

# Interview with George Sciadas, about his book ‘Number Savvy: From the Invention of Numbers to the Future of Data’<sup>1</sup>



Walter Radermacher



George Sciadas

**Abstract.** George Sciadas, a former director from Statistics Canada recently published the book ‘Number Savvy’, with the subtitle ‘From the Invention of Numbers to the Future of Data’. Though in recent years several books on data were published, this book is striking from several perspectives. It has been written by an insider, an experienced employee from a national statistical office, though in a style that allows a wide range of readers to understand and enjoy the material. Another reason for putting the spotlight on this book is its explicit structure, highlighting crucial elements of the production of statistics as well as its use, which makes it a tremendously useful book for those involved in training in statistical literacy. With this specific focus in mind, Walter Radermacher was found willing to interview the author George Sciadas.

**Walter Radermacher. Dear George, as a starter to this interview, and to help the readers understand where you are coming from, can you tell us a bit about yourself, especially from your career and experiences at Statistics Canada, known as one of the finest statistical offices worldwide?**

*George Sciadas:* Certainly. Statistics Canada became home to me for more than three decades. I’ve had opportunities to work in various areas, experience different statistical processes and expectations for analytical

outputs, and learn from very knowledgeable people. Over time, I’ve also had the opportunity to meet and interact with unique personalities in the organization and internationally, something that enriched my life. My whole career revolved around taking care of statistical

George Sciadas has worked in the public, private, and academic sectors. He’s well known in statistical circles in Canada and internationally, having worked for more than three decades at Statistics Canada and international organizations, including in several executive capacities. He has also taught at universities for many years. He earned his Ph.D. in Economics at McGill University in Montreal. He has led many national and international projects, with research teams on all continents. He has authored numerous papers and monographs and has been the editor of influential publications and compendia for many years.

<sup>1</sup>George Sciadas (2023); Number Savvy: From the Invention of Numbers to the Future of Data. CRC Press, ISBN: 9781032357218 (pbk), 9781003330806 (elk).

and analytical needs in areas with data gaps or new and emerging areas, typically with external partners and financing. The biggest example was StatCan's pioneering work in measurements and analysis of the Information Society, following the arrival of the internet. In the last ten years in particular, I was the Director of the Centre for Special Business Projects, a division that was built expressly for cost-recovery projects, primarily in business and economic statistics.

**Walter: This Centre for Special Business Projects seems to be very forward-looking and apparently has contributed a lot to your experience. Can you elaborate a bit more on how this works?**

*George:* As a centralized agency, there isn't much that StatCan doesn't measure – outside of weather and sports, as I used to explain half-jokingly. Still, regardless of the size of the organization chart and the volume of programs, as everything evolves there are always unmet needs. StatCan welcomes this challenge and takes on such work mostly on a cost-recovery basis. Projects extend from customized surveys to all kinds of data and analytical products. In recent times there are generally more demands for such work. My division has in fact grown from a handful of individuals to over 100 today. Our teams become the eyes and ears of StatCan outside, picking up useful signals for the evolution of programs. We also become the spokespersons for the partners inside the agency. Among other things, the relevance of the outputs is ensured as for every project a paying partner is waiting – and impatiently!

You can appreciate that these jobs are different from a typical job in a program with fixed periodicity, such as monthly or annual. For one, you can't do the job until you know well the existing data sources and the outputs of all other programs. Then, you work on an astonishing number of subject matters, moving fast and furiously among one-offs or occasionally repeated projects with unique timelines and without the benefit of a manual on the shelf. This makes it challenging to aspire to a routine conducive to work-life balance. On the flip side, you're never bored, no such thing as a dull moment. We used to joke that staff in our teams like to break their heads on a brick... but a different one each day. All that calls for a certain employee profile. Notwithstanding the uncertainties, the constant commotion and the deadline pressures, the multiple human interactions, the versatility of the issues dealt with, the multi-tasking involved, and the plunge into uncharted territory can be rewarding for the right personalities. This environment can also be a school for younger recruits. Experiences from early project negotiations to brainstorm-

ing for content determination, methodological choices, data collection or acquisition, processing, analysis and dissemination can sometimes all be condensed in the course of a few months. In a nutshell, such experiences shaped my appreciation of all kinds of statistics, the design of programs, and the nuances of data use.

**Walter: In reading your book Number Savvy, I got the impression that it seems to answer very much your very personal interest in data and statistics. Beyond this, it would be interesting to know how it relates to your career in general and in Statistics Canada**

*George:* StatCan has been home but I've also had opportunities to experience the world of data at large. I was invited for assignments at the OECD and the UNESCO Institute for Statistics, collaborated closely with Eurostat and other EU institutions, the UNESCO Network of Chairs in Communications (ORBICOM), numerous global and regional UN bodies, and led research teams in many countries, including for G7 and global summits. I've participated in capacity-building exercises involving several statistical offices on all continents. As well, I was invited for an assignment at the International Development Research Centre with the express purpose to help strengthen the statistical capacity of their research networks in the Global South. In parallel, I taught a variety of courses in several university departments for decades now, and through this I learned how to unpack and explain complexity. Naturally, my interactions with people from all walks of life over the years informed my interests and helped shape my views on statistics.

In truth, things weren't always rosy in the world of data. Not that long ago, data were something of an afterthought and much energy was spent to spread the word or for a seat at the table. While sometimes I felt openness for, genuine curiosity about, and occasional admiration for data, oftentimes I saw a lack of appreciation and respect, outright fear or even disdain. I've been on the receiving end of numerous questions regarding the production of data, their use in policies, in business decisions and elsewhere or the consequences of the absence of data. I experienced from close up the apparent paradox of why in an era of data abundance we also feel data shortages and witnessed firsthand the recent transformation of attitudes towards data. I've also been exposed to esoteric arguments or conjectures bordering conspiracy theories as to what data we produce, what we don't, who decides, why, and how we should move 'in with the new and out with the old'. Such matters intimidate many people and even confound mature

data connoisseurs inside statistical outfits. Obviously, they're of interest to data producers and users alike. A better understanding of how data come to life and what they mean, in conjunction with macro forces at work, can make deliberations on how our societies move forward more meaningful. So, the topics dealt with in the book reflect all those life experiences. The efforts to demystify statistical processes and popularize data are meant to remove misconceptions and fear by raising numeracy.

**Walter: I have seen in recent years several books been written on data. That makes me curious to hear why exactly was this book written**

*George:* Thank you for the question. It really goes to the heart of my thinking in the early days of writing, even before I realized it may become a book. I'll sum up the reasons in three groups: the continuous quest for improved numeracy, the desirability for meaningful communication between old and new data players, and the need for the perspective of a data practitioner.

First, our societies have done a good job in the area of literacy but, unfortunately, not as good in numeracy. Historically, we've experienced a pervasive lack of real understanding of numbers. Frequently, a certain amount of confusion and sometimes even fear surrounds numbers – and, by extension, statistical data. Even highly literate people still struggle with numbers, shrugging them off – regrettably, sometimes even with a sense of pride! (“*I'm not a numbers' person*”). However, for some time now, the functioning of our economies requires increasingly higher levels of numeracy. Thanks to many initiatives and the recent explosion of interest in data things are improving but our societies won't become numerate overnight. Those who work with data have a role and a responsibility in raising the level of numeracy. In that vein, this book adds to the long list of such efforts and aspires to contribute a proverbial stone.

Second, from an earlier use in rather narrow circles and relative seclusion data are nowadays all the rage. In addition to their geopolitical significance, the newfound love affair – even hype – with data implies a much bigger tent for those involved. Old faithful believers rub shoulders with born-again converts to the data cause. In that setting, mathematically, there's more heterogeneity. It's only natural that some know more about data than others or care about different things than others. Several distinct groups now vie for attention but communication is not always easy as they lack a common understanding and a shared nomenclature. The survey methodologist and the API programmer may both aspire for the newer title of 'data scientist' but such solitudes present chal-

lenges. By explaining how we got to the junction we are today and what it means this book aspires to facilitate much-needed communications among those under the new and wider tent in the world of data.

There is a third reason for writing this book. Many of the books I've come across have a mathematical bent or are technical in nature, some aim to put people more at ease with numbers and orders of magnitude, while others offer useful advice on the misuse of data to help fight misinformation or exalt the virtues of new 'big' data from search engines and social media. As an avid reader myself I've enjoyed many of these books. I benefited even from those I didn't quite enjoy if only because of the thinking they triggered. While many times I wished I could give the authors a pat on the back, other times I wished I could talk them out of something – but, of course, I couldn't step in the books of others. This came with the realization that my intended interventions had less to do with the contents addressed and more with the perspectives of the authors since most books I'm aware of have been written by observers of the data world, including academics and journalists. I felt that a particular gap existed in how data are produced and how they're linked to the socioeconomic notions they estimate. We know that data don't fall like manna from the sky nor do they magically appear on the shelves of a statistical outfit. Yet, both their existence and their *raison-d'être* remain a mystery to many, acting as an extra impediment to numeracy. I've been a data practitioner for most of my professional life. My affinity for data extends from their abstract conceptualization to their production and analytical use inside statistical 'ecosystems'. So, here I add my own perspective to the world of data.

**Walter: The structure of your book is very interesting. For example, you start the book by discussing the invention of numbers in the opening chapter. Why do you think this rather elementary chapter was necessary to include?**

*George:* Fair observation, it's an unusual chapter indeed in a book about data. Well, the early reason is so that we do not forget that obviously our statistical data are direct descendants of numbers! Differently put, both our numbers and our statistics today originate from the human need to quantify our world – in conjunction with the absence of a biological organ to tell exact quantities. It must be well understood that they're all human-made to serve our purposes. Then, a more important reason is that there are parallels between the development of numbers and the conceptual underpinnings of most socioeconomic statistics. Understanding our ancestors'

thought processes behind the abstraction of *numbers* and the tribulations that led to the eventual evolution of our number system in the first place serves as an intuitive bridge to comprehend the mental exercises that provide the foundations behind today's socio-economic measurements. In that sense, the narrative behind the origins of numbers and the conventions of our number system equip the reader well to explore the conceptual real estate of our socioeconomic statistics and the conventions made for their estimation.

**Walter: Another striking issue in the structure is the Fact-Checking Tips at the end of each chapter. What's the reason behind this, and how did this come about?**

*George:* There are a couple of reasons for that. First, we live in a period with instant access to massive amounts of information and knowledge. While liberating and wonderful, this also poses certain challenges. For instance, information overload at times overwhelms our capacity to absorb it. More relevant to your question is that, unfortunately, the accessible package also includes misinformation and disinformation – that is, whether accidentally or maliciously, truths are mixed with falsehoods. Examples abound, from science to politics to daily gossip. Clearly, in this environment, the ability to separate the real from the imaginary and tell truth from lies becomes a critical function. Indeed, individuals and organizations that perform fact-checking in general already exist. I expect this to intensify as information holdings expand by leaps and bounds.

This brings me to the second reason. All the above apply equally to statistical data too, for the simple reason that they're part and parcel of information holdings. It wouldn't be surprising to see the emergence of dedicated resources, specializing in fact-checking numbers. Whichever way it's done, the skills for this task are different from those needed for non-quantitative information. Being number savvy is instrumental for fact-checking statistics. Therefore, these sections of the book offer a bit of a primer with examples of how to hone our skills in data sources, methods, interpretation, and so much more.

And one more thing. In the final analysis, we live in a reality in which when confronted with a clear truth and an unambiguous lie a few people may still opt for the latter. This flat-earth syndrome is a serious problem – but unfortunately way above my talents and beyond the purview of this book. Aside from joking about some Ministry of Truth, I can only hope that there are people in our midst who can sort this out before we find ourselves in the abyss. Until then, we must do all we can –

to arrive at a clear distinction between what's true and what's a lie.

**Walter: Statistical literacy is currently a very important issue, in general, but also very relevant for policymakers using statistics. Your book clearly contributes to the discussion on statistical literacy, but in a very explicative way, by explaining in great detail the backgrounds and procedures in how statistics are produced. How do you think statistical literacy should be tackled?**

*George:* Statistical literacy is a heavy-loaded question. As we've already discussed, it's effectively been one of the reasons behind the very existence of 'Number Savvy'. So, on top of all the deliberations and initiatives I've been personally engaged in over the years, I've recently had the chance to reflect a bit more.

Much like literacy, I've always approached statistical literacy as a continuum rather than a dichotomous Yes or No. Under this view, there are different requirements for different functions or roles. For example, professional statisticians and data producers will possess skills to occupy segments higher up the continuum. Policymakers and heavy data users also require substantial skills to comprehend, manipulate, and even add value to the data. But statistical literacy is important for the average citizen to function in today's world too. A crude way to describe the differences is to think of an engineer who designs and builds a car engine, a mechanic who knows how to repair it, and a driver of the car. Although a very rough analogy, it conveys a key idea – that the objective is not for everyone to become a statistical guru! We don't put the bar that high for law, accounting, medicine, or other areas of knowledge, which is why we have lawyers, accountants, doctors, and other specialists. However, our societies need a critical mass of all those. It's also evident that the needed critical mass for statistical skills has gone up in recent years due to the explosion in the economic and geopolitical significance of data.

Now, moving up the continuum of statistical literacy isn't something that can be done overnight through some single activity. There's no quick fix, such as creating and delivering a miracle course. Drawing lessons from a closer inspection of recent progress would reveal a long-term process with persistent and concerted efforts by a multitude of institutional players. The key is to start early on. Appropriate interventions in the education systems at large, and school curriculums in particular, can improve comfort with numbers and statistics and boost skills over time. The media have a particular role to play in popularizing data. So do statistical

outfits, with the added responsibility of uplifting the statistical literacy of users. Books, articles, online resources in the spirit of lifelong learning and the like, all help. The continuous showcasing of, and exposure to, societal benefits brought about by the production and use of statistics will also foster a numerate populace. Only then, con artists would be out of business with no one to con.

Statistical literacy is also indispensable for policy-making. Therefore, all levels of government have a big role to play too – certainly through their hiring and training practices but also through their cultural attitudes towards data. Needless to say, the official statistics ecosystem has much to offer here, particularly through its interaction with ‘government’ at large. At the risk of sounding fancy, we can think of relevant initiatives as the ‘servitization of statistical outputs’. These include the cultivation of cultures based on a true appreciation of both the production and use of data, complete with the promotion of metadata. Depending on the country, practical implementations can include two-way exchanges of personnel, courses aimed at government employees, including at schools of public service, seminars, and specific initiatives, such as embedded data ambassadors, networks of departmental chief data scientists, or comparable initiatives.

I hope that the book can also make a modest contribution to overall efforts aiming at statistical literacy in many ways. Among them: it explains and promotes the importance of primary data sources and metadata, it discusses many of the issues involved and the skills needed to follow data as they move around from one source to another, it demystifies processes hidden to many by connecting the conceptualization of statistics with their measurements in practice, it makes a strong case for the appreciation of microdata, and shows what to look at when it comes to the analytical interpretation of data. It even extends the notion of statistical literacy beyond data, by calling for a societal re-think of attitudes towards privacy and confidentiality that will ultimately define future legal frameworks.

**Walter: The book is using a lot of examples and the explanations are very ‘down to earth’. Is this because you expect this book to be used also in lecturing in a general course in universities and maybe even secondary schools?**

*George:* Actually, let me clarify. If the use of examples and my explanations are ‘down to earth’ (which I take as quite a compliment), they probably reflect the perspective of my years of teaching. I wasn’t born with such innate talents but I do recall that early in my ca-

reer I was troubled by comments like “*I can’t explain more, it’s very technical, need to understand the math or the science behind*”. Such inability to explain some ‘scientific truth’ to lay people never sat well with me, to be honest. I strove hard to find ways and express in plain language research findings and unpack ostensibly complicated concepts to family and friends. To this day, I believe this is one of our main responsibilities as researchers. Over the years, I realized that the more I understand, the better I can explain.

Next, surely I can see course instructors including the book in their reading lists, whether in its entirety, a chapter or two, or just certain sections. In fact, although the book is fresh off the press, this has already started. Over the decades, I used all kinds of reference material in my own courses. However, my main intent in writing this book was to popularize data and contribute to the level of overall numeracy as we discussed. That’s why *Number Savvy* is a book, not a textbook for students or a manual for experts. Having said that, it’s not for everyone – in fact, I don’t think any book is for everyone! But it’s for many. It’s for those who produce, compile, or analyze data, employees at all levels of government, including national statistical offices, international and non-governmental organizations, and others. It’s expected to be useful and appeal to socioeconomic researchers, mature and intrepid. Finally, care has been taken that it doesn’t contain math, formulas, and technical jargon – with minor exceptions, mainly restricted to footnotes. So, in my opinion, it’s also written so that it’s within reach of anyone who wants to become more number savvy in today’s world.

**Walter: The book makes explicit connections between statistical data and socio-economic research. It discusses how technology, changes in the production of official statistics, and new data impacted research paradigms, and devotes enough time to their evolution from early in the 20<sup>th</sup> century until our times. Can you elaborate on the significance of this?**

*George:* Certainly. The relationship between socioeconomic research and statistical data is symbiotic, one feeds off the other. To illuminate issues or understand phenomena of interest, research uses existing data as inputs and produces new data as outputs. Over time, this relationship underwent a paradigm shift, and the book describes *how* and *why*.

In the past the compilation of desired statistical outputs relied on surveys, which became ubiquitous measurement instruments in the 20<sup>th</sup> century. However, what I called *Olympic ideals for data* many years ago – to describe light-heartedly how many of us internal-

ized the vocal demands for more and better, faster and timelier, more granular and cheaper data, among other got-to-have attributes – were antithetical to surveys. The gradual substitution of administrative and other data sources for surveys that had already begun, now intensified and became ongoing. Such steps are not only irreversible but one invites the next leading farther away from surveys. For instance, switching to scanner data for some CPI items is soon followed by data from smart meters, smartphones, or some other source for different items. A similar logic transmits best practices from one country to the next, following a demonstration phase. It's like our world is a giant laboratory where different jurisdictions try different things and see what sticks.

Data amassed through increasing digitization point to the same conclusion. In the past, knowledge of a particular purchase was between someone and a store owner, now many know – to fulfill, pay, and deliver to our doorstep. With all these data available, who needs a survey? So, if not our actions, interactions or transactions, what will be left to ask? Our inner thoughts? And would it be those we share with social media? In any event, without the need to proclaim the death of surveys, other things equal, they will become more targeted and rarer.

To illustrate with a stylized example, if you visualize the desired statistical output as a matrix, each and every cell was filled with data from a survey. Now, different cells of the output matrix can be populated from different sources. Data production morphed from a rather linear and controlled process, with a high data-to-information ratio, to a more roundabout and complex process involving multiple data sources and players. The ramifications of this paradigm shift are many and are playing out everywhere. Key among them:

- pressures inside statistical offices for new, faster, and more granular data
- some distance is created between data and research, in the sense that data compilation and curation assume an independent existence not necessarily linked to immediate use or awareness of a specific future use
- new research possibilities open up because the sheer availability of new data kindles freer thinking, no longer constrained within the confines of survey data

Many more detailed issues lurk below the surface and vie for attention, including matters of nomenclature and communication among the players in the new data ecosystem, data ownership, privacy concerns, and a

whole lot more. All that still needs to sip in and be absorbed by a critical mass of people.

**Walter: The book also devotes a full chapter to microdata. Why are microdata singled out, what's their significance in the world of data, and why should we care as we move forward?**

*George:* The main reason is a pragmatic one. Microdata are foundational to any future of data that I can imagine. Thinking a bit deeper about the paradigm shift just described, the move away from surveys to administrative and other sources points exactly to the direction of microdata. Whatever microdata were collected through sample surveys couldn't really stand alone and only served the construction of aggregate estimates. However, most of the new sources are census-like collections of microdata, and they can be used to craft alternative datasets.

I think there is a sentimental reason too since, in my view, microdata have historically been the underdog in the world of statistics. For instance, I've seen firsthand the shunning of microdata inside statistical offices. The book articulates several reasons behind such treatment. The pre-digital world of paper and print media was not conducive to microdata. Numeracy was quite low. Privacy matters were barely even discussed. Neither data producers nor data users could really handle microdata. Conveniently, everyone over-relied on aggregates. Each of those barriers that kept microdata out of the spotlight is either gone or in the process of being phased out. Digitization, including the huge drops in storage and processing costs, has had a transformative impact. Numeracy is generally improving. Approaches for researcher access to microdata, while protecting confidentiality and privacy, have emerged. By now many users are capable of manipulating and analyzing microdata, and some power users possess quite sophisticated capabilities. What's more, as discussed in the research paradigm shift, these extra capabilities breed extra demands for microdata, including through data linkages – which, incidentally, also point directly to microdata.

All in all, my reading of the forces at work is such that microdata will occupy a central place in the future world of data. Moreover, we've already started to see a lot more data below the micro level. Volumes of records from all kinds of sources, such as satellites, sensors and the Internet of Things require aggregation to arrive at today's microdata!

**Walter: In addition to explaining how data are really produced and what they mean, the book goes on to discuss the derivation of insights through the use of data in analysis. As this could well be the subject of an entire book, what are the key takeaways?**

*George:* There's a lot to unpack here. Data analysis is essential to move us higher in the pecking order of the pyramid, from data to information to actionable knowledge. Arguably, data analysis epitomizes the pinnacle of data work and is oftentimes seen as the reward for the pain to produce new data and document their metadata. Data analysis today should not be perceived as a singular activity, such as writing a 'paper' – a common output of the recent past. Nor should it be perceived as the final activity in some linear data life. We're better served to understand it as a series of value-adding steps in an endless circular loop of work involving data. Simply put, analyzing data can entail any number of things, such as digging deeper into datasets, transforming or integrating data series, and so much more. Disseminating analytical findings can equally assume many forms today. What really matters is that the insights gleaned during analytical activities feed back to earlier data work, including methodologies employed in data processing and the like, extending all the way back to the very conceptualization of variables. Particularly in repeat statistical operations, with or without a fixed periodicity, this feedback mechanism leads to refined concepts and improved data, which in turn improve subsequent analysis, and so on and so forth. That's one takeaway.

Another takeaway is more subtle, and not much discussed. It has a lot to do with establishing the factual basis of our world. Generally, analysis can be descriptive or inferential. In practice, the former is better understood as a way to organize and present data for dissemination purposes. Data can be aggregated or decomposed in any number of ways. Whichever way we choose, assuming they're error-free, these data are solid and irrefutable. They establish baseline facts. "*You're entitled to your own opinion but not your own facts*" comes to mind. Data that stem from inferential analysis are a different breed. Surely, they're derived from the same underlying real data but, regardless of the sophistication of the techniques to which they were subjected, they're not unique. This brings me to the relationship between inferential analysis and almost-philosophical matters of truth, probabilistic truth through hypothesis testing, and the like. The book argues that such twists and mind games are interesting over a glass of wine at night but not useful during our working days. In the process it touches on the peculiarities of the perceived meaning of probability by humans, beyond cold math.

An additional takeaway has interesting implications for the future. Data analysis typically tries to answer a question, policy or otherwise, test a theory, or validate a framework. The sheer abundance of new data and their

independent existence have given rise to talk concerning the 'death of theory', meaning that insights can be derived from analysis triggered without an underlying question. This will intensify and, ideally, it can be combined with imaginative approaches to data analysis – examples of which are discussed in the book.

*Walter:* **We live in a period inundated by data. As well, more than ever before, data have entered the public discourse and have become significant components in multiple and extensive conversations concerning the geopolitics of our times. The subtitle of the book includes "the future of data" – which is also the title of the book's closing chapter. So, what do you think the future holds for data?**

*George:* As I wrote in the book, I'm not a futurologist and I have no crystal ball. However, anyone working with data in the last twenty years is surely aware of forces underway that have already brought about significant change. We can peek more into the future by simply following these forces.

There's no doubt whatsoever that the future will be awash with data, the signals are everywhere. We're only at the end-of-the-beginning of the data revolution. Seen as a strategic resource, and central to the geopolitical AI races of our times, available data will be used. Therefore, the future will definitely contain everything related to *what* data will be involved and *how* this will happen. Transforming raw data into usable outputs involves a lot of code, which subsumes methodological choices. In that sense, algorithms will be the new currency. Multiple statistical outputs can be constructed from any data source and while, in principle, a processed data file could serve multiple needs, it'll be difficult to avoid duplication or mix-ups in the early go. For instance, processing scanner data for the purpose of CPI prices will not serve the needs of research aiming at the time of shopping, say by day of the week and time of day, or research looking at methods of payment, such as credit, debit or cash. The plethora of data sources, private and public, the two-way data traffic between business and government or business-to-business, whether legislated or market-driven, as well as the demands from the research community will be crossing each other and become unruly. How will all that be managed is a real issue – as is *who* will manage it? The European General Data Protection Regulation (GDPR) even calls for a distinction between data controllers and data processors. While still far from the elusive 'collect once and use for ever', the data space will be crowded and anarchic, reflecting data opportunism and lawlessness. Unless – or until – a better way is found. That's where we may

see the emergence of proper governance through *trusted third parties* that would house such functions.

It goes without saying that continuous technological evolution will also exert an independent and sizeable influence going forward. For instance, technology will undoubtedly enable higher automation in data production, as usable datasets are produced from multiple sources. But it'll do more. As data proliferate, it'll become more difficult for individual users to handle the volume. And, of course, AI will continue to be mainstreamed as it attempts to emulate and surpass human intelligence. All that means that machines will also become *users* of data, ingesting and further processing data produced by other machines – data by machines, for machines. Moreover, the cohabitation of many players in the new ecosystem will breed data 'frenemies', collaborating and competing at the same time. I also see the emergence of *data boutiques* as well as *data resistance movements*. 'Truths' derived from analysis will be more elusive, and we'll also have much more gaslighting – which must be fact-checked.

At the end of the day, as the book states, there's no karma leading to a unique destination. There's no future path for data independent of our responses – which are far from clear. A lot depends on how we deal with microdata, and who shares what with whom. Progress must be made on societal approaches related to issues of data ownership, markets for data, confidentiality, and privacy. Quoting from the book, if "*what lies ahead is a big leap from what's left behind*" existing legislation and professional ethics are necessary but not sufficient conditions to break through such matters, and a "*more supportive cultural reawakening with changes in societal attitudes*" is called for. This requires a much more informed and wider dialogue on the production, use, and value of our data resource.

**Walter: There's currently a lack of clarity about the future role of official statistics. What are your comments on this?**

*George:* First, I must confess that I haven't been a fan of the term 'official'. (In recent Eurostat papers, I've seen suggestions to replace it with 'Smart Trusted Statistics'). What really matters, though, is to contrast notions like good and bad statistics, appropriate/fit for use or not, with sound and transparent methodology or not, well-documented with metadata or not, and the like.

On the substance of the question, I believe that in the foreseeable future the fate of official statistics will mirror that of all statistics. Thanks to the explosion of data, and in the spirit of the tide lifting all boats,

it will be a period of growth all around. Demands for statistical products will intensify. Institutions of national statistical systems will grow in importance. In parallel, demands for data will also diversify, extending to non-traditional areas. This will have implications for data sources and production processes, as well as human resources and the composition of required skills. In other words, statistical institutions will feel more of what they're already going through for some time.

Another thing as we move farther into the future is that instinctively thinking of the statistical office when it comes to data will no longer be the case. This should not be approached sentimentally but as a mathematical fact. As the volume of data grows, in tandem with the proliferation of data sources and players, national statistical offices will inevitably have a smaller 'market share'. But no premature announcement of death is warranted! Even in distant times, so long as organized human societies exist 'official' statistical institutes will be present. It's also quite possible that in some countries they become those trusted third parties or assume some of the desired functions we discussed. It's not inconceivable that organizations of official statistics shoulder extra societal roles, such as helping manage data exchanges and data access at large or perhaps getting involved in certification and the wider setting of standards that would help fact-checking.

And one more thing. National statistical systems, whether centralized or not, are already in a period of introspection coupled with modernization efforts. Beyond functions and processes, it's worth prognosticating the future composition of the outputs of official statistics. Many such outputs don't serve ephemeral needs but have become fixtures for decades. This is a good thing as the continuity of outputs used in policymaking is reassuring to governments. However, outputs cannot be allowed to fall victim to inertia and overstay their welcome. This is not a good thing. Moreover, as evolving demands add new outputs to the mix, it's not sustainable. Simply put, we can't continue piling up indefinitely, something has to give. In the absence of well-functioning data markets, the onus for this is shared with policymakers. Prioritization can be guided by comparative relevance, lest we become enamored with our creations. Just think that for millennia humans were born, lived their lives, and died without ever meeting our most cherished statistics – never heard of the GDP, the CPI, or the rate of unemployment. There are reasons why all those and more are the creations of the 20<sup>th</sup> century. As the book says, rewinding back in time, many things we know today disappear and others we



don't know appear – there's no smartphone, film, or pensions but there's a kerosene lamp and a coal iron. As we now fast forward to peek ahead, looking back from there we must equally come to terms that the same will happen. Renewal should be the name of the game.

**Walter: At the very end of your book, you discuss the by some so-labeled 'end of theory' and the transition to an era in which analytical findings are fully driven by data. Doesn't official statistics also have to solve another problem, the rapid change in information needs and ensure that work program updates and priorities are developed and decided democratically?**

*George:* This is a complex question! To situate it properly, allow me please to take a step back and clarify a couple of things. First, open democracies are more conducive to the production and wider diffusion of statistics than authoritarian regimes that are generally antithetical to free information flows. The book mentions that even when ruling monarchs of the past produced data they were used for their own benefit rather than as a shared resource. Second, the question is focussing on official statistics, whose predicament can be influenced more by state institutions and actors than market forces. All data, though, can still be within the scope of an encompassing future framework.

Keeping these remarks in the background, statistics have historically evolved to serve the needs of the economies and the societies of their times. Presumably, in democracies these are expressed by elected governments. Indeed, the meteoric rise of statistics in the 20<sup>th</sup> century coincided with the rise of modern governments and elevated informational requirements. To remain relevant and sustainable not only our statistics but the entire statistical ecosystem must be responsive to change and subject to continuous renewal. In that context, the questions of how decisions are/ought to be made and who is making them are quite valid and worthy of public discourse. As a general statement, and if the past is a guide for the future, the evolution of statistics will continue to reflect the democratic choices of their societies. Some clarifications are in order, though.

For one, the desirability that our statistics reflect democratic choices is substantively different from their political manipulation. It is well understood that official statistics must be impartial, free from political interference, and not subject to short-term partisan politics. This applies to all aspects of statistics, from the establishment of programs and processes to the production of specific outputs. Statistical programs have generally not been accommodating to quick maneuvering and there-

fore have not been conducive to short-term interference. This was particularly so when they used to be survey-based, as the time it could take to develop, test, and implement a new large-scale survey could be comparable to the life spans of many elected governments. Political meddling could easier target statistical outputs directly, though. Sadly, there have been several publicized examples to that effect. . . these represent serious blows to official statistics and bode ill for the future – especially when alternative data sources proliferate. What's worse, not only it takes a long time to recover from such a reputational loss but it may lead to contagion, that is, rubbing off to the reputation of others.

Broadly speaking, it is important to take a longer view of how official statistics reflect democratic choices without falling prey to election cycles. This is precisely where the issue of the independence of statistical offices enters the scene, while still remaining integral parts of governments for obvious reasons. Then, democratic choices, as represented by elected governments, will dictate over time *what* to measure but certainly will not extend to the *how*. The application of scientific and methodological standards that underpin the integrity and credibility of statistics is best left to the statistical office. Under modernization efforts, statistical programs are becoming more agile and therefore also more prone to meddling – beyond the control of the purse strings through budgets alone. The 'implied', *de facto*, arms' length relationship between statistical authorities and the rest of the government is being replaced with independence through legislative actions – somewhat akin to the autonomy of central banks much earlier. This has happened in some countries, including in Canada recently to address unease in the air following an unfortunate incident involving the 2011 census. In addition to legislative acts, safeguards extend to broader governance arrangements inclusive of researcher communities and other societal voices through consultative committees, advisory councils, and the like. In the final analysis, societal choices will be made as in all other areas – through collective will, aided by (ideally) well-functioning institutions and processes, as well as evolving data cultures. Without getting into specifics, in Canada, we like to think of Statistics Canada as the *national* rather than the *federal* statistical office.

Walter: We have come to the end and I would like to thank you for the interesting answers. I am sure they reflect a lot of what you have written in your book, but I also feel they also add interesting background information about your motives and experiences for writing this book, which is especially helpful for readers and especially those that would like to use it in training in statistical literacy.