

March 2017

HOW BIG DATA IS ABOUT TO EXPLODE POLICYMAKING AS WE KNOW IT

THE RISE OF CIVIL ANALYTICS

Don Lenihan
Tom Pitfield

“In the post-fact era of mistrustful populism here is a compelling case for governments to double down on evidence, Big Data and predictive analytics. Lenihan and Pitfield argue that “civil analytics” holds the alluring promise of delivering powerful cures for today’s most wicked policy problems. It is a promise of effective, value-for- money government that everyone should welcome.”

Giles Gherson
Deputy Minister

Ontario Ministry of Economic
Development + Growth Ontario
Ministry of Research,
Innovation + Science

Canada 2020 is Canada's leading, independent, progressive think-tank.

Founded in 2006, we produce original research, host events and start conversations about Canada's future. Our goal is to build a community of progressive ideas and people that will move and shape governments.

For more, visit
www.canada2020.ca



ABOUT THE AUTHORS



DR. DON LENIHAN is Senior Associate, Policy and Engagement, at Canada 2020. Don has over 25 years of experience as a project leader, writer, speaker, senior government advisor, trainer and facilitator. Throughout his career, he has developed and led many research and consultation projects involving senior public servants, academics, elected officials, journalists and members of the private and third sectors from across the country. He is the author of numerous articles, studies and books, and was a weekly columnist for National Newswatch. He earned his PhD in political theory from the University of Ottawa.



THOMAS PITFIELD is President of Canada 2020, Canada's leading, independent, progressive think-tank. He is also one of the organization's Co-Founders. Prior to his appointment, Tom served as the Chief Digital Strategist for Prime Minister Justin Trudeau's federal election campaign, as well as his campaign for leadership of the Liberal Party of Canada. Previously, Tom served as Senior Policy Advisor to the Leader of the Government in the Senate, the Honourable Jack Austin. He also worked in China for the Canada Chinese Business Council and as a business strategy consultant, specializing in Corporate Governance, for IBM Canada. After co-founding IBM's Toronto Innovation Center, he worked as Policy Advisor to the Director of IBM's Business Partner Channel.

THE RISE OF CIVIL ANALYTICS – FOREWORD

BY MELANIE COULSON, CANADA 2020

How Big Data is about to Explode Policymaking as We Know It is the first of a series of papers Canada 2020 will release on data and policymaking.

In it, Tom Pitfield and Don Lenihan explain the shifts that will occur thanks to massive amounts of high quality data and a new capacity for data analytics.

Using the right analytics tools, and involving the right leaders, could be considered an answer to the post-fact politics that seem to be rising up all around us.

Civil Analytics, as Pitfield and Lenihan have defined the term, is a holistic approach to data, the tools that can be used to analyze it, and the various people who should be engaged to examine it.

As agencies and individuals with various interests are included in the process of understanding the data and creating policies, they will feel a greater ownership over them, which results in easier adoption.

Pitfield and Lenihan have big things to say about history, technology and politics, and this essay should be of interest to anyone watching where policy is going and what big trends are on the horizon.



TOWARD A NEW ERA IN POLICYMAKING

The history of modern government can be told as a story of the struggle to ground policymaking in knowledge and evidence, from the rise of scientific knowledge in the 19th century to the application of Big Data today. The so-called Green Revolution in the mid-20th century is an impressive example of what can be achieved through evidence-based policymaking.

The Green Revolution began with a handful of reforms proposed by agricultural experts to help modernize farming in the developing world. It included new, high-yield varieties of cereals and rice, chemical fertilizers, irrigation and modern methods of cultivation, including mechanization. The result was a spectacular success. By the late 1960s global crop production had almost doubled, an achievement that is credited with saving more than a billion people from starvation.

Success, however, can be a two-edged sword. While the Green Revolution has been celebrated for this growth in output, it has been criticized for some of its other effects: modernization undermined important traditional practices, such as the saving of seeds for planting; and it precipitated a shift in land ownership from individual farmers to large corporations.

There is an historic lesson here for governments: evidence-based policymaking is about far more than applying specialized knowledge to solve a problem. In the real world, the relationship between causes and effects gets complex very quickly. Changes to one part of an ecosystem or a community affect other parts, often in unforeseen and surprising ways. Managing this kind of change requires continuous monitoring, adjustment and learning to ensure the benefits are realized and unwanted consequences minimized.

Since the Green Revolution, policymakers have given a lot of thought to managing change and some very promising ideas have emerged. But getting them off the drawing board and into practice has been slow and difficult. Too often, a key enabler — high-quality data — has been in short supply. Good data allow managers to observe and react to change in real time. Until now, however, there simply hasn't been enough of it to truly put ideas to work. That is changing.

Today, massive amounts of high-quality data are coming online, along with a remarkable new capacity for data analysis or, as data scientists call it, “analytics.” The combination is explosive. Machine learning, for example, is the branch of analytics that uses special algorithms to allow computers to learn things they have not been explicitly programmed to learn. Experts have recently found that larger datasets not only allow smart systems to learn more things, they make them smarter learners.

For example, Google's word program uses machine learning to improve its grammar checker. At first, researchers applied the algorithms to a base of about a million words. Progress was slow and the system made lots of errors. By 2000, the computing power was available to vastly increase the size of the database — first to 10 million words, then 100 million and, finally, one billion.

Much to their surprise, researchers found that with each increase, the performance of certain algorithms improved markedly, making the system more adept at analyzing grammar. It was not just learning new things, it was getting smarter about *how* it learned them. The implications are fascinating: if machine learning is not just a function of the algorithms, but also of the size and reliability of the database on which they are working, what will happen as the availability of data soars?

Big Data is already rocketing analytics and machine learning to new heights. Soon machines will surpass humans in a wide range of intellectual tasks, such as medical diagnostics, accounting and legal analysis. The consequences are genuinely transformative.

These tools are also poised to push policymaking to a new level, where its effectiveness and reliability will vastly improve. This paper considers how and why Big Data, modern analytics and machine learning can transform policymaking and asks what steps must be taken to ensure success. We start by looking at two seminal achievements from the past quarter century — performance measurement and holistic analysis — then consider how they can be combined with Big Data and analytics to define a new approach to policymaking, one we call **civil analytics**.

MEASURING PERFORMANCE - A NEW PARADIGM

A few decades ago, a basic change in how policymakers work swept through OECD countries. These governments realized that, although they collected a lot of information on how they spent their program budgets, they collected almost none on the impact the use of those resources had on society — the so-called *outcomes* they achieved. Without such information, how could they know if their policies and programs were realizing their goals?

A new approach called “performance measurement” (PM) arose to fill this need¹. PM included three basic tasks. First, managers were supposed to set clear program goals. Second, they were required to identify progress indicators, collect information on program performance and use the indicators to assess progress toward the goals. Finally, they were to adjust a program’s design to tighten the “fit” between program goals and actual outcomes.

¹ Performance measurement/management was a key part of a hugely influential movement from the 1980s and 90s known as the *New Public Management*, the main features of which were succinctly stated by Christopher Hood in ‘A Public Management for all seasons’, *Public Administration Review*, Vol 69, Spring 1991 (3–19).

Once the changes had been made, officials would launch the next round of program delivery, wait a bit, then evaluate its performance again, make more changes, launch another round, and so on. Ideally, the fit between program goals and outcomes would get tighter after each cycle.

The great promise of PM was that it created a virtuous cycle of **continuous learning and improvement**, where each round was supposed to yield more effective programs and services. They called this the **learning cycle**.

It is hard to exaggerate the impact PM has had on modern governments. It was, in the truest sense, a paradigm shift in policymaking that quickly spread beyond OECD countries.

Today it is the basis for program management in governments around the world, including most of the countries in South America, Africa and Asia.

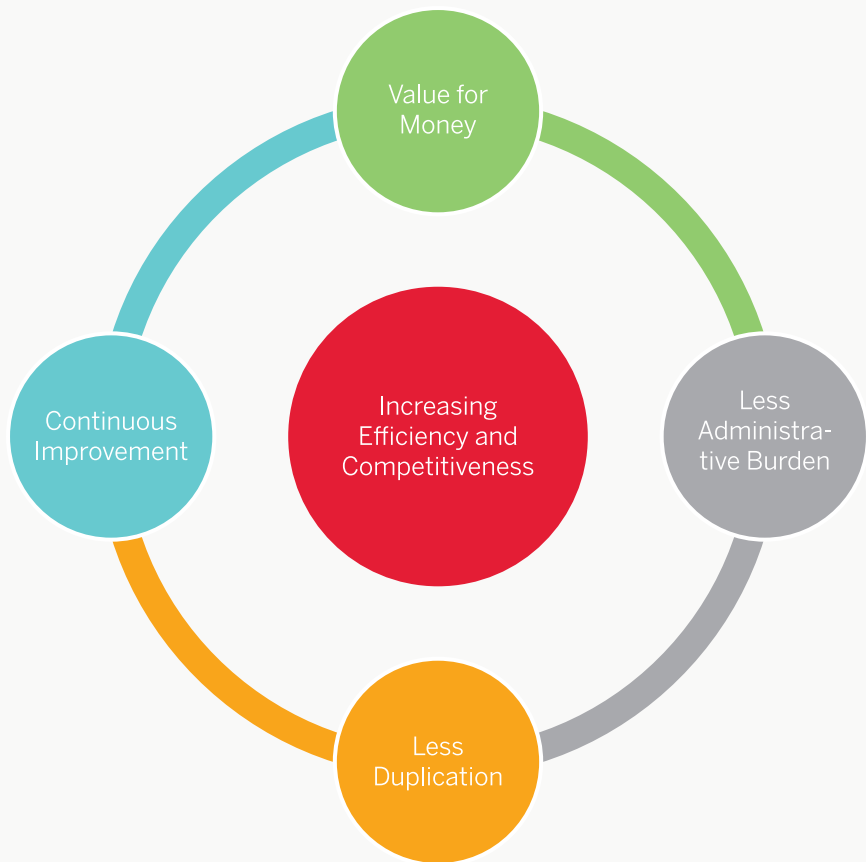
THE HOLISTIC TURN – FROM SILOS TO SYSTEMS

In hindsight, however, PM didn't pan out quite the way its designers had expected. As managers focused on clarifying their program goals, they realized that programs often have multiple goals. Consider the Government of Canada's approach to regulation², which holds that better regulation should lead to:

- value for money
- less administrative burden on businesses
- less duplication
- continuous improvement

This policy has four main goals, which is not unusual. Some have many more. As managers considered the implications for PM, they also noticed that multiple goals can converge on a broader, overarching goal. For example, the four goals above could be sub-goals of the broader one of “helping Canadian businesses become more efficient and competitive,” as the diagram below shows:

² This is set out in *The Cabinet Directive on Streamlining Regulation* (2007), available at: <http://civication.gc.ca/site/eng/309288/civication.html>

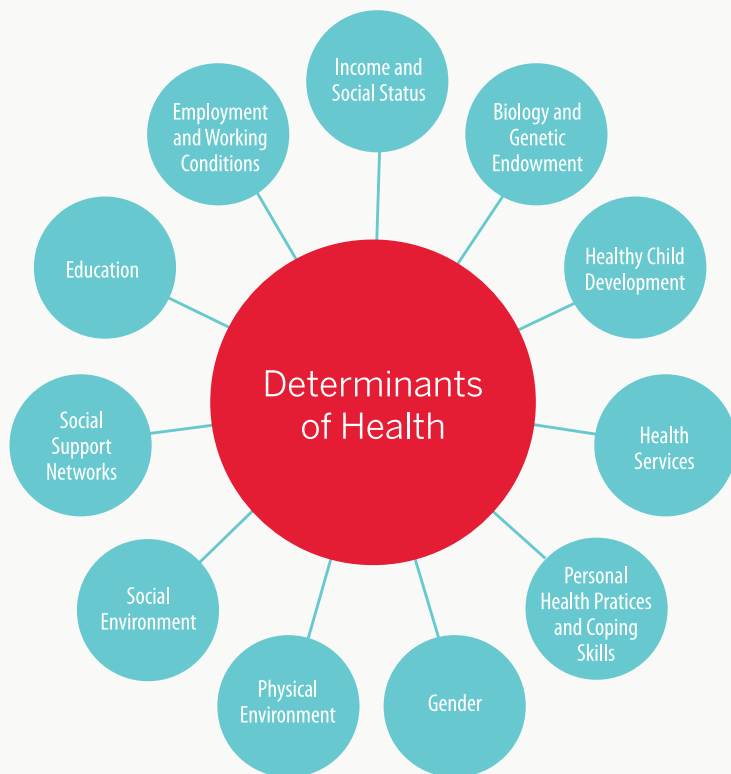


Identifying overarching goals turned out to be an important and useful way to help fill in the bigger picture around a program, such as why it exists, how sub-goals should be ranked and how the program relates to other programs and issues.

The more managers used higher goals to integrate clusters of lower ones, the more impressed they were with the benefits. This work culminated in what is now called *holistic* policymaking³. Determinants of Health is perhaps the best-known example.

³ It is also widely known as “systems thinking” or “the systems approach.”

In this model, health is the overarching goal or outcome of the system. Health is seen as resulting from much more than physical factors, such as fitness or nutrition. A person's or community's health is also determined by their economic conditions, cultural background, education levels and a range of other factors, as the next diagram shows:



[Ontario Healthy Communities Coalition](#)
Coalition des Communautés en santé de l'Ontario

This broad-based approach was a major departure from the old “siloes” approach to policymaking, where data and information on, say, income weren't likely to be relevant to a discussion on health. However, viewing these factors as interdependent parts of a single system encouraged experts from different fields to begin sharing data, information and knowledge to see what light it would shed on health-related issues.

It changed how policy is made. Sharing data and information across the silos allowed officials to connect the dots between different sets of issues, which, in turn, uncovered all kinds of new links between them. They learned, for example, that people with low incomes are particularly susceptible to a variety of diseases, such as diabetes. Exploring these hitherto hidden links has been a major driver of discovery and innovation in the field of health ever since.

Holism, as we shall call it, was not confined to health. Experts from other fields quickly proposed similar goals to create holistic systems of their own. The 1987 Brundtland Report famously called for a new multilateral commitment to **sustainable development**, based on the idea that the economy and the environment were inseparably linked⁴. The concept of **life-long learning** was framed to underpin a holistic approach in education. Others followed suit and within a decade holistic thinking had become the norm in virtually every major policy field.

It is no accident that holism emerged when it did. Its success was closely linked to the rise of digital technologies. The old, paper-based system that supported departmental silos was far too slow and labour-intensive to investigate more than the most immediate and conspicuous links between issues. Policymakers may well have suspected that all kinds of interesting links existed, but governments were not organized to handle the connections and lacked the tools to gather and use the data and information to explore them.

³ The report was produced by the United Nations World Commission on Environment and Development and named after the commission chair, former Norwegian prime minister, Gro Harlem Brundtland.

Digital technologies and the emergence of the internet changed all this. Suddenly, it became possible to gather, share and integrate data and information from a wide variety of sources, pushing governments toward a more “networked” organizational structure and rendering this kind of policymaking practical.

The 1990s and early 2000s were a period of high excitement and adventurous redesign for governments everywhere: filing systems became electronic, websites made government information more accessible and governments began providing services online, from renewing drivers’ licenses to filing taxes.

The combination of PM and new digital tools spawned a new vision of government. The challenge of “e-government” or “government online,” as it was often called, was to transform a loose collection of policy silos into an integrated information system. Traditional organizational boundaries became less clearly defined and roles and responsibilities became more flexible.

Performance measurement may have started out as a strategy for increasing the effectiveness of individual programs and services, but by the end of the millennium it had evolved into a much more ambitious effort to share data and information across government to promote broad holistic goals, such as healthy populations, sustainable development, safer communities and innovation.

PROBLEMS WITH PERFORMANCE MEASUREMENT AND THE LEARNING CYCLE

As the proverb says, “Two steps forward, one step back.” If PM promised to transform policymaking, things haven’t quite worked out that way. Measurement and evaluation turn out to be a lot harder than people thought. When policymakers face a pressing challenge, such as, say, unemployment or chronic traffic congestion, the time it takes to complete the learning cycle can cripple the effort to respond. Data get drawn from a variety of sources, but is often out of date, incomplete or both. Collecting new data can be costly and time-consuming, as is its analysis. Unintended consequences often flow from a new policy or program, leading to new problems that require more data and analysis.

In short, the learning cycle is often slow, costly, complicated and unreliable. While PM is now institutionalized in governments around the world, the capacity to do it effectively is notoriously uneven. But there is light on the horizon. A new era in performance measurement may be dawning, one that could make the learning cycle much faster, far more integrated and vastly more effective. This, in turn, has the potential to transform the way policy of all sorts is made and implemented, from bicycle lanes to famine relief.



For reasons of space, our focus here will be on one aspect of this revolution: how Big Data and analytics will allow policymakers to move away from “reactive” approaches to policymaking and toward what we call “proactive” ones. A more comprehensive account of the implications of Big Data and analytics for policy-making must wait for another day.

FROM REACTIVE TO PROACTIVE POLICYMAKING

Much of traditional policymaking is *reactive* in the sense that it responds to events *after* they happen. Thus, law enforcement reacts to crime and the health system reacts to illness. Over the years, governments have found themselves building ever-larger police forces or more hospitals to combat crime or illness, to the point that many policymakers now see this as a treadmill that governments must get off, if only to halt the escalating costs. Is there an alternative?

Perhaps the most appealing option is to put far more effort into *preventing* crime or illness from happening in the first place. An ounce of prevention, after all, is worth a pound of cure. Of course, the idea of prevention is not new and governments have been active on this front for decades. To avoid diabetes and heart conditions, for example, doctors recommend low-fat and low-sugar diets. To prevent crime, social workers call for better recreational activities for youth, such as sports or social clubs.

As we will see below, however, initiatives like these are at best limited forays into disease or crime prevention. The challenge ahead is to develop comprehensive prevention initiatives that are both effective and affordable. Indeed, proactive policymaking should not only seek to prevent bad things from happening,

but to encourage good ones, such as healthy living, sustainable development, inclusive communities and innovation. Big Data and analytics bring this within reach, but success requires a very different approach to policy analysis. The Ontario Ministry of Correctional Services and Community Safety’s (MCSCS) *Community Safety Initiative* provides a convenient example.

Research shows that while there are identifiable risk factors that contribute to community crime, no single one is responsible. They do not work in isolation. Rather, they interact in ways that increase the likelihood of delinquent or criminal behavior⁵. The data show that key risk factors exist at four basic levels:

Risk Factors			
Society	Community	Family/Peers	Individual
Cultural norms supporting violence	Crime in area	Abuse	Behavioural problems
Social disorganization	Few social services	Few economic resources	Poor educational achievement
Negative media messaging	High poverty concentration	Neglect	Poor mental health
	Poor housing	Negative parenting	Prior criminal behaviour
		Poor peer influences	Racism/marginalization
		Parent/sibling criminality	victimization/abuse

But if there are key risk factors, there are also “protective factors” that can reduce risk. Promoting these protective factors contributes to a community’s overall well-being, which, in turn, reduces crime. These protective factors also exist at four levels:

³ Crime Prevention in Ontario: A Framework for Action, Ontario Ministry of Correctional Services and Community Safety.

Protective Factors			
Individual	Family/Peers	Community	Society
Personal coping strategies	Adequate parental supervision	Housing in close proximity to services	Low social tolerance of violence
Strong attachment to adult	Parent(s) engaged in child's life	Cohesive communities	High awareness of the determinants of well-being
Positive school experience	Positive peer influences	Recreational facilities for youth	
Self-esteem			
Self-efficacy			
Sense of responsibility			

The MCSCS plan is to gather data on how these factors are affecting key social networks at each of these four levels, analyze how they contribute to crime and prevention, then formulate a plan or strategy to build on the protective factors in ways that help neutralize the risk factors.

This approach is now used in a wide range of areas, including health promotion, poverty reduction, social inclusion, environmental protection, innovation and more. But if the approach is sound, making it work remains a challenge. To see why and what must be done, let's take a step backward in time, which will help us learn an important lesson for the future.

A LESSON FROM THE PAST

In the early days of modern medicine, doctors struggled to identify basic illnesses. Symptoms were at times seen as random or coincidental, and at other times as the illness itself. Consequently, healing often focused on the effects of the illness — the symptoms — rather than the cause.

Gradually, doctors realized that symptoms, like a fever, spots or a cough, were manifestations of an underlying cause, which was the real illness. Symptoms became signs or *indicators* pointing to what was wrong with the patient. This was a huge step that, in turn, suggested a methodology. The idea was to use the symptoms to diagnose the illness, guide the intervention and allow the doctor to monitor and assess the patient's progress.

This methodology also suggested a research program. Doctors collected data to identify which symptoms were associated with which illnesses. As they did, they learned that different illnesses can cause similar symptoms; and that a single illness can have multiple symptoms.

The new method greatly accelerated progress and, over the years, doctors devised legions of tests, from x-rays to blood tests to ultrasound, to explore and map the impact of disease on the human body.


They learned that the body is a system of systems, which interact in complex ways. Thus, the circulatory system, which is composed of the heart, blood vessels, arteries and capillaries, interacts with the respiratory system — principally, the lungs — to release carbon dioxide and to circulate oxygen to the body.

Thanks to this research, the medical community's view of the relationship between causes and effects, and diseases and symptoms, is vastly more sophisticated today than in the early days—as is its ability to diagnose and treat disease.

Indeed, an exhilarating new development — the mapping of the human genome — has opened a whole new window on the relationship between causes and effects in medicine. For example, sometimes a genetic causal relationship is clear and direct, such as the chromosome that determines eye colour. At other times, it is conditional, such as genes that predispose women to breast cancer. In these cases, other factors at other levels can enhance or reduce the risk. The questions now are which factors, when, where and why?

Moving forward, researchers will seek to answer these questions by taking the new datasets on the genome and using analytics to explore the causal connections between an individual's DNA, the various systems within their body and their environment. The results promise to transform modern medicine.

This historical sketch of modern medicine offers a glimpse into the trajectory that lies ahead for proactive policymaking. Analysts who seek to prevent some harm, such as crime or poverty, or to promote some good, such as public health or sustainable development, are struggling with methodological problems much



like those faced by doctors in the early days of medicine, but they appear to be at a turning point.

Research documents like the MCSCS tables above are an important milestone in the journey. These lists of risk and preventive factors provide the basis for a methodology, much like the early medical approach based on symptoms and causes. Just as doctors used that methodology to explore and map the systems of the human body, policymakers are now poised to begin exploring and mapping how these factors interact within a community. Unlike the early doctors, however, today's researchers can use Big Data, analytics and high-powered computers to help them. The way forward is clear and real progress is within reach, as the debate over "root causes" clearly illustrates.

ROOT CAUSES AND DATA MAPPING

In the root causes view, real progress on preventing illness or crime requires digging deep, finding the underlying causes, then pulling them out by the roots. The view has much in common with the distinction between symptoms and disease. It contrasts the deep underlying causes of a social condition like crime — the disease — with superficial ones, which are more like its symptoms. Thus, while some people point to drug use as a key cause of crime, the root causes view treats it as an effect. The real causes are much deeper, such as poverty, lack of education or social stratification. Further, like the systems of the body, root causes are interconnected. To make real progress on an issue like crime, we need to understand how they are interacting.

This is the key challenge now facing proactive policy-makers. Although there is lots of talk about attacking the root causes of issues like crime or poverty, these initiatives often fall short of the mark. The interconnections between the causes rarely get explored and are even less likely to be acted on holistically. Policies and programs to reduce crime, for example, usually fall back on the idea that root causes can be dealt with one at a time. Thus, policymakers create a bit of affordable housing here, and a new program to reduce family violence there, as though the roots they were trying to pull out were of the carrot or turnip variety.

They are not. Root causes are more like the tangled roots of an oak tree: they are part of a single system that produces the effect. It is thus a mistake to speak of “the” root cause of an issue. Root causes are multiple, not singular and they are *interdependent*, not *independent*. They behave like an ecosystem, where a major change in one part of the system affects the system as a whole — often in surprising ways. Trying to isolate individual causes and effects within such a system, then base policy on them, is unreliable at best, misguided at worst.

This is NOT to say that programs that focus on a single underlying cause — say, housing or family violence — are unhelpful. As noted above, doctors and social workers have been prescribing these kinds of preventive measures for decades. Rather, the point is that other factors are usually also at play. The impact of single measures on a complex phenomenon like crime or poverty thus will be limited. To make real progress, we need to address a critical mass of the factors that are entangled within the issue space and we need to do it *in a way that reflects their interconnectedness*.

Like work on the genome, this kind of analysis is often highly complex and requires Big Data, state-of-the-art computing tools and sophisticated analytics. The right combination of these inputs can be a game changer that provides analysts with the information they need to identify:


- the constellation of risks and protective factors at work in the community;
- how they are interacting to produce crime, poverty or some other issue; and
- where the key links within the network lie.

This kind of information, in turn, puts performance measurement on a new footing. While the old model saw evaluation as the last stage of the learning cycle, “data mapping” allows for a much more dynamic and engaged approach, where learning occurs as the cycle unfolds. By constantly updating the map, a manager can monitor the project’s impact on the system in real time.

He/she can thus use the intervention to affect how the risk factors and preventive factors are interacting. Ideally, the person begins to steer events within the system by continuously adjusting the intervention to the changes around it, much the way a navigator’s hand on a tiller continuously responds to changing currents, waves and wind acting on the vessel. This kind of feedback system thus makes the manager an active participant in the system who can help shape trends as they develop.

Now, the idea of a continuous feedback loop where interventions are used to shape the system is not new. As we’ve seen, this is what doctors do when they monitor the healing process by monitoring the patient’s symptoms. The point here is that carrying out this kind of intervention across an entire community is a far more complex enterprise that requires the right data — which must be robust, accurate, timely and relevant — and the right analytics, which must be able to comb through the data to provide real-time information about the system.

We call this approach **civil analytics** because it combines Big Data, analytics and new policymaking techniques, tools and skills. As we’ve seen, this approach is already being used to achieve historic breakthroughs in the field of medicine. Experimentation



is also underway in areas like crime prevention and environmental management.

There is every reason to think these methods can be further improved and that they can be applied to a wide range of other policy areas. This wasn't an option until very recently because the data either didn't exist or managers lacked the raw computing power to use the algorithms to analyze it. For the most part, policymakers clung to old-style methods of finding linear causes and building policies and programs around them. But the data environment is changing, and the question now is whether policymakers, stakeholders and citizens are ready and willing to change along with it.

CIVIL ANALYTICS, PUBLIC DEBATE AND DEMOCRACY

We've argued that civil analytics raises performance measurement to a new level so that highly sophisticated evaluations of the causes, correlations, solutions and impacts around issues like crime or public health are not only possible, but essential to progress.

In future, with the right data, civil analytics could become the policy equivalent of an MRI scan, which uses powerful magnetic fields and radio frequency pulses to produce detailed pictures of organs, soft tissues, bone and other internal body structures.

Some will wonder if all this talk of data and algorithms threatens the place of public dialogue and debate in policymaking. We think not. While we agree that civil analytics will change the nature of policy discussion, we believe it will do so in ways that expand the opportunities for meaningful debate, dialogue and collaboration.

Conventional policy debate is usually framed as a winner-take-all contest where one side is right and the other wrong. Thus, if one person says that the best way to control the spread of drugs is through stronger

penalties and someone else says it is through safe injection sites, people assume that one of these must be wrong. Debate is supposed to sort out who is right and who is wrong.

From the holistic viewpoint, this kind of debate is misguided. Reactive analysis focuses on what has happened in the past. If a case can be found where stronger penalties were a decisive factor in reducing drug use, then it can hardly be argued that this view is wrong. The same holds for safe injection sites. In fact, both could be right.

However, nothing here assures us that things will align the same way in the future. Holism recognizes that many factors are at play in such issues and that a shift in their alignment could easily change the outcome.

Civil analytics can hugely improve our capacity for such analysis, but this does not mean that computers will decide everything for us. Discussion and debate must play a central role at every stage. For example, deciding which datasets to use is a critical choice. Similarly, the interpretation of data often requires searching discussions of the social, cultural, economic or environmental circumstances around the issues. Turning these findings into workable policy is yet another task that requires discussion and debate.

Getting the right mix of people around the table to sort through these issues is a critical condition for the success of civil analytics. These people won't be just data scientists. Depending on the issue, the process will need experts of various kinds. Ordinary citizens are important too. In a democracy, they are the "experts" on values and priorities. In our lexicon, "Open Dialogue" refers to the set of skills, processes and tools needed

to make such exchanges fruitful. This is a rich and important side of civil analytics that we must leave for another day.

So, nothing here suggests that civil analytics turns policymaking into a subcategory of data science. Big Data and analytics will not eliminate debate from the policy process, but they will change it. Ideally, they will make it more *evidence-based*. We think that is a good thing.

However, there is a darker side to civil analytics that must be acknowledged here. We believe that societies like our own are all but certain to use Big Data and analytics to support evidence-based decision-making. The benefits are too great to ignore. We also realize that this does not guarantee these tools will be used to enhance democracy. Policies can be evidence-based, yet undermine democracy. Authoritarian societies, whether mild or strong, can make aggressive use of Big Data and analytics for their own ends.

Early on in this paper, we had to set aside for another day questions on the role various kinds of analytics will play in civil analytics. A few paragraphs above, we did the same thing for Open Dialogue. Now we must add democracy to that list, not because we think it is unimportant, but for lack of space. For the moment, we will simply declare our belief that an ambitious commitment to civil analytics should be accompanied by an equally ambitious commitment to enhancing the transparency and accountability of governments and their respect for basic rights and freedoms — what is now widely known as “Open Government.”

THE AGE OF BIG DATA

In the past, when governments collected data and information, they were usually for specific purposes and they stored them in silos designed for that purpose. There were no universal standards for data collection and storage. They were archived in all kinds of ways, often just put away somewhere in a file, so that no one would even know they were there.

The information management systems of the day were not designed for large-scale data sharing. No one foresaw the need. As a result, accessing these data for policy purposes today is often expensive, difficult and time consuming. In addition, there are questions around how to overcome barriers, such as privacy, security or the protection of commercial advantages.

This is a problem for civil analytics. If a dataset remains in its silo, the story it might have to tell when linked to other datasets will remain unknown. That is the big lesson of the holistic turn. Datasets that are seen by their curators as having a singular purpose often turn out to be highly illuminating when viewed by other experts in other contexts. To extract the value from this resource, data must be freed up so they can be accessed, analyzed, and compared and contrasted with other data.

In response, governments are not only making their data reserves available to the public — Open Data — they are adopting new shared standards that will ensure

that data are collected and stored in ways that make them accessible in future. In addition, governments are redesigning their IT architecture to transform their data holdings into a vast reservoir. Rather than a collection of concrete silos, the government of the future will be a digital platform from which data can be easily accessed and whose open features will allow access and connectivity to all kinds of other data sources outside government. The new vision is of government as the steward of this valuable public resource and the agent that will make it accessible to everyone.

Realizing this vision will not be easy. This is about more than adopting new technologies or making organizational changes. Civil analytics also requires changes in our policies and the attitudes behind them. Personal privacy is a good example. Civil analytics could result in far better public policy, but policymakers will need access to high-quality data. Sometimes this will be personal. Does that mean that civil analytics involves a trade-off with personal privacy?

We believe that a better balance between access and privacy must be struck, but it is wrong to see this as a trade-off. The challenge is one of refining our views on privacy protection. Current privacy rules take a one-size-fits-all approach. They could be much more nuanced. For example, people who are unwilling to share their address or birthdate might be willing to share information about their gender, hair color or some other easily identifiable data point. Why not give them the choice? Citizens already willingly disclose all kinds of personal information to rewards programs, discount cards and online services. Ironically, however, the huge data resources accumulating from these exchanges are rarely available for public purposes, mainly because of

government restrictions on *data mobility*. The worrying result is that a handful of private data companies now control most of the reliable public data.

If this is a bad precedent for the future, the rules around privacy are the main culprit. As things stand, the same laws that prevent privately controlled data from being shared for public good also prevent citizens from agreeing to pool their own data in ways that would lead to public benefits. Yet, as the private sector experience shows, citizens are often quite willing to exchange personal data for better services.

Over the next decade, privacy rules are likely to become a serious issue. In the age of Big Data, the free and consensual exchange of personal data among individuals, companies or data brokers — what we call data mobility — will be a critical factor in building innovative, prosperous and informed societies. Pressure is already growing to ensure the rules reflect this. For example, some people favour creating a government-regulated market for commercial data. They believe that allowing companies to buy and sell data this way could greatly enhance data mobility, while ensuring privacy protection.

Whether this is the right solution or not, data mobility is an essential condition for better decision-making, both in our institutions and among our citizens. In the coming years, new tools and practices of all kinds will evolve to support it, such as common “data dictionaries” that will give analysts a reliable tool for comparing large datasets. Expect citizens to become increasingly associated with a single ID so their personal files can be shared — with their permission — across organizational boundaries. Policymakers will need new skillsets, as they debate the format, scale and scope of government datasets as hotly as they do trade agreements or fishing quotas today.

We are on the threshold of a new era in policymaking. The journey has taken decades and involved much effort. The establishment of performance measurement was a defining moment along the way — a huge step in the direction of better decision-making. In hindsight, however, we can see that, while the idea was right, the execution was flawed. Breaking the process into stages made it slow and clumsy and the learning from it was often outdated. But governments couldn't have done otherwise; they lacked the tools and the data to really put the ideas to work.

That is changing. By 2020, some 50 billion devices will be streaming data into the internet. We are entering an era where data are abundant, as are the computing power and analytics to process it. The impact is already visible, from self-driving cars to smart cities that adjust the power supply to meet demand or the location of police resources to control crime.

And this is just the beginning.

Managing trade, fostering innovation, protecting the environment, ending poverty and halting crippling diseases — the possibilities for transforming our physical and social environment are limitless.

The age of Big Data is here and we have the tools to make it an era of enlightened policymaking and informed political participation. The question now is whether our governments and our citizens have the imagination and the will to make that happen. We are hopeful.

Canada 2020 is Canada's leading, independent, progressive think-tank. Founded in 2006, we produce original research, host events and start conversations about Canada's future. Our goal is to build a community of progressive ideas and people that will move and shape governments. For more, visit www.canada2020.ca