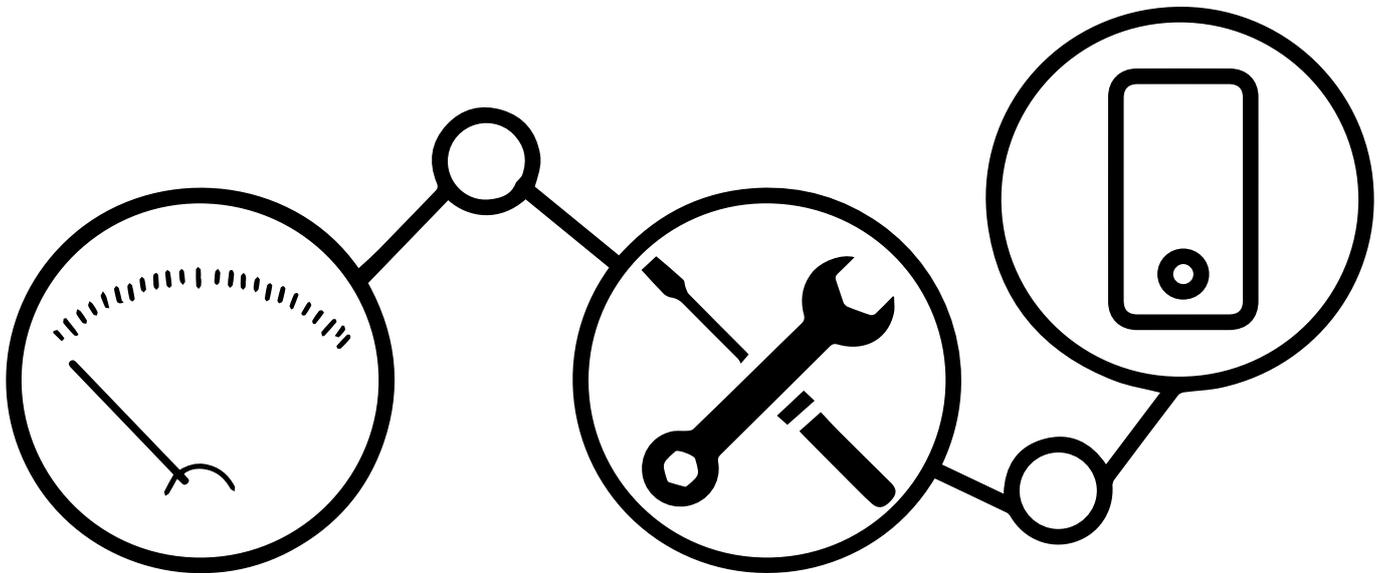
Big data predictions are also applicable to the inner workings of an organization. Data from the Internet, social media, phones and devices accurately measure our performance and use this knowledge to predict future performance, identify and assess new opportunities and risks.

These applications are here to transform the structure of the organization and make it leaner. Just in time to join the accelerating change and constant adaptation to new—and continuous—market demands. And just in time to replace outdated measuring, **move beyond performance indicators (KPIs) and control panels and embrace performance analysis and data engineering.**

Managing and supervising the execution of corporate strategy will now require evolving from the business process management methodology, thanks to the increasing amounts of available

Taryn Sullivan ▲
CEO of Efficiency Exchange

Stefan Bungart ▲
Leader of GE Software Europe at GE
Global Research

"The point is to be more efficient, save costs and make things easier for GE employees".

data at present. To this end, innovative financial management, predictive analysis and business intelligence tools must be combined with big data analysis (predictive and causal) into an integrated management platform. This platform must be able to convey performance results to interactive, visual, digital reports that are native to mobile devices.

This opens the door to new opportunities, thanks to our ability to use and analyze big amounts of data (such as web registries, data in our smart phones and other devices, statistics from search engines, etc.) and less structured sources of data (conversations in social media—Facebook, Twitter—videos, etc).

There are today examples of companies that integrate a traditional approach with innovative analysis to analyze performance and support data-driven decision making. Efficiency Exchange (EEx) helps many of these companies. Particularly in manufacturing. **Taryn Sullivan,** founder and manager, says that "it is very powerful, yet it lacks visibility nowadays: most people do not realize there are many more data about manufacturing than anything else". Results speak for themselves: "We have some **300 manufacturing plants in China that save up to 11% in electricity",** she points out.

What is the recipe? "We use data to drive change among the human resources inside the factory, interacting with all employees in all contexts (including sports and leisure time), to understand them deeply and build an interface based on their actual problems", explained the founder of EEx. "I think the human aspect is essential, although understanding the user gets lost amidst the big data conversation", she maintains.

"When people ask me what technology makes us unique, I answer that the tools are really there, but unused. This has to do with experience, with the skills of whoever is looking at the data. Therefore, the focus should not be on the technology, but rather on its **management,** on how people use data to run their factories", said Sullivan.

"We don't know what we don't know! But we do know that this is the path to generate truly powerful competitive advantages".

According to this executive, the key to run operations more efficiently in factories involves understanding the cost of electricity beyond the energy cost. "That's the type of thing that works everywhere, taking into account not only energy efficiency per se but also efficiency of resources, HR management and all utilities. It is about finding a scalable model, one you can apply not only to a factory but through the entire supply chain", she says.

Sullivan points out that the main benefits reaped from applying her method include process automation, more polished planning and factory operations, less waiting and idle time. "Unfortunately it is only applicable to certain types of manufacturing", she says, "although they include most consumer goods manufacturing processes", added the founder of EEx.

General Electric (GE) is another company applying big data to its operations to improve the internal performance of the company. **Stefan Bungart,** general manager at GE Software, part of GE Global Research, says that they have prioritized projects that "generate cost savings in-house and improvements for service companies—efficient and effective sales force, engineering design optimization, monitoring and diagnosing equipment online—without unplanned time off". "The point is to be more efficient, save costs and make things easier for GE employees", he says.

Bungart sees these projects as "a fast track to extend efficiency pervasively throughout the company". The process involves obtaining valuable data about failures or potential operational mistakes, projecting them into the future and forecast what would happen if things continued the same, and

then send the technical support team to fix it before the problem takes place. It's sort of a preventive application of big data in order to optimize operations, boost productivity and help the technical support team be more effective.

According to this GE executive, this project is possible thanks to a combination of computational learning and human interaction, so that an algorithm created by a person is constantly being improved in a feedback loop.

Bungart supports as well exploratory data analysis: "We don't know what we don't know!" he highlights. But we do know that this is the path to **generate truly powerful competitive advantages.**

Innovate
to transform

2.4

Lita Sands
Global Head of Digital
Transformation at Novartis

Big data can help GE optimize internal processes and approach consumers from a new perspective too, in order to engage in a different type of relationship that builds value beyond a specific product offering. Bungart knows this is the key to move from traditionally iterative innovation to disruptive innovation. The first step undoubtedly is shifting towards an open corporate mindset. Open to tackle organizational problems and find solutions together, leveraging on mass intelligence through crowdsourcing and open innovation. And a change in the corporate mindset to start targeting individuals and citizens, instead of customers.

That is exactly what GE has undertaken through several recent projects. For example, working together with Alaska Airlines, they set out to save on the airline's fuel cost. "When we started observing contextual data, looking for a way to reduce global delays, we found out our plan was lacking: we could not possibly do it alone; we didn't know where to begin or what we were looking for", says Bungart. "So we decided to compile as much data as possible, upload them to the Internet and **open the problem to all data scientists worldwide,** so that anybody could analyze and create their own scenario", he explained. In exchange, the person or team who

would enhance their algorithm to make reliable forecasts and optimize the problem in a scalable solution would be rewarded with $100,000.

The result: the winners of the GE Flight Quest reduced all flight time by 2.5 minutes, adding up to **$26 million in fuel and staff savings yearly,** by freeing up two aircrafts in a 120 plane average fleet. 200 teams from around the world took part in the competition through Kaggle, the biggest community of data scientists focused on crowd solutions to problems.

Along with Kaggle, GE has launched other projects that have saved costs and increased productivity, besides creating value for end-customers. Some

of these projects involve oil and gas pipeline integrity, railway performance or managing the radiation dose in healthcare. According to Bungart, the success of these initiatives lies not in the profitability obtained, but in the engagement and passion of participants. "You can see the fire in their eyes, the thrill of being part of something big", emphasizes the GE executive.

The pharma company Novartis has experienced a profound shift in its corporate culture. **Lita Sands,** director of Digital Transformation, shares her case: "I want to discuss the road we have been traveling for the last three years. In only three months since we started the process, we got rid of paper completely (and replaced it with notebooks and tablets for the sales force) and integrated all sales systems digitally. It was thrilling and terrifying at the same time", acknowledges Sands.

How did we do it? "Apple style. We realized we needed to do something simple enough for anyone and everyone to embrace the change. We looked for inspiration in newspapers—as simple as reading stories—for our launching platform. When you buy a car, you don't have to learn how to drive again. That was our motto", says Sands.

"The entire sales force shifted overnight from working on paper to receiving a system that enabled digital control over what is shared, what are doctors interested in, which should be the next transactions and what should we be doing as a marketing brand", says the top executive. "While everybody seemed to be initially delighted, when we looked at actual data at individual country and regional levels, we saw that this did not reflect their situation".

"We realized we were only trying to solve the why, because we wanted a digital transformation", accepts Sands. Providing employees with tablets and thinking that was it wasn't enough. **We had moved to action with the why in mind but we had forgotten about the how:** "Digital change is everything. People realize it is necessary and we need to make an effort to start a real discussion

We had moved to action with the why in mind but we had forgotten about the how.

We decided to compile as much data as possible, upload them to the Internet and open the problem to all data scientists worldwide.

Harvard Business School
Professor John Kotter's 8 step
process for leading change.
It involves creating
a sense of urgency (need to
approach the change), build a
guiding coalition, form a vision
to lead, communicate and
convey the strategy to follow,
empower everybody to act
accordingly, generate short-term
wins, consolidate improvements
and institute new methods for
continuous development.

in order to move from paper-based to paper-less and allow work to continue flowing". Based on this premise, they thought about how to approach change, combining together why and how.

The theoretical model to follow was Harvard Business School Professor John Kotter's 8 step process for leading change. It involves creating a sense of urgency (need to approach the change), build a guiding coalition, form a vision to lead, communicate and convey the strategy to follow, empower everybody to act accordingly, generate short-term wins, consolidate improvements and institute new methods for continuous development.

They were inspired by Burberry's case, whose digital transition was led by the then general manager, Angela Ahrendts, now vice-president of retail and online sales at Apple. "Ahrendts saw the importance of using social media and having a good CRM platform to digitalize retail stores", says Sands. "She renewed culture among employees, set up all sorts of committees to help understand change and transfer their work skills to the new needs of creating and managing contents, connecting on social media and building a community—as well as the back office", says Sands.

The supermarket chain Tesco provided another inspiring experience. In its effort to lead in South Korea without building new stores, they identified a main issue for individuals there: the lack of time to shop. So **they decided to take stores to them through a virtual shopping system in the subway system.** The system—called Homeplus—recreates supermarket shelves on the walls of the subway stations so that commuters in transit can use their smart phones to select the represented products. After shopping "on the go", the supermarket delivers the order at home. The success of Homeplus won Tesco the leadership among online supermarkets in South Korea, and the second position in the offline ranking.

Based on Kotter's theory and Burberry and Tesco practical cases, Novartis restarted the digital transition, moving from why to how. "We created

an operational committee and strong governance structure and worked with each region to focus on their main concerns", specifies Sands. "We put ourselves in the shoes of the sales force to understand their frustration, and we realized that access to data and data analysis was always a problem", she adds. So "very simple" dashboards were designed, and the first thing they did was explaining employees how to use them. "In Italy, adoption among employees went from 1% to 30% in a month, and same happened in China and other countries", points out Sands.

"We realized we were overlooking very simple human behaviors, so we focused on them to facilitate

> The transformation will continue to involve doctors, patients and employees, to continue building on this beautiful user experience. We will continue prodding the mind and reaching to the hearts of employees so that they join on board.

the transition", explains the Novartis executive. "We identified the skills, which ones were unique and useful in the future, and we basically found out two things: curiosity and passion. By working with our people and training them we are embedding "digital" into our corporate DNA", she says.

"We have now adopted a new system to monitor every click and the doctors' responses, so we are creating an indicator that correlates the feeling of the client with the prescriptive behavior", explains Sands. "Thanks to that, we have seen a 1% increase in customer satisfaction correlate with a 6.5% increase in sales, which is very powerful", she adds. Real time analytics have enabled them to **improve user experience and obtain information to continue optimizing it.** "This knowledge enables us to make data-driven decisions, and act more efficiently", says Sands.

"When you do all these things together, you can make your organization grow faster thanks to exponential opportunities presented by digitalization, which already is a strong suit of the company", assures Sands. "In three years, we have transformed the organization in such a way that 50,000 individuals are truly using their digital skills, being proactive and innovative", highlights the Novartis executive. "The transformation will continue to involve doctors, patients and employees, to continue building on this beautiful user experience. We will continue prodding the mind and reaching to the hearts of employees so that they join on board", concludes Sands.

In short, quantifying data to obtain precious information for the business, predict and change behaviors, optimize operational performance in organizations and innovate to transform and gain a truly competitive advantage—and grow benefits—are just a few superpowers currently brought by big data. However, this has only but started, **many challenges and barriers remain ahead of us,** as you may read in the following chapters.

Big data, big money. Or so they say.

From the people,
for the people

3

From the people, for the people

Vicki Seyfert-Margolis. Founder & CEO of My Own Med, Inc.

⊙ **Almost without our knowing it,** we as a society have been participating in a cultural revolution started by the internet and driven by companies, such as Amazon, Google, and Facebook, which are now seamlessly integrated into so many of our lives.

The social contract that we have entered involves the convenience of instant access to information, goods, people and connectivity in exchange for information about us. And while the value to us, as end users is great, the value to the companies is even greater. These big centralized computers are holding our data, the assest upon which the extraordinary values of such companies sit. So the question becomes- is the social contract equal and is the data from the people, which is driving the value for a few that hold the data, an equal exchange in services for the people?

One could argue that to date, this question has been evolving- the value of the companies increased dramatically as the usage of the tools spread. And this occurred almost imperceptibly to each person contributing the data. We now find ourselves in a

Is the social contract equal and is the data from the people, which is driving the value for a few that hold the data, an equal exchange in services for the people?

place where we question who gets to own and use our data, and is what we get back as contributors enough? As big data moves into healthcare, arguably the most personal of our data, this question is now being raised as a privacy issue. Ironically, our participation in online retail, social media and the app economy has already ceded a huge amount of our personal data, but health seems to be a catalyst that is calling into question this contract.

With the growing use of wearable sensors, electronic medical records, health and wellness apps and even human genome sequencing- our deepest "selves" that is what makes us who we are, now can be captured in real time through a device as simple as a watch or armband, or as sophisticated as glucose monitors and insulin pumps that now automate management of diabetes. This data tracks where we go, how much we move, how much we sleep, our heart rates and so much more, and is increasingly being combined with our medical record information to define our current health status and predict our health trajectories to assess risk. On its face, this is the ultimate big data from the person, for the person, as we can begin to dream about real time health interventions and preventative medicine in ways we never could before.

But what of the social contract? The question before us as individuals and the collective will become one of risk and benefit to our most personal decisions about our health as defined by big data. As one contemplates having big data and sensors calculating our health status, one has to ask- how much more data are we willing to hand over and in exchange for what?

Fundación
Innovación
Bankinter

Introduction

3.0

⊙ Limiting the discussion to big data analysis business applications would be very shortsighted. At the end of the day, what truly matters is the benefit people draw from big data. In fact, each of us is a very relevant source of data—object of many measurements and analysis tools.
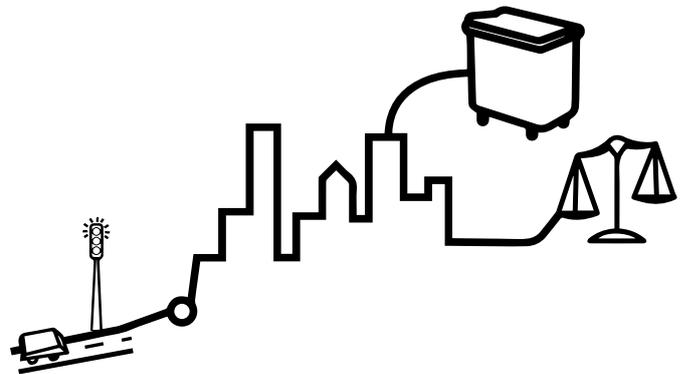
There is enormous potential for these data extracted "from the people for the people" and their interactions, ranging from health care to government or human resources, in order to optimize and strengthen our performance and uild safer, more livable cities. There are several scenarios to apply big data "from the people for the people" and their interactions, be them public officers and citizens, patients and doctors, executives and their employees, coaches and players or law enforcement agents vs criminals. During the FTF, workgroups were set up to discuss each interaction. You will find below their thoughts and our commentary.

Public officers and citizens

3.1

Technology is accelerating the ability to listen to concerns.

As the experts gathered in this group explained, new technologies based on data collection and processing empower policy-makers and city managers to listen, access and learn about citizens' concerns and needs like never before. And communication is now enabled to go both ways too. Communication can flow from the bottom up and top down, so that citizens can actively participate in collective decision making. In short: the dream of direct democracy—utopia until not so long ago—has become true.

"We hope much of the new data collected will speak for themselves and the cities will act in response", said some FTF experts. They explained that passive data collection points to specific issues (such as there are troubles in the local park), which will enable a more mature government, more receptive to needs and ready to answer. However, it also raises questions regarding how officers prioritize these concerns, and whether they will focus on the low hanging fruit or on what really adds value in the long term.

"Technology is accelerating the ability to listen to concerns. In exchange, the administration is expected to react promptly", said the experts. This pressure might lead to hasty decisions in order to solve what might be secondary or less relevant problems, versus more time-consuming concerns. Besides, citizens often come up with unfeasible or less than optimal solutions based on their lack of knowledge about the resources needed and other options, or simply because they appear to be the quickest and easiest solutions to a problem. However, officers would be wrong to yield to this type of pressure and provide makeshift solutions as opposed to long term solutions.

The need of this type of space for citizen engagement.

As mentioned at the FTF, the challenge lies in analyzing how to offer valuable feedback to the citizens regarding these important discussions, so that they can make better informed decisions and opinions regarding the course of action to follow. Transparency is a key issue. The attendants agreed that an open governmental data policy grants citizens access to information, and hence is an essential tool to make decisions. "Besides, it is a weapon against reputation-destroying rumors—a use most people do not consider" added an attendant regarding the reservations of certain social classes to open their data.

So, how could we gather all these data and devise the means for citizens to voice their concerns regarding the city? Some members of the workgroup are proposing an Amazon of sorts online marketplace applied to user-friendly public administration. "At present, in the face of a problem, we write a complaint to the mayor and nobody else finds out about it. On Amazon, when a customer reviews a product, other people can read it, identify with it and add to it", pointed out an expert. "Besides, the company is particularly sensitive to public reviews, which are normally addressed quite swiftly in matters of improvement of a product or delivery system, depending on the problem identified", he added.

The discussion regarding the creation of this type of platform heats around the proposal of having citizens star-rate political issues. "Allocating stars to political issues cannot be a good solution", said an attendant. However, there was agreement regarding the need of this type of space for citizen engagement.

In fact, there are several similar initiatives already at work. Canada or the United States provide some live examples of open government platforms. In Europe, these programs are often channeled through open data portals, which are but a fraction of what open government really is. However, there are multiple regional and municipal projects relevant to governance. Irekia in the Basque Country or the example of Zaragoza illustrate the best practice in Spain. The national open data portal datos.gob.es was created in March 2010, although there are many local and regional projects.

However, these platforms provide a poor user-friendly interface, heavy with text and cumbersome access to the information sought—far from Amazon's style. The FTF workgroup believes more efforts are needed regarding transparency and empowerment of citizens. "The alternative is to maintain the status quo, casting a vote once every 3 to 4 years. Is that the type of democratic dynamics we want?" wondered an attendant.

Doctors and patients

Vicki Seyfert-Margolis
Founder & CEO of My Own Med, Inc.

◄

Big data is part of the healthcare system reinvention.

Big data is part of the healthcare system reinvention. That is the first point raised by FTF experts in this workgroup regarding the impact of big data measurements and analysis applied to health care and the relationship between healthcare professionals and patients. "Some areas of healthcare organization and provision are undergoing great changes" There are very disruptive changes in business, patient care and the industry as a whole", they said.

"There are innovations at several levels that impact the work at hospitals and the work of nurses and researchers... For instance, data have changed the traditional patient follow up process by appointment for a continuous process through a digital platform or app where professionals can track the patient's evolution minute by minute and make constant measurements. A continuous flow of data processing takes place, and when this happens in any industry, business changes", they said.
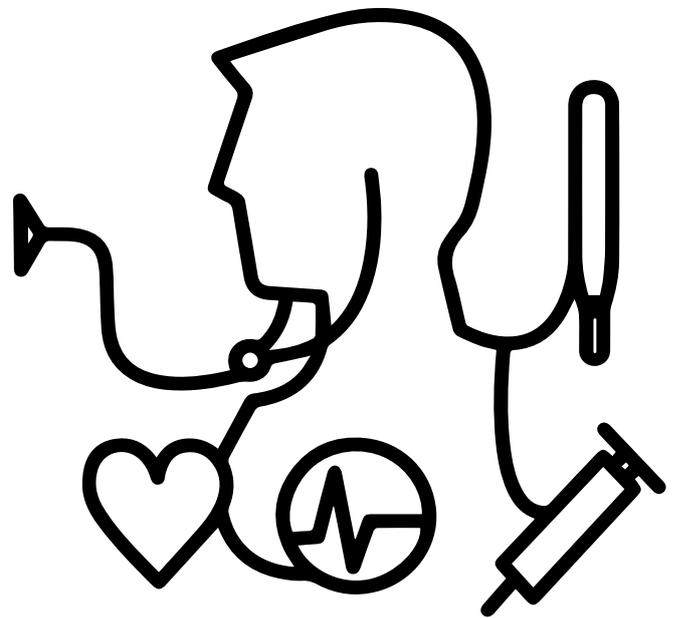
Sensors embedded into our smart phones, bracelets, watches or other wearables provide biological feedback and can be of assistance in the follow up and healing process of specific pathologies and the overall health of the patient. However, **Joseph Kvedar** believes three conditions must be met to

"Observing data is not the point, but rather, connecting the data within the patient's context, situation and actions".

successfully use these applications: predisposition to change, adoption of connected health tools and practice. In fact, this is the goal pursued by the Center he leads at Harvard Medical School: to promote an open attitude among individuals vis-à-vis new technologies in health care.

The integration of these technologies means changing the culture and behaviors among healthcare professionals and patients and the role they all play in the care. The doctor-patient relationship is affected, as well as the relationships within the family and the community. "The role of communities—both online and physical—is a critical factor of change, because data can point out a failure and enable them to act and modify unhealthy behaviors", stated the FTF group.

"This is linked to the roles perceived as important in a community. Those who play important roles are listened to by their peers, as they would listen to their nurse or doctor. That is why they must promote attitudes that improve the group's quality of life, whose cultural identity usually puts more pressure than the authorities", added an expert.

Vicki Seyfert-Margolis, former advisor to the President's Council of Advisors of Science and Technology, agrees that the community exerts great impact on what we end up doing healthcare-wise. Her experience in this field led her to found My Own Med: a platform where patients, caregivers and healthcare professionals share one interface, access the same data and work on the same page thanks to cloud architecture. "Its mission is to empower communities and healthcare professionals, connect them to their population in an ecosystem built on transparency and trust", says the founder.

"From a technological perspective, My Own Med combines sensors, big data and all available tools to provide a comprehensive vision of the patient and team management around them. Based on other big data efforts, we know these tools are useful both to show objective data and learn from the feelings and the state of an individual at all levels. This is how we can provide micro-segmented information that moves people to act", explains Seyfert-Margolis.

In this expert's opinion, "observing data is not the point, but rather, connecting the data within the patient's context, situation and actions". "This is a critical integration for the next generation of healthcare solutions. There is huge potential, and we're just scratching the surface so far", she added.

Executives and their reports

3.3

New organization and relationship models between executives, managers and employees.
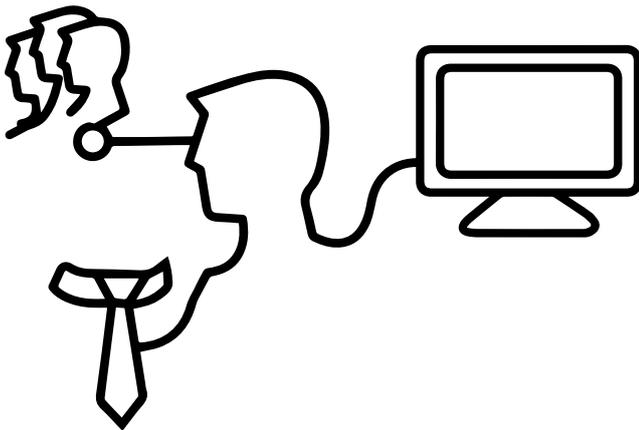
This is a no-brainer for **Stefan Bungart:** "Money does not drive employee motivation, but rather importance, value, actual meaning, and feeling part of a project that will change how the world works. That is what really drives people to give it all", highlighted the executive at General Electric.

However, traditional, highly hierarchical business structures do not address these desires and are becoming obsolete. As pointed out by the experts at the FTF, we are currently in a transition period towards new organization and relationship models between executives, managers and employees. Big data can make this transition more efficient by setting up an open data policy, for instance.

Information is power and the more power held individually and by corporate groups, the more empowered they will be to make decisions—provided that is the actual goal. Besides, self-sufficiency leads to independency, which in turn leads to greater accountability of one's actions. This self-sufficiency does not mean isolation: the point is to work individually within a group, under a multi-tasking structure based on transparent work and processes.

"How does this drive an organization's performance?" asked **Julia Li,** Chief Advisor for MIT Media Lab and founder of HCD Global, a company in the business of setting up experimental learning programs through gamification and digital technologies. "Governance used to shape companies, but the scenario is bound to change, and collective leadership will soon take the spotlight", said Li. She believes that adapting

requires "an open mind, vision of the present and the future and swift action, since data make for more efficient processes".

Li is convinced that we need to destroy structures based on individual, hierarchical power: "It is necessary in order to create a sense of urgency so that people sacrifice their own independence, apply their know-how and be accountable for the change. And promote an open mindset and collaboration". And especially transparency: "When information flows and budgets, performance and corporate actions are public, everybody can share their opinion and participate. "We must realize we are working with high skilled, experienced workers who can run the company on their own", she pointed out.

Maarten Dan Braber underscored the concepts of self-sufficiency and accountability: "It is linked to a very different attitude to leadership and management. The CEO authority model represents the company. His role is to convey to employees that they can really do things on their own, no need to ask for permission". Another attendant supported this vision of a leader as a facilitator, not as an authority, and compared it with a conductor "who conveys to musicians where they are going and extracts from them their best performance".

However, democratized decision-making is far from simple. "Even for companies with a very advanced vision of the future, this type of open-data-driven process is a big leap forward, because an open decision making process and its implications require a change in the corporate culture", said Dan Braber. It is hard, but not impossible. Zappos, the leading online shoe retailer, is a good example. Ever since it announced its transition towards a holocratic organization in late 2013, it has become an example to follow. Holocracy distributes authority and decision making in self-organized, self-sufficient teams within the organization.

This system will not only delight employees, but will also (in combination with access to massive amounts of open data) promote intra-entrepreneurship and

Julia Li
Founder & CEO of HCD Global, Chief Advisor for MIT Media Lab

innovation-based transformation both in front and back end processes. Besides, it is a precious tool to gain understanding of and keep tabs on the competition and the direction of the sector; which in turn leads to new business development ideas.

Still, there are those who believe in the value of big data but are not convinced by flatness. Some experts in attendance at the FTF highlighted that "democratization can slow down decision-making" and counter the optimization brought about by big data. "There must be some balance, some way to use data to reinforce rational decision-making—which remains the task of the CEO. The CEO must have sufficient information to know what the next step is, because they are the one held accountable at the end of the day. Besides, investors will not hold the team responsible", they said to close the debate.

Coaches and players

3.4
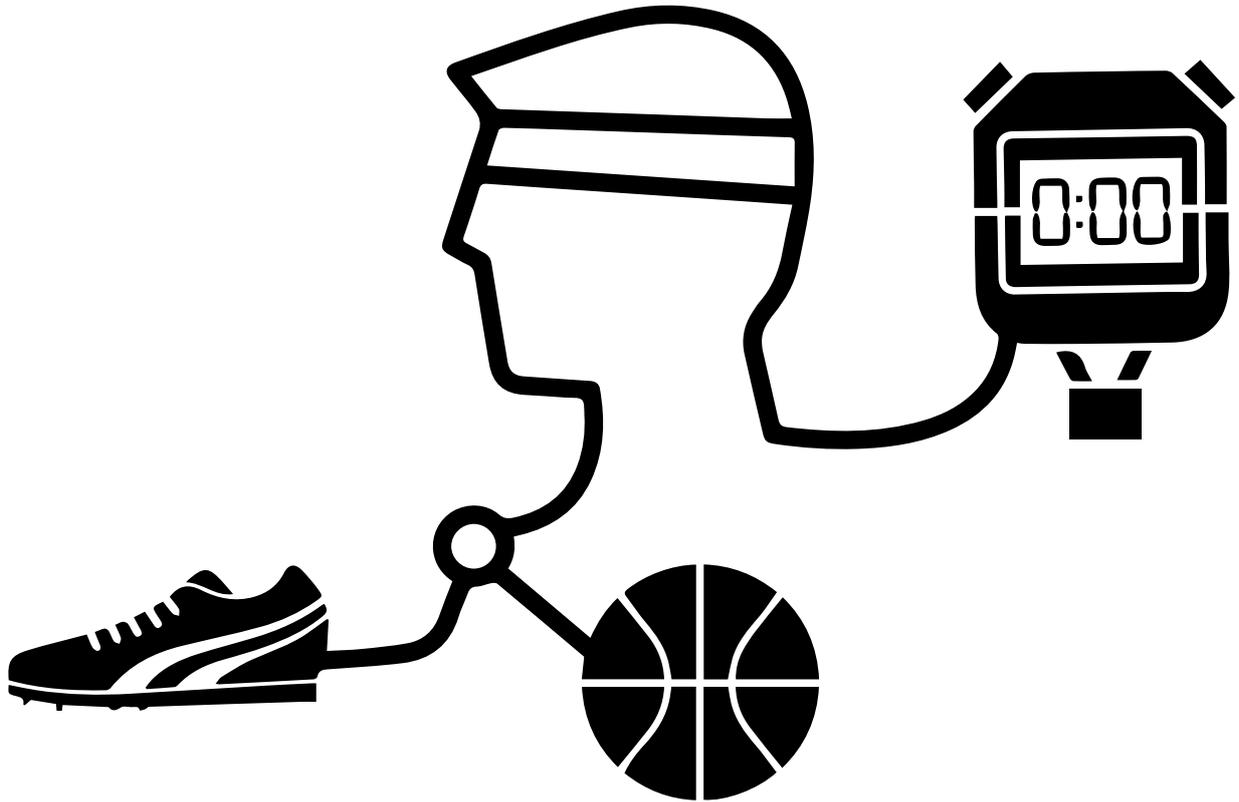
Can you imagine a telecommunications engineer training one of the best soccer teams in the world? It is not far-fetched, and the reason is big data. Analyzing game patterns or knowing the optimal plays are just some applications in the field of sports. That is why **Pablo Rodríguez,** a director with Telefónica I+D, is completely dedicated to a pioneering data project in collaboration with the FC Barcelona soccer team.

According to Rodríguez, the executives of FC Barcelona realized that maintaining the structure and each player's role on the playing field is beside the point: rather, keeping the ball when you have it or recovering when you don't is. Computational science could be of use if you conceive the game not as a group of individuals but as a network where the better the connections (passes) between the points (players), the better the performance. Therefore, new strategies could be defined for the playing field by combining data analysis and the theory of networks.

"It goes beyond knowing where the players are (or should be) to focus on where they can be. Up until now, coaches acted based on intuition and trial and error, but now they can do better thanks to big data

They can do better thanks to big data and computational modeling based on statistics. Thanks to that, passes, hotspots and what have you can be visualized in order to anticipate the strategy and react while playing.

and computational modeling based on statistics. Thanks to that, passes, hotspots and what have you can be visualized in order to anticipate the strategy and react while playing", explained Rodriguez.

This is obviously interesting to improve coaching strategies and the players' game. However, both stakeholders are reluctant to bring data analysis technologies to tactical decision making. On one side, coaches might feel somewhat threatened by a technology that performs part of their tasks and could see it as a replacement instead of a complement. On a different note, without good data

scientists and analysts that can teach how to use and apply these tools, they will be useless in practice.

"It is a matter of educating both data analysts and coaches and players. There must be a democratic relationship, which is not easy in a traditionally vertical structure. Democratization requires changing the role of coaches to some extent and making them part of the data cleaning process—which accounts for 99% of the job. And of course, they are key to ensure that the value obtained is conveyed to players", said **Wilfried Vanhonacker,** Coca Cola Professor of Marketing and Dean of the OSB at the American University in Beirut.

"They must open their minds to data and their analysis, and that requires explaining it all to them: what are data, how do you build the model… It is the only way for them to realize that they can improve their results collectively" said Vanhonacker, based on his experience working with soccer teams. He believes the focus should not be in knowing what could work: "that's easy to find out. You just need to identify any improvement in the players' performance; unfortunately the coaches of the opponents could also do that and find out your soccer playing strategy".

In short, plays can become very predictable. "Therefore, to a certain extent, we need to go back and consider reintroducing old tactics strategically". "In short, designing new coaching methods, understanding how past decisions were useful to win or lose the game", he concluded.

Pablo Rodríguez ▲
Researcher and Director of
Innovation at Telefónica

Wilfried Vanhonacker ▲
Coca Cola Professor of Marketing, Olayan
School of Business, AUB & Bankinter
Foundation Trustee

Law enforcement agents vs. criminals

3.5

Mobile devices are more vulnerable.

Technology reproduces the crime. Frank Abagnale, the most infamous impostor in history says so. Steven Spielberg made a movie about this con man (Catch Me If You Can, featuring Leonardo Di Caprio), who now works for the FBI and private companies to fight online fraud. "If you cannot beat the enemy, join them", reads Sun Tzu's celebrated book The Art of War. And war today is waged on the Internet. Thanks to the new information technologies, identity theft is exponentially easier today than when Abagnale worked on the other side of the law, as he has repeated in many public statements.

Cybercrime and computer crime are multiplying at a pace that leaves little room to counter it. "As citizens and users, we expect every company to provide trustworthy, attack-free services, but that's probably unfeasible", they said at the FTF. "Hence, the industry of cybersecurity is flourishing. Providers are specializing, and new businesses based on online security are emerging", they added. It could even become a matter of national security, as proven in late 2014 by the cyber attack on Sony Pictures Entertainment, allegedly committed by the Government of North Korea in retaliation to the imminent premiere of the movie The Interview, which mocks the North Korean leader.

While cybersecurity companies and experts have proliferated, new and creative crimes continue to sprout on the Internet. Data kidnapping is increasingly popular: it consists of encrypting files (pictures, videos or valuable documents). Ransomware, a notorious type of malware, makes it automatically and asks for the ransom digitally.
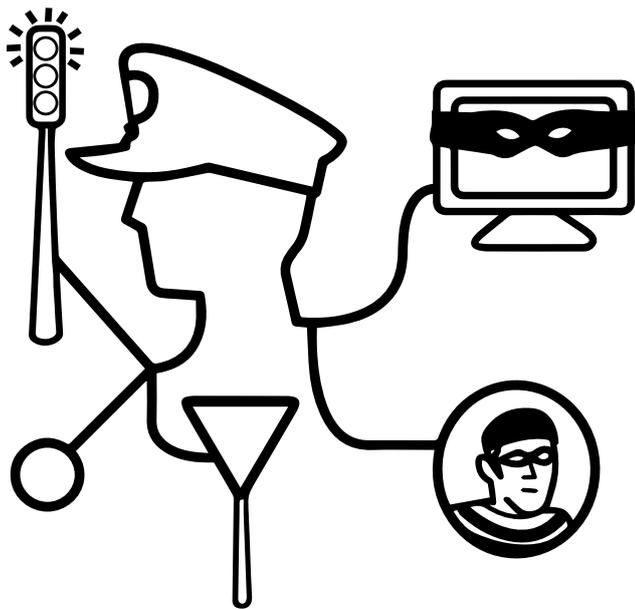
Big data enables better access to everything, and their intelligence can be used for good… or evil", was pointed out at the FTF.

If you want them back, pay up. It is complicated to prevent them, since we often lack the tools to protect from these attacks. Mobile devices are more vulnerable. "Is it secure to store data on my smart phone?" was asked at the FTF. The truth is that it isn´t. The only 100% fail proof way to avoid an attack is to avoid storing in it whatever it is that you don't want others to see or steal.

Online fraud is not the only type of fraud on the rise thanks to technology. Criminals, terrorists and drug traffickers are also using the new technologies. "On the other side of the law, data are used to professionalize, unify communication channels and create increasingly bigger criminal networks. The good news is that we now have newer, better tools to keep them in check. The bad news is that they can also use those technologies. Big data enables better access to everything, and their intelligence can be used for good… or evil", was pointed out at the FTF.

There is a mismatched fight between law enforcement agents and criminals, as pointed out by several FTF experts. There are two variables in the favor of criminals: the economics and the law. "Organized crime often deals with massive amounts of money. A trafficker might have $20 billion in cash at home whereas the total budget of the police is $4 billion". Add to this the lack of legal boundaries, which the police must abide by. "Big data without boundaries favors the development of smart, criminal super-organizations, feeding their network and supply chain endlessly. Therefore, the legal system must change to allow the exchange of data and information between all law enforcement agencies, for instance. Otherwise, they will always have the lower hand, however unfair", was said at the FTF.

Experts participating in the discussion also pointed out the need of establishing a big data brigade of sorts against crime. That is, a data analysis unit that integrates data scientists in law enforcement agencies. "Out of 49,000 policemen in New York, there are only 880 trained in data. They need so many more!" cried out an attendant. "It is increasingly urgent to respond quickly and automatically to cybercriminals. Without skilled staff to do so, it just will not be possible to counter them", they concluded.

Including a Photo of a Woman

Including a Female Face increased response rate by as much as a 3% increase in season (up to a 4 month loan)
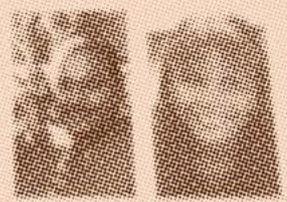
Remaining
challenges

4

The Industrial Internet

What remains to be done?

Stefan Bungart. Leader of GE Software Europe at GE Global Research

⟳ **As manufacturers** continue to seek productivity and profitability gains, continuously driving efficiencies and effectiveness of their operations, one subject in particular has garnered much attention. The industrial internet, the connection of production technology and machines (OT) to the IT environment of the company and between companies and their supply chain partners, has received a lot of attention in recent years.

There is the promise of several significant benefits from connecting machines and production equipment in this way. Access to more and richer data from every aspect of the operation and every individual process is one, the ability to update firmware and software on machines much more regularly and frequently is another.

Data can be used to better understand individual machine parameters like wear and tear, and the combination of data from several machines can be used to better understand the interaction between processes and machines. Equipment does not need to be co-located to achieve this benefit, but can be located in different manufacturing sites entirely.

Frequent firmware and software updates allow machine and equipment functionality to be updated, improved or even changed. Machines that can be so changed are generally referred to as software defined machines.

The internet generally and IP protocol specifically allow combinations of machines and cloud applications promising to deliver the productivity and profitability gains that companies are seeking.

The adoption of industrial internet applications in manufacturing today is mostly limited to pilots and proofs of concept. Most larger manufacturers are testing the water. Smaller manufacturers currently adopt a wait and see strategy. Any disruption of the manufacturing or production process immediately results in loss of production, and has repercussions throughout the wider supply chain. In some industries the loss of production from disruption can be several million US$ per hour.

The industrial internet technologies, both hard- and software are in a very early stage of development today, and are inherently more vulnerable to fault and attack. For the adoption of the industrial internet to accelerate several important improvements needs to be made.

Data Analytics and Data Strategies

While individual data applications have proven benefits in maintenance cost reduction and improved uptime of equipment, much remains to be discovered and understood in using large quantities of detailed data for productivity and profitability in manufacturing and supply chains. This requires both the data strategy experts to think through the use cases and applications, and the data analytics technologies to extract and construct the knowledge from data.

Security

Connecting machines to the internet makes them vulnerable to attack from the outside. Current IT security strategies are less able to protect critical production assets. New security strategies and technologies specifically aimed at manufacturing OT environments need to be developed.

High Availability

When the internet goes down in the office environment nobody is happy but most of the time work can continue. Networks are not built for 100% availability because of the cost involved. When the internet goes down in an industrial internet application the cost from loss of production can be very significant. The requirements in production are very different and require different technology strategies both for OT and IT technology used in companies.

Openness

The real benefit from the industrial internet comes from combining machines and data from different vendors, across different locations, and possibly across different jurisdictions. Many vendors of production equipment use proprietary data formats, communication protocols and languages and machines don't generally "talk" to each other. Different jurisdictions have different requirements for data protection and even data distribution across borders. Industrial internet applications must be flexible to adapt to this environment and the ever-changing technology landscape. Open standards and specific industry standards can help to improve openness.

Manufacturing companies continue to invest in industrial internet applications and will learn how to derive benefit from its applications. Vendors will continue to develop both hard- and software adapted specifically for the needs of the industrial internet. New challenges will undoubtedly surface and will need to be resolved. The industrial internet will not come as a revolution, but rather as an evolution. It will bring a new way of thinking about manufacturing and production.

Fundación
Innovación
Bankinter

Introduction

4.0

❯ **Imagine you are a chef,** and a new ingredient—called big data—falls in your hands. Everybody is speaking very highly of it, it is said to be a gastronomic revolution. First, you need context: where is it coming from, to what food family it belongs, how it tastes and with what flavors it pairs well.

You received it in bulk, so you will need to clean and dispose of trash data and just keep the tasty data. You must cook a nice dish with the selected data. Aided by the very best cooks, you must convince your guests that they will enjoy our data lasagna. It must be presented with charm, with layers clearly distinguishable to the naked eye.

Given the tremendous success in the kitchen, and having enjoyed our sophisticated data delicacy, you are confronted with several dilemmas: should you share it with others? Should you reveal your secret recipe for others to use—and enhance

perhaps—even though that means losing exclusivity over the recipe?

We are clearly dealing with bigger questions than answers, as our chefs at the FTF have claimed. As an attendant pointed out, it would be dramatic not to see the benefits of our stellar ingredient: big data. Fortunately it is a matter of attitude, so it is on us to win or lose the battle with the three great challenges identified: good data measurement and cleaning, correct application and balance between what is shared and what remains private, and how is it shared so as not to impair user privacy.

What is measurable and what is disposable

4.1

How can a social network, Facebook for instance, know how we feel without asking us? Once again, the answer is big data. In late 2014, a paper published by the scientific journal Computers in Human Behavior proved that it was possible to establish Facebook users' sentiments with 83% accuracy. In order to do so, Spanish researchers at the Autonomous University of Madrid have developed an application called SentBuk, a hybrid of semantic analysis techniques and computational learning.

The success of SentBuk relies to a great extent on leveraging the context: it classifies users based on variables such as the sentiment polarity in their (previously analyzed) messages, detected emotional changes or reactions to friend's feelings. This type of method is exactly what the FTF experts consider essential to obtain valuable data. "The human component is essential, data in and of themselves are worthless, because they are not neutral. It is extremely challenging to find the exact algorithm for tool interaction with human beings", said **Juan Carlos López,** from the Roche Innovation Center.

"It is essential to understand the context and cause-effect relationship", said **Emilio Mendez,** Director of the Center for Functional Nanomaterials of the US Department of Energy. Therefore, the most important part of working with massive amounts of data takes place before they are even processed. First, there is some margin of error introduced by

Juan Carlos López ▲
Head of Academic Relations & Collaboration at Roche Innovation Center

Emilio Méndez ▲
Director, Center for Functional Nanomaterials at the U.S. Department of Energy's, Brookhaven National Laboratory, 1998 Prince of Asturias Prize & Bankinter Foundation Trustee

the actual intention of the user when sharing a post. "Volume and operational and transactional value vs. declarative data. In other words, what people do vs. what they say, post or tweet", was said at the FTF.

Additionally, several participants at the FTF underscored the quality of the data above the analysis per se. "Virtually, all data are very descriptive. We need to connect them through causal

inferences to understand why superficial data are what they are. Out of context they can be alarming or misleading", was said at the FTF.

Antonio Damasio said the following about misleading data: "Devices developed so far are very simple and their level of error is considerable. For example, the reliability of blood pressure measurement devices depends on the situation: the data in and of itself is useless, you need to link it to the situation—are you getting up or resting. Measurements depend on your wellbeing, health, etc. at each point in time" said the neurologist.

That is exactly what **Vicki Seyfert-Margolis** was making reference to in her contribution to the workgroup about doctors and patients: "The idea of passive monitoring out of context is incomplete. We cannot lose the emotional component, the behaviors, because they reflect what people really need and want to use and how they connect to other people".

So we need more context and less data... And just like chefs, data scientists must also throw dirty or bland data to the trash bin. This is the most important, work-intensive part of their job. What Méndez defines as "distinguishing between the signal and the noise, what is important from what is not, what is relevant from what is not". That is, distinguishing between fine or thick, deep data. It is an arduous tasks, because the amount of data we can compile is unlimited. Working with massive amounts is a double-edged sword. "We must consider how many data are just too many, because we currently just don't know" said López.

Once we have all the necessary data—pointed out Chris Meyer—the next question is: "How much do we need to learn from these data? The answer will determine the data cleaning process. "Once again, we are facing a challenge: How can we make sure we are cleaning data really well, and not working with the disposable data. The veracity of our data depends on it" added López. He believes the quality of the data "is measured by how many predictions we can make with them". The power of the crystal ball. "The value

The idea of passive monitoring out of context is incomplete. We cannot lose the emotional component.

The experts of the FTF agreed on the need to have more staff trained on data science. They said it is essential to harness the potential of big data.

Carla Brodley
Professor and Dean of the
College of Computer and Information
Science at Northeastern University

of big data lies in creating subgroups of data based on the findings from the initial groups obtained", said the Roche executive.

Carla Brodley added a point: "IT guys hate the term 'big data': it is empty". Data by themselves are no evidence of the value. "Some companies see a need for data, and there is tension between what is needed and what these organizations are doing about it", added Maartin den Braber. "We are working on deep networks, which will become a reality if we manage to include these processes within the group guidelines that define the different user access settings of an organization's IT system" finished Brodley.

"The lack of standards and unified formats in data is a barrier. "Data are disorganized, incomplete, inconsistent and presented in different formats. Therefore, their management becomes a key skill. It requires a corporate and business strategy. And it also requires an IT infrastructure befitting a comprehensive approach to developing prediction platforms and their integration as financial KPIs, staff training and dealing with the legal aspects of data rights", summarized **Stefan Bungart.**

Another issue regarding the approach of companies— key to correctly select and clean data—is the lack of human resources highly skilled in these tasks. The experts of the FTF agreed on the need to have more staff trained on data science. They said it is essential to harness the potential of big data.

Understand and apply

These tools are very useful to convey dense, complex information, since visual information adds clarity and order.
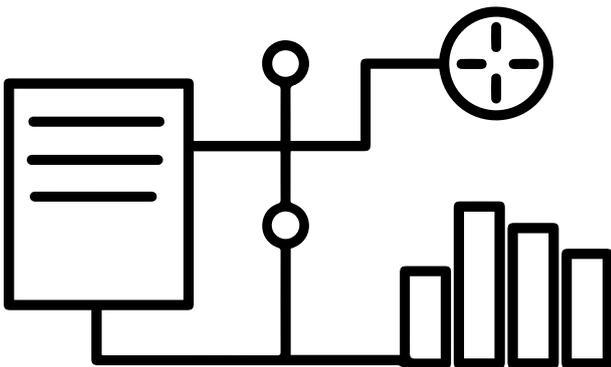
Data professionals are necessary to extract useful information from data and discard useless information. They are also essential to understand and convey them to whoever needs to apply them in order to improve in-house performance, business development or a broader transformation. "The truly important point—besides having compiled the right data—is being able to explain the results to other co-workers that will be using them", said Brodley. Otherwise, the information before our eyes might be illegible.

As pointed out at the FTF, that is the tragedy of not understanding the value of data. "You must be able to understand the value. If you cannot do it, find someone who will. It is not a purely technical issue: applying massive data analysis to drive results requires new expertise that is often unavailable in-house. In that case, it should be hired outside and employees in-house should be empowered via a broad learning program", said **Stefan Bungart.** "It is the only road to create investment-driving success stories—and the story begins where value is most easily identified and captured. This is no joke. It could result in hundreds of millions of dollars saved or revenues from one single client", added the General Electric executive.

In short, it is about transforming the business in an era of massive data by harnessing their potential and driving efficiency and cost-saving. And it requires outsourcing this complicated professional task to external consultants who will speed up the transformation. Attendants to the FTF maintain these measures are necessary to identify on what employees need to focus.

Miguel Arias
COO at CartoDB

Visualization tools complement this task. Images are much easier to process, and they are very useful in the face of the information saturation that results from managing massive amounts of data—potentially overwhelming for the non-initiated. These tools are very useful to convey dense, complex information, since visual information adds clarity and order.

One interesting application are the reports of results of big data analysis that help understand strategic results. **Miguel Arias,** director of operations at CartoDB knowingly assured that "they help business professionals understand big data". His company is in the business of massive data visualizations, they have worked on wide-ranging projects, from health care to education, banking and finance, business intelligence, governance, architecture or journalism. By way of example...

Let's say we are Real Madrid's marketing team. We want to know our impact on Twitter during the Champions League final. Let's say that instead of a long explanation describing the traffic of tweets geolocalized per country and region, we get an

interactive, comprehensive map where we can zoom in to go into more details. That is exactly what CartoDB did for the Champions League final between Real Madrid and Atlético de Madrid in Lisbon, spring 2014. And the result: a clear example of how mapping and visualization tools can be effectively used to understand data.

But these tools are not enough... **Taryn Sullivan** rubbed salt in the wound: "Now we have beautiful analytics and sophisticated graphs. But to what end? How can we apply them?" Sullivan believes that is the great challenge, especially in industrializing countries, namely China. "You must dig in the data and act", she said. And Lita Sands added that there is no rush in doing so: based on her experience leading the digital transformation at Novartis, she knows these transformations take years. "We were doing great at first, we created the right mood for change, but we got it wrong when we thought we could do it all in three months".

Sands remembers that even if they technically implemented the changes over that period—they gave tablets to the sales force and moved from paper to digital formats)—this was not the case from an HR perspective. "It is a years-long journey. It is more than considering if you want it done and then making it happen in the blink of an eye. It is important for organizations to understand they need to build the skills in the organization and yearly review their plan and messages for constant support and maintenance. And performance is often measured on a quarterly basis at corporate level", she explained.

"The only big data people are interested in involves personal data".

Make data construable, integrate humans in the learning process and do so responsibly, while preserving privacy.

There are different approaches to this type of process. According to Brodley, multitasking learning is "the biggest trend right now, it takes up a third of all scientific papers published about computational learning processes". "Let's say I'm analyzing the model of a well-managed hospital (A) to apply it to another hospital (B). I cannot do it automatically or immediately due to the multiple practices, treatments, etc. between them. At the same time, I don't want to gather all data from the new hospital (B) because it is very expensive. So I take the model I've learnt from hospital A, I gather just a few data from hospital B and I transfer the knowledge", that's how Brodley illustrated the multitasking learning model.
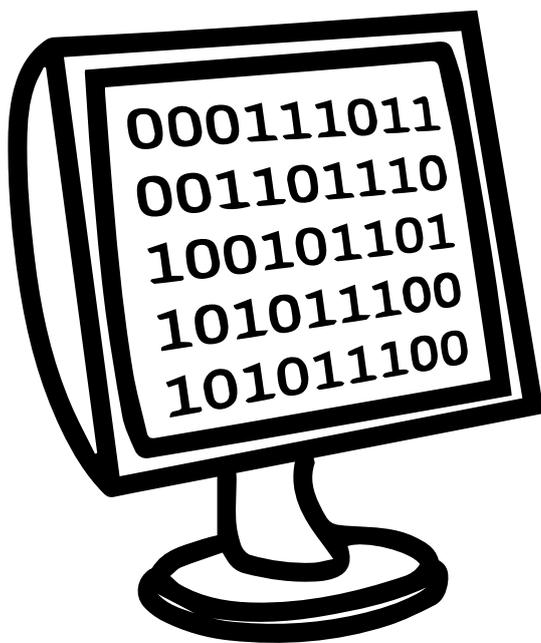
Another discussion regarding how data analysis results are applied beyond industries involves their usefulness to individuals. According to Den Braber, "the only big data people are interested in involves personal data". He's referring to obtaining behavior patterns, for example, "not necessarily to act accordingly but just to know what they are". Therefore, how people's beliefs are modified by new technologies and access to information is a big challenge. That is, how to extract the potential benefit derived from sensors and activity measuring apps already used widely and apply it to public health.

The experts at the FTF also asked themselves how data from these and other more specific applications that track the day to day of patients with certain pathologies could integrate with hospital data. "Let's say I'm a doctor and I have all these data before me, what do I do with them? What do I use them for? A better approach to the patient? Learn what works and what doesn't in the system? Where does data compiling from patients stop?" asked Brodley.

The answer is unclear, although the objective is obvious: "We must use these data to reinterpret current clinical measurements of success and include perceived improvement in quality of life, health and results that matter to patients regarding their ability to function", was said at the FTF. To this end, it is important to "make data construable, integrate humans in the learning process and do so responsibly, while preserving privacy" finished Brodley. That last point is precisely another big challenge faced by big data, as you will read below.

Property vs. privacy vs. security

4.3

"You are being watched. The government has a secret system: a machine that spies on you every hour of every day. I know, because I built it. I designed the machine to detect acts of terror, but it sees everything. Violent crimes involving ordinary people; people like you. (...) but victim or perpetrator, if your number's up... we'll find you". These are not the words of any expert gathered by the FTF, but the voice-over of the main character in the TV show Person of interest.

This situation is not the reality (yet) but it does portray the immense power of video surveillance and facial recognition systems when combined with data mining. This type of technologies—although they are not as advanced as in the series—already exist. In fact, the UK announced last summer 2014 that the police of Leicestershire would be using a new facial recognition software called NeoFace. This software compares measurements based on a picture of a face with the 92,000 faces in the Leicestershire police data base, and is "quite successful at it", they said, as gathered by the BBC.

NeoFace compares its measurements mostly with images from the video surveillance cameras already scattered throughout cities. They continue to generate controversy and heated debates regarding citizen's right to privacy and anonymity. On one side of the scale there are the privacy of individuals and the security of their data, and on the other, their physical security when a law enforcement action could benefit from those data.

Individuals and companies are reluctant to share data.

The concepts of property and its limits come to play as well. Who owns the personal data collected by governments, law enforcement agencies and other organizations? Where does their right to use them start and the right to protect one's information end? "Access to data is a huge concern, as well as privacy. Therefore, there is a conflict of interest, and abusing data is a large concern. Who must be reminded "not to be evil"? Governments? Corporations? The military? Mass media? All of them?" wondered out loud a participant at the FTF.

The implications of big data for property, privacy and security are therefore unquestionable. "I believe responsible data management requires the utmost attention from the IT community in order to preserve consumer privacy, especially in connection with patients. Most people don't understand what

Identified 90% of over a million anonymous consumers by analyzing some data from their credit card operations.

it means to click on "Yes, I like this comment on Facebook" and do not realize the implications for their own privacy. Even if they do understand them, they are carefree and reckless, because their desire to use technologies is above the resulting interferences with their privacy. But I do not believe this to be the case with medical data. We need to create algorithms to manage data while preserving privacy", said Brodley.

Health care is not the only field subject to our attention. Financial data, especially data from payment cards, are very sensitive. A team of researchers at the MIT proved so: they identified 90% of over a million anonymous consumers by analyzing some data from their credit card operations. The scientific review Science has reported about the study in "The End of Privacy".

In the light of this information, no wonder there are companies such as Abine who provide services through a system called DeleteMe, to delete personal data from the Internet. "Clearly, this is the game of big data now: privacy, individuality, and anonymity, and how to approach them taking into account that their value stems from using different sources and integrating users along the way", was heard at the FTF.

Esteban Moro questions who owns the value of data gleaned from citizens. He believes "the value is so high that society should benefit from it". "The United Nations already have initiatives so that companies must share data in cases of emergency", he said. He's referring to the program Global Pulse, which encourages sharing on big data opportunities that will drive development through data exchange alliances between the public and private sectors.

Among others, Global Pulse uses data in real-time, as well as social network monitoring, cell phones, blogs and online shopping platforms to obtain indicators that might alert about social unrest or a potential natural disaster. The objective is to prepare a suitable prevention plan, one that "could save lives", according to the United Nations.

Andreu Veà
President, Internet Society
& Member of Advisory Board to
the Internet Hall of Fame

Attendants to the FTF believe that despite programs like this, individuals and companies are reluctant to share data—either due to privacy concerns or competition/economic reasons—in the current environment of uncertain regulatory change and exchange of value. This will stop the realization for society of many benefits and promises of big data.

An attendant to the FTF said that "there are good intentions but bad regulations". "People in organizations do not understand how they can protect data while allowing innovation. While the value of sharing data through industry verticals is misunderstood, it would be tremendously beneficial for everybody", he adds. Moro and **Andreu Veà** (President of the Internet Society and Digital Champion of the European Commission for Spain) go a step further and consider that guaranteeing an open data system is essential for the public good, beyond the purely business exchanges. Arias and several other attendants agreed and talked of high-quality massive data accessible to everybody. "Otherwise, we run the risk rendering data available but unused", explained Arias.

It is a dual problem: access and regulation. Access because it continues to be highly fragmented and uneven. Not everybody has the same opportunities to access data, be it for financial, technological or legal reasons. And regulation, because there is no framework that establishes the rules of the game. "Everybody is afraid of regulation!" claimed an attendant to the FTF. Therefore, many agreed that companies must start discussing data property rights with stakeholders "before the government does".

Legal uncertainty is increasing in this area, according to Bungart, who is convinced that the legal complexity is also bound to increase. If the data gathering problem does not seem complex enough, some say that "it is by far much bigger than we can see". Therefore, the challenge of balancing property, security and privacy runs the risk of becoming a barrier to harness big data. A legislative barrier, as you will read on the next chapter.

This will stop the realization for society of many benefits and promises of big data.

A key
to the future

5

A key
to the future

J.C. Herz. Product designer and strategist engaged with Big Data

⊘ **How much** of my behavior and its artifacts do I own in toto? When does a social or commercial transaction cease to be mine.

A deep discussion of technology often reveals tensions and trade-offs that have less to do with technical systems than with our own individual and collective values. People's enthusiasm or fear about technology and its implications is reflective of their deeply held beliefs about the role of the individual vis a vis society - whether the individual is the prime actor whose freedom and agency must be preserved above all, or whether common or collective good implicitly requires public participation. Discussions about privacy, cyber-threats, government and commercial visibility into people's use of technical system, and the promise of "big data" analytics in health care hinge on this question of "how much of my behavior and its artifacts do I own in toto? When does a social or commercial transaction cease to be mine." An individualist demands decision-making authority and participation in the economic value generated by his "digital exhaust." A corporatist or collectivist views de-identified data about individuals as a natural resource to be mined and refined.

These are normative questions that drive legal and legislative decisions, business practices, and industry standards. If information is power, the decision to share it - or require that it be shared - is essentially a political decision, whether they are literally a political issue or a tug-of-war between corporations or even within a corporation. The decision that "thou shalt share data" between health providers is an expression of political will on the part of a national health system (or in the case of the US, the government payer), even more than an issue of the technical nuts and bolts. Fights and foot-dragging about industry standards or APIs are almost invariably conflicts about resources, competitive advantage, footholds for disruption vs. the entrenched position of incumbents. The fact that this kabuki theater about data plays out across domains - media and e-commerce, finance and healthcare, transportation, education, national security and the process of governance itself - underscores the fact that competing visions are driven by competing value systems. The technological landscape, as it shifts, gives us a proxy discourse for the shifting of much deeper non-technological paradigms that are far more difficult to "rev" than hardware, software, or APIs.

Introduction

➥ **As this publication** comes to an end, the loop is linked back to Stephen Hawking. He has been "the first serious scientist to consider time travel practically", he declares in The Universe in a Nutshell. If someone achieved time-travel, "he could change the course of history and master the world"— says the scientist—"but just a few of us are crazy enough to work in such a politically incorrect subject".

Jessica Bland
Senior Researcher in
Technology Futures of Nesta
◂

Crazy or not, attendants to the FTF are about to travel into the future of big data... you will not be able to escape it if you dare continue reading. The answers to great big data science questions are now within reach. Because along with the challenges we have discussed, there are several **legal, human, market, format, access and skill-related barriers.** The question is: Will we be able to overcome them? Will we attain the Promised Land where the potential of big data is realized? Will everybody get there, or just a few?

Taking the current situation into account, there are several possible milestones in the near future and in the medium to long term:

Legislation and legal framework

The new regulation passed will focus strictly on fighting fraud and identity theft.

Esteban Moro is excited at the thought that as soon as 2016, governments, companies and United Nations agencies will be pushed to an agreement to **share data in the fight against the great scourges of humanity, such as poverty,** access to food, epidemics or organized crime.

That same year, according to the product designer **J.C. Herz,** Apple will set the direction to follow once again; in this case, in the field of privacy, security and trust from mobile app consumers. How? By requesting every app on iTunes to reveal how and what data do they gather from users, and how they are monetized. A change in dynamics "more powerful than any new bill passed".

These data would be displayed on the classic "About us" tab on the iTunes Store. "Some sort of Charity Navigator [a non-for-profit organization that rates NGOs in the USA] where we will see to whom are we giving our money", assured Herz. She attaches the greatest value to the fact that "if this becomes the social norm, and everybody can easily understand what is being obtained from your mobile app, **data will be part of the equation for consumers".**

Ana María Arboleda takes up a more protectionist stance. The managing director of Atlantis Healthcare in Spain says that all types of businesses (big, small and medium sized companies) will have a Data Protection Policy by then; something that will become widespread by 2017 thanks to the **European data-protection regulation,** added **Jessica Bland.** According to this researcher in new technologies at Nesta, the regulation "will

The lack of trust due to the invasion and access to all type of data will take us back to a conservative mindset, where isolated data cells and connectivity will rise.

J.C. Herz
Product designer and strategist
engaged with Big Data

◀

Ana María Arboleda
General Manager Atlantis
Healthcare Spain

◀

standardize access to data and privacy legislation in the region, and will probably be accepted as a global standard". **Maarten Den Braber** believes the data protection standard could be transposed to national laws by 2023, "being optimistic".

Going back to 2018, machine learning algorithms will greatly contribute to control privacy issues in IT by anticipating them, says **Carla Brodley.** Nonetheless, the President Emeritus of George Washington University, **Steve Trachtenberg,** predicts that politicians and particularly the US Congress will be increasingly fearful of the direction taken by big data, and will see "Orwellian implications in the regulation of information used".

This will happen in 2020, when the digital natives are in power and they implement all sorts of measures against big data barriers, as predicted by **Taryn Sullivan.** Nevertheless, a massive terrorist

piracy act will destroy the web in 2020. The lack of trust due to the invasion and access to all type of data will take us back to a conservative mindset, where isolated data cells and connectivity will rise, foresaw an attendant.

Miguel Arias believes something along that line might happen too, although he places it in 2025: "A global campaign orchestrated by supercriminals will destroy the Internet. We will rebuild the legislation and the way we think about big data from the ground up". In this case, Arias believes that **the new regulation passed will focus strictly on fighting fraud and identity theft.**

Silos

5.2

One definition of 'silo' given by the Oxford dictionary of English is "a tall tower or pit on a farm used to store grain". The very core of big data contains equal access to data and tools to process, structure, analyze and finally utilize data. **Julia Li** considers the first milestone in this regard will take place in 2017. It has to do with the readiness of cities and companies to work together in order to **build a data ecosystem,** while individuals embrace a more open mindset towards this science and its implications.

Another task to be accomplished by the industry will be to establish programming standards and taxonomies for APIs—**Chris Meyer** calculates it will happen in 2019. In the same line, Arias estimates that wide-spread access to data tools will be available by 2020—such access is now the privilege of big companies. His view is that "data scientists and individuals will talk the same language" thanks to this.

And probably as a result of that, **"everybody will do big data on their mobile phones,** with access to open data and crowd-creation tools" probably by 2025, thinks Moro. "Working with massive amounts of data is a problem today because we need infrastructure, but this will not be a problem in 10 years time", he assured. In fact, he offered a headline: "A group of kids discover a potential treatment for an orphan disease using free data management, open visualization and crowdsourcing tools during a hackathon"

Moro seems to be right on target, since a 7-year-old girl has hacked a public wifi network in 11 minutes in the present day. That happened in this experiment, organized by the security company Hide My Ass!, as reported by mass media from the world over.

Build a data ecosystem, while individuals embrace a more open mindset towards this science and its implications.

Formats, standards and quality

5.3

Access to data is closely linked to access to a work infrastructure around massive data, which will become widespread by 2021 the latest, said the experts. According to **Julie Freeman,** artist and curator of the Culture Art commission at the Open Data Institute, this will take place before tools become free. It will be in 2017, when "open data certificates are adopted around the world and formats are aligned", said Freeman.

Two years thereafter, in 2019, **Juan Carlos López** says a **common format for medical records** will become universal "to be shared by all scientists in order to facilitate interactive care". "This will come to fruition when companies realize the cost-reduction involved and people become aware of the potential benefits, and both are therefore motivated by the efficiency and their necessity, respectively", said the executive at the Roche Innovation Center.

According to López, this will lead to an international agreement on a standard set of rules to be observed when reporting about patient health and data. This will be in 2021, and a year later, (2022), "there will be a **legal framework around what patient data must be shared by companies".**

Another great milestone will take place slightly before that: the **entire world will use one single data standard** by 2021, said **Andreu Veà.** "Just like with TCP today, which has become a great enabler". This new standard will lead to new technology that will "enable analyzing and understanding high-quality data vs. low-quality data". According to **Richard Kivel,** investor and trustee of the Bankinter Foundation of Innovation, this will take place by 2024.

Julie Freeman ▲
Artist, TED Senior Fellow, Head of Data as Culture Art Programme at the Open Data Institute (ODI)

Richard Kivel ▲
Entrepreneur & Investor: Technology, Bio/Pharma, Diagnostics and Healthcare & Bankinter Foundation Trustee

That same year, we will start enjoying a much awaited technological breakthrough: **connected cars.** Several research teams from around the world and big companies such as Google have been working on it for years. Kivel believes that higher-quality data and the advanced sensors available will enable using driverless, connected cars. "We will basically see a transformation where service providers—not individuals—will own most vehicles in operation, and insurance, parking and the like will be rendered unnecessary. Cars will become a rental service that drives you, drops you off, leaves and comes back for you or sends another car back for you", predicted the investor.

The human factor and the reluctance

5.4

Eden Shochat
Founder of Aleph & Bankinter
Foundation Trustee

What do people have to do with all this? How does their positive or negative preference contribute to the progress of this science? Facts prove that there are reasons to distrust giving data to third parties. Paradoxically, users overcome distrust when they are very much interested in accessing a specific technology or suffer the drag mass effect. This is the well-known case of social networks such as Facebook—although distrust towards these applications continues to increase. That is why **Eden Sochat,** founder of Aleph (a company that supports Israeli entrepreneurs to build scalable businesses) foresees a collaboration between Google and Facebook in 2016 "to create a start-up that will develop a **new membership-required, identity system** to analyze disclosed data and feed the Facebook graph".

Joseph Kvedar foresees another agreement with Google, in this case, with Apple. In 2019, both technology giants will agree to create a simple, intuitive interface where consumers can share

individual data from their private stock of data. That same year, another factor will facilitate sharing data too: **global connectivity**—the whole world has access.

This will "increase the capacity to provide data, available in increasingly massive amounts, for the world to work together", said Freeman. Another enabler for run-of-the-mill users will be the computational learning tools made available for personal use. Brodley said this will be a reality in 2020 and "will close the gap between scientists and ordinary people" thanks to "a universally accessible, user-friendly interface".

Another expert present at the FTF focused his forecast on machine learning too: he talked about the appearance of **"new educational disciplines** around data science and their connection to human behavior". He believes they will rise "supported by remote and self-certified learning and on the sidelines of mainstream university and school systems" which "will be a revolution for education as we know it today".

But there are always two sides to a story, and access to simple tools that enable everybody to analyze data will lead to many jobless data analysts in 2025, said Li. "These tools will end up becoming a boring toy and human intuition will take the higher hand again, minimizing the value of big data".

Skills and Resources

5.5

The need to have skilled workers immediately will give rise to a leading field in business schools and universities.

Widespread unemployment of data analysts, as foreseen by Li, would nevertheless be a consequence of a long term process starting this very year, 2015. "Competences and requirements to become a good data analyst are being defined at present", said this expert in experiment-based learning. She believes it is not a purely scientific profile; it is actually closely linked to business, e-commerce, gamification and behavioral science.

"This skill set will be helpful to better understand data, and the need to have skilled workers immediately will give rise to a leading field in business schools and universities, which will have a full range of intensive, express training programs", assured Li. Therefore, in two years (2017), some of the leading universities in the world will become well-known for their training of experts in data analysis with a focus on e-commerce and gamification.

Sullivan has a different—although compatible—take on this: the founder of EEx believes that data science "will become integrated into every discipline of study by 2018, enabling more individuals to think of ways of creating value with the information at hand and implement their designs based on human interaction with data". Brodley's similar vision predicts that "all college students will have received exposure to data science at some point in their studies" by 2019.

This interaction might be the very reason why Wilfried Vanhonacker foresees disappointment. "We will realize big data is not the beauty we all thought. "There will be some positive applications, but not on par with our expectations at this time of over-excitement in 2015", said Vanhonacker. "Being a data scientist myself, I usually say that a good data expert can create any data and make anything of it, so let's make sure we are aware of that", he said ironically.

Vanhonacker's perspective might change if Emilio Méndez's idea becomes a reality: he expects "an IBM application to be developed by 2019, enabling the extraction of the 1% of useful information from any data set and automatic disposal of useless data".

If this increases the potential applications of big data, it would make sense to continue investing in the training of data workers. As Vicki Seyfert-Margolis said, it could then happen that the United States, China and Europe sign an agreement in 2022 to create an education campaign about big data: an open policy "to enable the exchange of talent and increase the presence of valuable human resources around the world". And those people might use quantum computers, which according to Shochat, will be available by 2025 for data scientists.

Underdeveloped markets

5.6

In the chapter Show me the Money, we made the point of the existing data market. But not many FTF experts shed light on this aspect of the future. **Puneet Batra,** data scientist and co-founder of the start-up LevelTrigger, did contribute his two pence. In the medium term (2018), he envisions the development of a **corporate marketplace to buy and sell data.**

The two key enablers, Batra thinks, are the fact that "the purchaser understands what they mean and how they are going to be used" and "the seller is aware of the privacy implications". "This will lead to a smoother data market", he said.

Iyad Rahwan, associate professor of the Masdar Institute and currently moving to MIT's Media Lab, suggests something more disruptive. Rahwan is an expert in finding the needle in the data haystack, and confidently gives the following headline for 2025: **"A fully automated algorithm successfully manages a city for a whole month".** It will fulfill operational tasks: processing the payroll, finding problems and allocating resources... all this supported on big data, sensors and the available tools. Rahwan believes that strategic decisions will be the only task outside of the scope of the algorithm. However, he pointed out that "people will start questioning what type of administration is possible when this scales to other types of decision making processes".

These are the two milestones in the future of data commercialization, as seen by the FTF experts. Meyer believes that the lack of contributions reflects distrust in the value of exchanging data. However, the actual reason might be that **opportunities will**

arise from structural shifts previously described, "particularly in regulation, data access and format ", he believes. "We can expect other reactions in these fields, such as companies making things easier and big data being something understandable for everybody".

These are just a few future possibilities contemplated by FTF experts, scenarios where individuals, corporations, governments and institutions will necessarily **work together to build upon transparency and trust, in order to overcome the barriers to big data.** And—why not—to make the world a better place. Just like Freeman said, if data prove that forecasts regarding climate change, for one, are real, "everybody will need to re-focus on how they use the Earth's resources and the environment".

And clearly, things might also take a turn for the worst if data science and the sophisticated tools involved are used with criminal and terrorist intent. As we all know: technology is not good or bad, it all depends on how we use it. The future is not in its hands, but it is in ours.

The future of Big Data

The future of *Big Data*
Time line

01.Jan. > 01.Jan.
Big Data Analysis Competencies.
Competencies required for quality big data analysis are defined.

01.Jan. > 01.Jan.
"Fast-lane" Training for Big Data Analysis.
A few leading universities and business schools start fast-lane training to build talent for big data analytics.

01.Jan. > 01.Jan.
New federated identity system.
A new federated identity system augments the Facebook graph.

01.Jan. > 01.Jan.
Agreement for use of big data to fight poverty, crime and other problems.
Governments, big companies, and UN agencies sign agreement to share their data to fight poverty, food supply problems, epidemics, and organised crime throughout the world.

01.Jan. > 01.Jan.
Global Adoption of Open Data Certificates.
Open data certificates are globally adopted.

01.Jan. > 01.Jan.
Europe enacts data protection regulation.
Data protection regulation in Europe harmonies data access and privacy legislation which becomes adopted as the global gold standard.

01.Jan. > 01.Jan.
Leading university gains name as trainer of big data scientists.
One of the world's leading universities becomes well known for producing data analysis experts for e-commerce and gasification.

01.Jan. > 01.Jan.
Machine learning algorithms provide for privacy.
Machine learning algorithms become accountable for/ provide for privacy concerns.

01.Jan. > 01.Jan.
Realisation that big data is not what it is cracked up to be.
Realisation that big data is not what it is cracked up to be.

01.Jan. > 01.Jan.
Marketplace to buy and sell data.
A marketplace to buy and sell data is created that assesses the privacy impact of a trade and assesses algorithm lift of a trade.

01.Jan. > 01.Jan.
Application that identifies useful data.
IBM announces an application that allows for the extraction of 1% of useful information in any data set and automatically disposes of the useless data.

14.Mar. > 19.Apr.
Equal connectivity / internet access worldwide.

29.May. > 24.Jul.
Universal patient data / history format.
Universal patient data / history format - with tracking controls and interactive care (Alert / event driven)

24.Jan. > 27.Mar.
Congress fearful of the direction of big data.
Congress fearful of the direction that big data is going and seeing orwellian implications will regulate the use of information with an eye on application and privacy.

03.Feb. > 9.Mar.
Machine learning tools for personal use.
Machine learning tools for personal use with easy interface and accessible for everyone - closing the gap between data scientists and mainstream.

09.Jan. > 15.Feb.
One single standard.
(as it happens with TCP).

18.Feb. > 27.Mar.
International agreement on healthcare data standards.
International agreement on the standards that must be observed to report healthcare / patient data.

14.Jun. > 20.Jul.
New education disciplines.
New education disciplines / paradigm across High schools and Universities.

14.Jul. > 20.Aug.
Remote learning credentials.

13.May. > 19.Jun.
Legal guidance on what patient data must be shared by corporations.

25.Jul. > 31.Ago.
Joint big data education campaign by US/China/Europe.
Joint big data education campaign by US/China/Europe with open policy of work VISAs to enable exchange of talent & human resources.

23.May. > 19.Jun.
Worldwide data privacy design principles.
Worldwide data privacy design principles are now embedded in laws in every country.

27.May. > 02.Jul.
Shift of Citizens and Compnaies.

08.Aug. > 09.Sep.
Big data predictions.
Big data predictions for climate change meet reality.

11.Jan. > 16.Feb.
Connected Cars.
Driverless sensors; owned by service provider, not people; no need for insurance or parking.

08.Jan. > 13.Feb.
Jobless data analysts.

09.Jan. > 24.Feb.
Quantum computer for data analysis.

10.Jan. > 15.Feb.
1st fully automated algorithmic municipal government.
1st fully automated algorithmic municipal government successfully runs a city for a whole month.

10.Jan. > 15.Feb.
New global strict regulations approved.
New global strict regulations on fraud & identity theft are approved.

18.Jan. > 24.Feb.
Mobile data mining and open collaborative creation tools.

Año	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025

01.Jan. > 01.Jan.
Data Protection Legislation Passed.
Data protection policies/ legislation implemented in all kinds of business.

01.Jan. > 01.Jan.
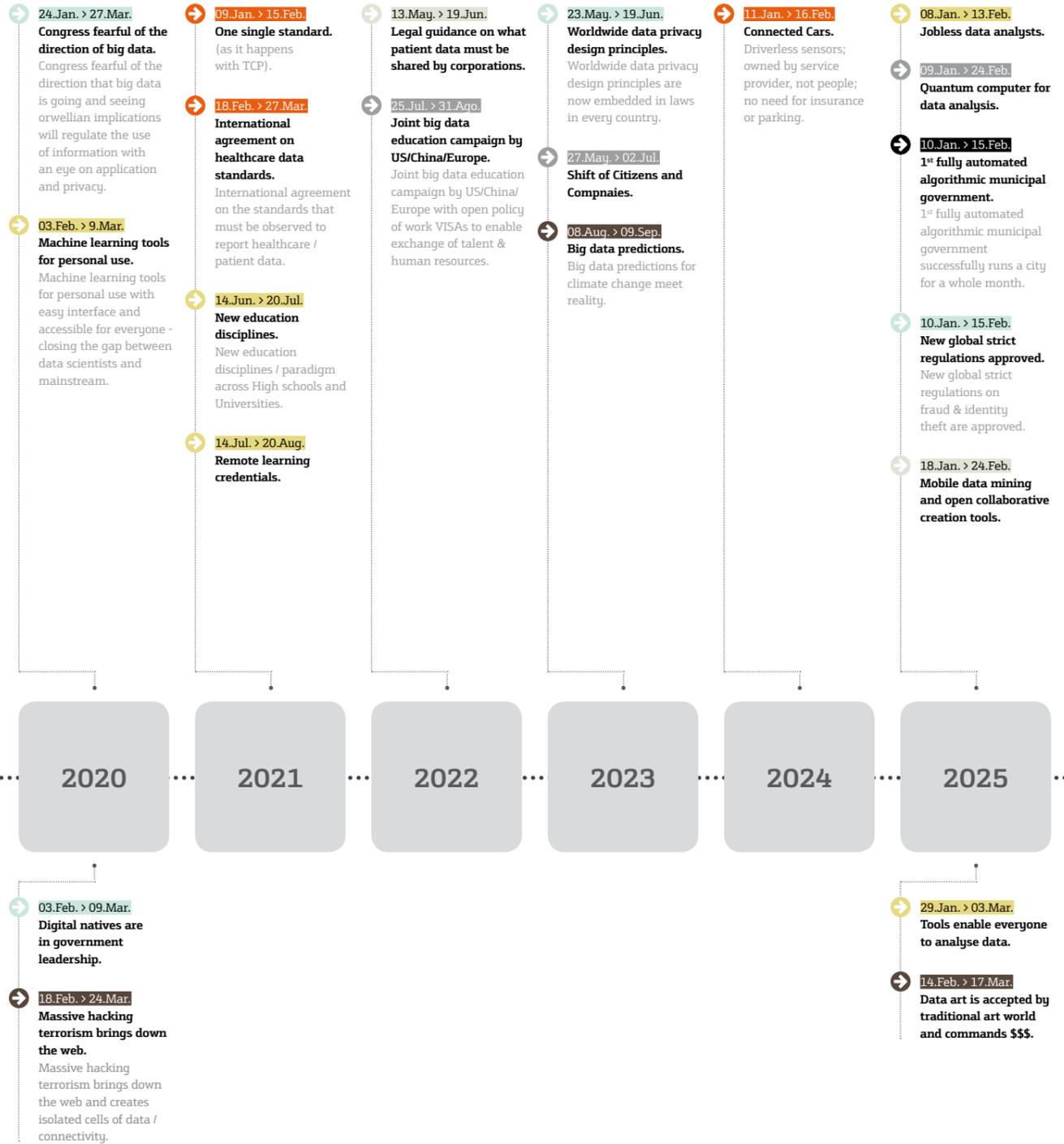**iTunes reveals what user info it collects.**
Apple requires that every app in iTunes reveal what user information it collects and how it monetises that data, on a standard "about" screen on the iTunes Store Page.

01.Ene. > 01.Ene.
Increased open-mindedness and ecosystem.
Individuals become more open-minded while cities and corporations build ecosystem.

01.Jan. > 01.Feb.
Human-data interaction design becomes part of curriculum.
Human-data interaction design education is integrated into every discipline of study, enabling more individuals to figure out how to create more value with data.

11.Jul. > 17.Aug.
All college students have exposure to data science.

03.Feb. > 09.Mar.
Digital natives are in government leadership.

18.Feb. > 24.Mar.
Massive hacking terrorism brings down the web.
Massive hacking terrorism brings down the web and creates isolated cells of data / connectivity.

29.Jan. > 03.Mar.
Tools enable everyone to analyse data.

14.Feb. > 17.Mar.
Data art is accepted by traditional art world and commands $$$.

➡ **Undeveloped Markets milestones**

➡ **Data formats, Quality & Standards milestones**

➡ **Human factor & Resistance milestones**

➡ **Legislation & Legal Framework milestones**

➡ **Skills & Resources milestones**

➡ **Silos (Access, API) milestones**

➡ **Other milestones:**

Fundación Innovación Bankinter

Glossary

6

1. **Real Time Big Data Analytics:** It is the capacity of the new technological tools to analyze big data sets as they are being created.

2. **Predictive Big Data Analytics:** They result from extracting and analyzing the historical knowledge found in massive data to outline potential patterns, results or future trends.

3. **Application Programming Interface (API):** The API is the contact point through which the IT staff can access an application without knowing how it actually works, by using the computer language.

4. **Automatic Machine Learning:** Automatic computational models based on artificial intelligence. They are launched off comparative analysis of structured, semi-structured and unstructured data. The often require human interaction with an analyst and in fact feed from it.

5. **Multi-task Learning:** It is an automatic machine learning model that analyzes several related problems or tasks together in order to extract the common, most useful points for all and enhance the starting point.

6. ***Back-office:*** In computer lingo, the back-office is the software that processes corporate information, such as data bases, records, operations, inventories, etc. It is in charge of maintaining the back-end.

7. ***Big Data:*** The group of technologies, techniques and tools to gather, process and analyze massive data sets and visualize the results. The objective is to transform the information in these big data sets into something useful—statistics, behavioral patterns, performance analysis, etc.

8. ***Crowdsourcing:*** This technique involves the masses in the quest for solutions to complex problems. From a corporate perspective, it requires disclosing the corporate data necessary for the project to outsiders. Crowdsourcing is an application of open innovation.

9. **Dashboard:** A screen that shows all relevant KPIs in one place. It outlines an overview of the indicators and facilitates the identification of trends.

10. **Dirty or Thin Data:** These are big, useless data sets that do not add valuable information when analyzing big data.

11. **Clean, Thick or Deep Data:** High quality data that will determine the results of the big data analysis.

12. **Group Policy Object (GPO):** A collection of IT infrastructure settings following each organization that define what users can do with their equipment in a computational environment.

13. **Gamification:** Applying game theory techniques to interactions among several groups (internal and external) in an organization, seeking to draw their interest and motivate them. Some examples are

adding rewards to encourage action before an open challenge, using status and activity indicators with friends and colleagues, classification tables and the possibility of jumping levels, performance data, progress bars, etc.

› **14. Hackathon:** It is a very intense programming marathon that lasts many hours. The objective pursued is finding solutions–such as an API prototype—to one or more open challenges.

› **15. Holacracy:** It is an organizational system that distributes authority among several independent groups or circles that identify, manage and solve objectives and challenges independently, although in tune with the rest of the organization.

› **16. Key Performance Indicators (KPI):** These measurable parameters set the objectives of an organization and measure and control the strategic and operational performance of a company over time. This is how shortcomings are identified and the potential for improvement is visualized in a business process.

› **17. Open Innovation:** It is a type of innovation based on integrating individuals and communities external to the organization into its innovation chain in such a way that it utilizes external resources to find the fastest, leanest way towards new and better solutions to corporate issues and challenges.

› **18. Business Intelligence (BI):** It is a set of methods to improve decision-making in business by using a fact-based support system. When applied to big data, it can be defined as a technique followed to make data-driven decisions in a business.

› **19. Intra-entrepreneur:** A new term that describes entrepreneurs within an organization. They

proactively contribute with ideas and in fact implement them from conception to realization.

› **20. Malware:** Malicious software designed to damage a computer system.

› **21. Responsible Data Management:** Using data while respecting the privacy of the individuals who shared it.

› **22. Data Mining:** the science that finds the needle in the big data haystack and extracts value from the massive amount of information.

› **23. Deep Networks:** Computational development system that mirrors the way deep neural networks work, ultimately enabling a machine to detect, identify, remember and respond just like a human brain. They learn based on data representations and models learnt from these representations.

› **24. Transmission Control Protocol (TCP):** The Transmission Control Protocol is a system that uses common protocols so that different IT systems can exchange information over the Internet.

› **25. Data-Driven Decision Making:** It is the real-life application of the knowledge gained from the big data analysis results. It is the final link in the chain, after having extracted the value and analyzed the data. For example, if it shows changes in the purchasing patterns of consumers and forecasts certain trends, the decisions made will include adapting to the said changes and anticipating the realization of the trends forecasted.

Resources

7

7.1 Reports and Studies

> The Report 'Are you thinking too small about big data?' by IBM gathers the growth forecasts of the value of big data for 2015: http://www.ibm.com/smarterplanet/us/en/smarter-enterprise/perspectives/big-data-and-analytics.html

> 2014 annual report of the start-up Splunk: http://files.shareholder.com/downloads/AMDA-RWDLH/3981313229x0x657273/5C7E9352-D951-4F14-BE2E-6B513C307C05/Splunk_Annual_Report_FY14.pdf

> 2013 annual report of the company Acxiom: http://d3u9yejw7h244g.cloudfront.net/wp-content/uploads/2013/09/2013-Annual-Report.pdf

> Study published in the journal Computers in Human Behavior regarding the possibility of knowing how Facebook users feel at the moment with 83% accuracy by conducting a big data analysis of messages and interactions between users: www.sciencedirect.com/science/article/pii/S0747563213001751

> Set of research and papers published in Science's special report 'The End of Privacy' regarding the opportunities and dilemmas presented by the use of data published on the Internet by individual users, and the violation of privacy and anonymity of individuals: www.sciencemag.org/content/347/6221/490.full?intcmp=collection-privacy

7.2 Articles and News

> Article that discusses the progress leading to storing all the data in the world in a DNA hard-drive the size of a spoon in the future: http://singularityhub.com/2015/02/20/worlds-data-could-fit-on-a-teaspoon-sized-dna-hard-drive-and-survive-thousands-of-years

> Article on an analysis carried out by Jawbone based on data regarding the sleep quality of people exposed to the same earthquake: https://jawbone.com/blog/napa-earthquake-effect-on-sleep

> Article regarding the capacity to predict our psychological profile with big data: http://elpais.com/elpais/2015/01/12/ciencia/1421084469_835718.html

> Article about big data's predictive capacity regarding the people who matter to us: www.technologyreview.com/news/533536/att-builds-an-assistant-app-with-social-skills

> Article regarding the police use of facial recognition software, NeoFace, to identify criminals. Already used in the United Kingdom: www.bbc.com/news/uk-england-leicestershire-28307938

> Article regarding an experiment conducted by the security company Hide My Ass! In the experiment, a 7-year-old girl hacked a wifi network: www.dailymail.co.uk/sciencetech/article-2919762/Hacking-Wi-Fi-s-child-s-play-Seven-year-old-shows-easy-break-public-network-11-minutes.html

7.3 Websites

> Website of the Quantified Self community, created to keep abreast of the latest self-measurement tools and apps and help people understand them: http://quantifiedself.com

> SHIFT corporate website. Maarten den Braber, a co-founder of this company, is an expert present at the FTF: http://shiftbsp.com

> Website of the Fitbit smart bracelet to measure daily activity: https://www.moves-app.com

> Website of the UP by Jawbone smart bracelet to measure daily activity: https://jawbone.com/up

> Informational website on the book 'The Victory Lab', by political analyst Sasha Issenberg, where it explains how the Democratic Party used big data to win the presidential elections in 2008 and 2012: www.thevictorylab.com

> Website of the GE Flight Quest competition: www.gequest.com

> My Own Med corporate website. This company has been founded by **Vicki Seyfert-Margolis,** expert present at the FTF. A platform where patients, caregivers and healthcare professionals share one interface, access the same data and work on the same page thanks to cloud architecture.

> DeleteMe website. It will delete your personal data from the Net. This service is part of the product portfolio of the company Abine: www.abine.com/deleteme/landing.php

> Website of the program Global Pulse of the United Nations This program encourages sharing big data opportunities that will drive development through data exchange alliances between the public and private sectors: www.unglobalpulse.org

7.4 Applications and Platforms

> Kaggle web platform. It is a community of data scientists focused on crowd solutions to problems via open competitions: www.kaggle.com

> Canada's Open Government Portal: http://open.canada.ca/en

> US Open Government Portal: www.whitehouse.gov/open

> Basque Country's Open Government Portal: http://www.irekia.euskadi.eus/es

> Open Government Portal of the Zaragoza province: www.zaragoza.es/ciudadania/gobierno-abierto

> Spain's open data portal: www.datos.gob.es

> Moves application. It measures daily activity: www.moves-app.com

7.5 Other

> Graphic representation of a man's heart rate as he is proposing in marriage: http://imgur.com/mbOPX2L

> Opening video of the TV series 'Person of interest': http://youtu.be/WOnQ8CD3v4g