

Catalogue no. 11-633-X — No. 034
ISSN 2371-3429
ISBN 978-0-660-40764-7

Analytical Studies: Methods and References

The System of National Quality-of-Life Statistics: Future Directions

Release date: January 31, 2022



Statistics
Canada

Statistique
Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "[Standards of service to the public](#)."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

© Her Majesty the Queen in Right of Canada as represented by the Minister of Industry, 2022

All rights reserved. Use of this publication is governed by the Statistics Canada [Open Licence Agreement](#).

An [HTML version](#) is also available.

Cette publication est aussi disponible en français.

Table of contents

Executive Summary	4
1. Introduction and summary	6
1.1. Why change? With what payoffs?	6
1.2 The system of national quality-of-life statistics in summary	10
1.3 Managing constraints and risks	11
1.4 Assumptions about an uncertain future and the need for flexible planning.....	11
2. Outcomes: the purpose and scope of the evolving system of quality-of-life statistics	21
2.1. Supporting policies and programs	21
2.2. Supporting public and private decision-making	22
2.3. A statement of vision and mission – and a new title	25
3. Outputs: statistical products	27
3.1. Progress in building the interactive web site.....	27
3.2. Access to tabulated data: progress to date and next steps	30
3.3. Access to microdata: progress to date and next steps	30
3.4. Partnered ‘what is likely to work best’ databases.....	32
4. Processing: transforming inputs to outputs	33
4.1. The conceptual framework.....	33
4.2. Managing and processing the data	35
4.3. Data documentation.....	36
5. Inputs: the multiple sources of data	36
5.1. Responsiveness to requests for new statistics	36
5.2. Response burden and hybrid collection vehicles.....	37
5.3. Obtaining new administrative data	38
5.4. Incorporating geographic information in individual records.....	39
5.5. Inputs describing social programs and services.....	40
6. Managing constraints and risks	41
6.1. Skill shortages	41
6.2. Building understanding and support in the user community	42
6.3. Ethical issues: building trust	44

The System of National Quality-of-Life Statistics: Future Directions

Executive Summary

In response to current demands, Statistics Canada is engaged in a long-term exercise to greatly strengthen the statistics that it produces to measure the quality of life in Canada. Compared with current arrangements, these statistics will be more timely, more granular, more responsive to the needs of users, and more easily accessed.

In line with directions set out in Budget 2021, the system that is being developed will provide comprehensive, useable evidence on:

- The inter-related factors, including non-economic factors like health, housing, environment, and safety, that result in a high quality of life, including the effects of the policies and programs of all orders of government on people as they move through the course of their lives.
- Equality, showing the distribution of outcomes and opportunities across places and people, taking account the great diversity of the Canadian people and the need for analysis, and action, at the level of small population groups including those facing barriers to a high of quality life.
- Sustainable outcomes, assessing how current actions will support high quality of life in the future.

Over time, the system will also begin to provide evidence that can improve decisions about the particular social interventions that are likely to work best for specific individuals – with that information being accessible at the time when these decisions are being taken, based on analysis of what has worked best in similar situations in the past.

The resulting evidence will support government budgeting and accountability, the design of effective programs, and efficiency in their administration. Decision-making by individual citizens and service providers will be supported with practical ‘what is likely to work best’ data. The outcome will be major improvements in the operation of labour markets, health, learning and other social dimensions of life in Canada, both on average and for all population groups – and direct benefits to individuals as they make big decisions in the social, health and labour domains of life.

Unlike the existing social statistics, which are based around particular data sources such as surveys, censuses and summarized administrative records, the new system will provide far more detailed and timely information by drawing on data from multiple sources. These sources will include a much more flexible suite of surveys, finely-grained information about operation of social programs, and anonymized data about individuals that currently exist in administrative records. A multi-source system is possible as a result of the major improvements in digital information processing technology that have been introduced in recent decades, with promise of even more powerful tools and data sources on the horizon.

The foundations for this transformation are already in place and the priority on speedy implementation has increased as a result of COVID-19 and the directions outlined in Budget 2021. Many payoffs become visible over the next several years. Others will take longer to be fully realized.

One of the most visible early products will be the creation of an interactive website that will present high-level quality-of-life indicators that reflect current policy priorities, as well as a series of practical dashboards of indicators designed to support a wide range of current policy agendas for users both in and out of government. Current statistics, such as the latest unemployment rates, would be released on this site where they would be analysed along with indicators related to other dimensions of wellbeing.

The site will provide direct access to interactive tables along the lines of those on the present Statistics Canada site, except that there will be many additional tables with newly-created data, based on consultations with users, that would be drawn from different sources. This is in opposition to the present practice of presenting tables based only a particular survey or other data source.

In addition, the new techniques for combining data from multiple surveys and administrative data sources without any of the resulting data being linkable to real people opens up the possibility of developing powerful new kinds of data products for release on the web site. These are known as standard, multiple-source microdata files. These contain particularly rich data that can be used to provide a new understanding of the factors that affect the quality of people’s lives over time, including the role played by different social programs and services.

Another early payoff will be greater flexibility in responding to requests for new statistics. The interactive web site will, itself, provide far more information than is currently available. As well, many requests for new data will be met by combining existing data from different sources, without time-consuming and costly new data collection. When new data collection is needed, a flexible suite of new survey vehicles is being developed to allow fast collection of information that is finely-tuned to current needs.

Other statistical products will take longer to produce. For example, some will require developing consistent, finely-grained ways of describing the existing social, learning and health services. Such standard descriptors are needed, for example, in order to provide consistent ‘what is likely to work best’ information that will support service providers and individual citizens in making the best choices among the interventions that are open to them. In some areas, the needed information already exists. In others, much development work will be needed.

This paper describes the system as it may look in some five or so years time assuming the continuation of present plans. It does this by examining the likely future inputs, processes, outputs and outcomes of the system, and compares them with the arrangements in the traditional single-source system of social statistics that has, until recently, been the norm.

The paper also describes the main constraints and risks that will be faced as we move ahead and how these will be managed.

In terms of resource constraints, funding will come from internal reallocations and new funds outlined in Budget 2021. More powerful information technology (IT) resources will be required, but these are already being used in other applications. The biggest constraint is likely to be the shortage of people with the skills needed to produce accurate statistics from multiple sources of data, often involving large data sets and complex statistical manipulations, while ensuring privacy. New approaches will be developed to increase the size of the pool of people with these skills and to maximize the use of those skills in both Statistics Canada and in the user communities.

New ways of doing business are being proposed. The largest risk to be managed during this period of rapid change relates to the need to maintain and reinforce the trust and confidence of the users of statistics, the individual Canadians who respond to surveys and the agencies who provide the needed administrative data inputs. Managing this risk will require different kinds of outreach to explain how the new arrangements will work, how privacy and confidentiality will be assured, and how all parties will gain over time from the new approaches.

As well, the risks associated with the introduction of a new product line – i.e., the use of predictive analytics to produce ‘what is likely to work best’ evidence in support of decision-making about particular social interventions – will be mitigated by a gradual, experimental approach, working in partnership with different service providers.

In terms of social acceptance and trust, Statistics Canada has recently put in place a world-leading framework (called the Necessity and Proportionality Framework) for balancing use, privacy, and response burden. Sophisticated tools have already been developed that ensure privacy protection when linking data from multiple sources, using techniques that split out individual identifiers from the data that describe those individuals. Response burden will be reduced by lessening the demand for repetitive inputs. New emphasis will be placed on the role of Statistics Canada itself in producing statistics that are based on multiple sources of data, with built-in assurance of both quality and privacy protection, reducing past reliance on external users to carry out some of these functions. Greater use of synthetic data will simultaneously produce useful new data and assurances of anonymity.

Perhaps the most important new payoffs in terms of building acceptance and support will lie in the production of new statistics that will be directly useful, not only to researchers and policy-makers, but also to those who administer social programs and services and to individual citizens. A more immediate and transparent balance can be struck among response burden, usefulness and the protection of privacy – strengthening the implicit social contract that exists among those who provide, process and use data for statistical purposes.

A separate paper, entitled “The System of National Quality-of-Life Statistics: Conceptual Framework”, will be released as a complement to this paper that will describe the conceptual framework that underlies the new approach. It will show how the standard definitions that are used to describe the individual items of data that are collected from different sources will result in a flexible system that allows integrated analysis across all the domains that affect quality of life, including linkages with economic and environmental statistics, compatibility with the frameworks used in many academic disciplines, and comparisons with other countries.

1. Introduction and summary

Deep changes are taking place in national social statistics. This paper presents a picture of what the system might look like in some 5 or 10 years should Statistics Canada continue along its current path of rapid evolution. Possible actions in the shorter term are described, as are strategies for managing the constraints and risks that inevitably arise in making such large changes.

A supporting paper will be available shortly that describes the changes from a more technical perspective. It will describe the conceptual framework that underlies the new system and how it results in a system that is both integrated and flexible. It will show how the framework is consistent with those used in economic and environmental statistics and in different scientific disciplines.

This introductory chapter is structured as follows:

- Section 1.1 summarizes the nature of, and rationale for, these changes.
- Section 1.2 summarizes the features of the system once it becomes mature and compares it with current approaches.
- Section 1.3 explains how the constraints and risks will be managed as the system evolves.
- Section 1.4 discusses the need for flexible planning that can adjust to the introduction of new data sources and technologies.

1.1. Why change? With what payoffs?

Deep changes are taking place in the collection and use of national social statistics, which – for reasons that will be explained later – will be referred to as the ‘National System of Quality-of-Life Statistics’.

- The system is shifting from the traditional approach where data collection, and the release of the resulting statistical information, was based on single data sources, such as a survey.
- It is moving towards a system that is based on consistent data obtained from multiple sources and that will, in consequence, be able to produce integrated, timely information about the inter-related dimensions of social well-being. That information will be released in an easily understood way that shows our progress towards a better quality of life and that identifies problems that need to be addressed by policy.

The new information obtained from multiple sources is changing our understanding of how societies work, particularly as we dig down beyond broad averages and examine the diverse needs and circumstances of people in the many groups that comprise our population. It is starting to provide a far better understanding of how people change over their lifetimes, including the role played by our social institutions and programs.

Box 1**The rationale for the change: some bottom-line Q&As****What new directions were launched in 2019?**

In order to meet new demands for statistical information, and as part of a larger modernization exercise, Statistics Canada launched a reform of its social statistics that makes use of current technology to shift from a traditional single-source system to one where statistics would be drawn from multiple sources, including much greater use of administrative records and a more flexible suite of surveys.

- The purpose was to better meet the currently expressed needs of users – producing statistics that were more relevant, timely and disaggregated.
- The new processes would also minimize response burden, increase trust in the statistical system and use state-of-the-art tools and methods.
- The ultimate goal was a higher quality of life for Canadians as a result of the improved effectiveness of policies and programs, and the improved decision-making in all dimensions of social life, that would result from the use of this new statistical evidence.

What changed as a result of the pandemic and Budget 2021?

The need for rapid responses to COVID-19 resulted in dramatic increases in the development of new data sources, the speed of turnaround times, and the production of disaggregated information that showed which population groups were at greatest risk. It demonstrated the validity and power of the new way of doing business.

Budget 2021 strongly reinforced the emphasis on producing policy-relevant, disaggregated statistics. It also stressed the importance of producing evidence about the multiple factors that impact on quality of life and about the effects of policies and programs in supporting a better quality-of-life, including effects on equity and sustainability.

Who will gain from the new system?

Policy makers and analysts will have much better evidence to support planning, budgeting, evaluation, and accountability.

- In the shorter term, the production of more disaggregated data for small population groups will result in major gains in understanding the needs and circumstances of people with diverse characteristics and of the effects of programs in addressing those needs.
- Over the longer-term, as the system becomes mature, new evidence will be created that shows how the many social services and income support programs of different orders of government work together in influencing the subsequent lives of the individuals being served.

Some of this causal information exists today for income support programs, but there is little evidence about the effects of service programs or how different programs act in combination.

Box 1

The rationale for the change: some bottom-line Q&As

Program designers and administrators will have much better evidence about the characteristics of the intended audience for their programs, the actual effects of existing programs and which program designs are likely to work best in the future.

- In the shorter term, the main payoffs will come from better evidence on which programs are working best for people in diverse circumstances and in making adjustments to program designs or administrative process to better meet those needs.
- As the system matures, there will be steady improvements not only in the evidence about what programs designs and processes are working best on average, but also which components of the program are working best for whom, and which components need improvement.
- Even more important in the longer-term will be new predictive information that can be produced for some types of programming. This will show which kind of program interventions are likely to work best for particular clients during the time the referrals are being made. For example, a counsellor or case manager would be able to use this kind of information in advising clients or referring them to additional services.

Researchers in many disciplines will greatly benefit from the transformative increase in the amount of data that will be available, particularly longitudinal data at the micro level which will allow analysis of the factors that affect the course of people's lives in their social and geographic contexts.

Businesses and organizations will, of course, gain from the better outcomes that will result from more evidence-driven policies and programs. In addition, they will be able to access more disaggregated information for marketing purposes, namely the characteristics of the potential audiences for their products and services.

Individuals will similarly see their lives improved by programs and policies that will be shaped by more disaggregated evidence that focusses on quality of life. This will be especially felt by people in visible minority groups or in particular parts of the country where needs may be quite different from those identified by more aggregated statistics that produce only broad averages.

In some areas there will be, over time, very direct uses of social statistics by individuals. The 'what is likely to work best' information that will help counsellors and case managers will also be available directly to individuals on interactive websites. These will provide information on the options that are available for making decisions in many of the social domains of life and the probabilities of success associated with different options, tailor-made to the individual's circumstances and needs.

Factors that are driving change

The main factor driving these changes has been increased demand for statistical information that cannot be easily met by further investments in traditional single-source source collection and processing methods. There are demands for:

- More granular, timely statistical information to support public policy – information that takes account of the many interacting determinants of quality of life, including the role played by combinations of public programs and policies over the course of the lives of individual Canadians.
- Information that reflects the diversity of the population and the inter-related dimensions of social wellbeing, taking account of the need for analysis at the level of small population subgroups.
- Predictive analytics that can directly support both individual citizens and the employees who administer social programs with information on which types of actions in the social domains of life are likely to work best in the current circumstances, based on what has worked best in the similar circumstances in the past.

This kind of information requires detailed descriptions of social programs and services. It requires detailed micro-level measures of the changes that are taking place over the lifecourses of individuals, with individuals seen in their differing social, economic, and geographic contexts.

That information cannot come only from traditional survey and census sources taken in isolation. The parallel driver of change has been the current availability of new data sources, new processing power, and new statistical methods that can be leveraged to fill these new demands in ways that are affordable, minimize response burden and protect privacy and confidentiality.

The 'Bottom-line Q&As' in Box 1 provide an overview of the rationale for the changes and how different audiences will benefit.

Reform is already well launched

In response to these changing demands – and to exploit the potential of new ways of doing business - the field in Statistics Canada with prime responsibility for the collection and release of health, labour and other social statistics launched a future directions exercise as part of the work on modernization that is taking place within Statistics Canada as whole.

Important first steps were taken in 2019, including a reorganization that separated responsibilities for dealing with the substantive content of the statistics from responsibilities for collecting the statistics. This was a necessary step in the shift to a multiple-source model.

COVID-19 accelerated development. Plans for improving responsiveness and flexibility in meeting the needs of users were put to immediate test with arrival of COVID-19 in 2020. The response of the statistical system in providing Canadians with the information needed to deal with COVID-19 has been astonishing, marked by innovations that would have been unbelievable only a few years ago. This was accomplished while learning how to deliver in a completely changed work environment with people working from home. These innovations not only met current demands, but also represented important steps in the desired longer-term directions.

The energy devoted to COVID-19 did, however, mean that some longer-term plans were put temporarily on hold, including the articulation and communication of a clear statement of where the social statistics system was headed. This document attempts to fill that gap.

Budget 2021 provided funds and reinforced the directions. The budget allocated funds to Statistics Canada to take concrete action along the lines initiated in 2019. Particular emphasis was placed on timely, disaggregated evidence to support policies and programs. It also stressed the importance of measuring the multiple dimensions of quality of life and of developing ways of using these measures to shape policy-making.

The role of a paper that shows where the system is heading

Many users and partners are not aware of the huge potential for improvement in social policies and in social decision-making that will be opened up by the shift that is underway.

In the absence of a shared understanding of future directions, both inside the statistical agency and among users and partners, it seems unlikely that the frenetic pace of change in 2020 can be sustained. Yet there is still much work to do. It is important to understand that, despite the considerable recent progress, we still do not have a system of national quality-of-life statistics in the ordinary sense of the word 'system'.

- Rather we still have a series of partly coordinated, but mainly siloed, approaches to measurement in different social domains such as education, health, justice, labour markets. There is need for greater integration with economic and environmental statistics – all of which impact on our quality of life.
- Most statistics, whether aggregated data or microdata, are still organized and accessed according to the source of that data – e.g., information from the Labour Force Survey, or from the Census, or from public use microdata based on the original source of that data.
- The tools that will hold a multi-source system together – consistent definitions that are independent of data sources, shared statements of vision and mission, coordinated outreach to users and partners, outputs that reflect the power of linking multiple, overlapping data sources – are only still partly in place.

1.2. The system of national quality-of-life statistics in summary

The chapter will end with a series of text boxes that compare the system that is being built once it is reasonably mature, perhaps after some 5 or 10 years of development, with a traditional single source system that has been the norm until recently.

Box 2 sets the context. The shift to a multi-source system in the statistical system is one dimension of a much larger shift in government towards better use of digital technology, openness, and data management. The Box shows:

- How the statistical system fits into that broader context in way that results in virtuous circles through much stronger data flows and linkages among the program and policy departments that use the statistics and produce administrative data.
- How trust and privacy protection lie at the heart of the evolution that is taking place.

Box 3 describes the outputs of the mature system, the statistics that will be produced.

- The most visible and short-term change will be the creation of an interactive website or portal that presents high-level quality-of-life indicators that reflect current policy priorities, as well as a series of more practical dashboards of indicators designed to support current policy agendas.
 - ▶ The site will provide direct access to Common Output Data Repository (CODR) tables (previously called CANSIM tables) along the lines of those on the present site, except that there will be many additional tables that are addressed to particular topics of interest – with the data drawn from many sources, including tables based on linked microdata files and including high quality synthetic variables.
 - ▶ The site will also facilitate access to a series of standard multiple-source microdata files that would be constructed by the statistical agency to shed light on analytic topics of general interest. These would be documented and anonymized.
- As well, a new output would be gradually introduced. These are databases that can create predictive analytic ‘what is likely to work best’ information that will support the decision-making of individuals and service providers as they make specific choices in the use of different types of social intervention.

Box 4 describes the inputs to the mature system, i.e., the sources of incoming data.

- There will be much heavier reliance on administrative data sources.
- Major changes will occur in the use of traditional survey and census data, particularly the use of hybrid vehicles that are designed from the outset to gather data from both survey and administrative sources.
- There will be detailed descriptions of social programs and services that can be used at the micro-level of particular interventions.
- Information will be captured about the economic, environmental, and social characteristics of the geographic areas in which the individual resides, and that information will be incorporated into individual micro-records.
- Eventually, feedback information from the new ‘what is likely to work best’ uses will provide information that will be especially useful in causal analysis of the effects of social programming.

Box 5 describes the processes that convert the inputs to outputs.

- These will make use of the sophisticated data processing and analytic tools that are now available, including techniques for data linkage that ensure privacy protection.
- A major addition will be an explicit conceptual framework to describe and integrate individual data items independent of their source. This will be an empirically-based micro-level framework, quite different from, but consistent with, the macro-level System of National Accounts that provides integration in the area of economic statistics, or other macro accounts in the environmental and social areas such as those related to health or education. As noted, a separate paper is being drafted that describes the conceptual framework.

Figure 1 is a logic model that shows how inputs, processes and outputs are inter-related and how they work together to produce the outcomes of the mature system – the uses to which the statistics will be put.

1.3. Managing constraints and risks

Chapters 2 to 5 will elaborate on the systematic description of the future system as set out in Section 1.2 and in the boxes at the end of this chapter.

The final Chapter 6 will address the main constraints and risks that will be faced during implementation and how they can be managed.

The largest resource constraint is the relatively small number of people in both Statistics Canada and in the user community with the skills needed to make full use of the potential of systems where data is drawn from multiple sources.

The largest risk would be a failure to maintain and reinforce the trust and confidence of the users and suppliers of statistics as we move to quite new ways of doing business.

1.4. Assumptions about an uncertain future and the need for flexible planning

In the real world, the statistical system will never be mature. It will always continue to evolve and will always be messier than some hypothetical model used in planning exercises. And, of course, the future cannot be predicted with any certainty, as the COVID experience has so clearly demonstrated. For this reason, an ongoing annual planning process is envisioned. In addition to updated statements about longer-term directions, the planning process will also include concrete action plans covering the following three years.

This paper is a draft of the long-term directions as they might appear in the 2022 version of such a strategic plan. It examines the system as it might exist about five or ten years into the future based on a continuation of existing reform initiatives. Given that time frame, this 2022 version assumes:

- **No large organizational or governance changes** will be made that would affect the operation of the system of social statistics. That could change in the future and subsequent plans will be adjusted accordingly.
- **Only currently available technologies and readily available data sources are referred to in this version**, although exploration and developmental work on possible future technologies and data sources will continue to be examined. However, it is likely that future versions will reflect major changes that are on the horizon.
 - ▶ For example, it is easy to imagine a future where a main source of social statistics will be portable personal devices (that combine health monitoring and GPS features) that would allow an individual to voluntarily provide anonymized information to the statistical system about their activities, contacts with others, and physical and mental wellbeing as the individual goes about the business of daily living.
 - ▶ Steps along the route of asking people to use devices to share information voluntarily have already been taken. However, there is too much uncertainty for a planning exercise to assume that this will become a mainline way of collecting data within the next several years.

Predictive analytic tools are, however, included. It might be thought that the use of predictive analytics to provide ‘what is likely to work best’ data belongs to future versions of the plan given the current lack of the use of this tool in official statistics. However, it is assumed that the statistical system will be given the mandate to work in partnership with service providers on experimental work over the coming years to create portals that will allow the individual employees who provide the services to have real-time access to information on:

- The interventions/choices that are available (content coming from the service providers).
- Which interventions/choices are likely to work best in particular cases (content coming from the statistical system).
- The same kind of ‘what is likely to work best’ information could also be provided directly to individual Canadians to help them in their decision-making.

It is included in the 2022 plan because:

- The payoffs from early action would be high on many fronts. The eventual result would be a giant step towards social policies that are truly based on evidence. As well, many more people, including individual citizens, would gain directly from such applications, increasing overall trust in the system.
- There are risks of moving in this direction, but they are manageable. Implementation would be gradual, based on pilots. Accordingly, failure on this front would not have a large, negative impact on the rest of system.

What works initiatives refer to the sharing of evidence about the effectiveness of programs, often based on evaluations and random control trials. The UK, for example, has a What Works Network covering different programs areas. The ‘**what is likely to work best**’ initiative in this paper is quite different. It describes an approach that embeds predictive analytics right into the operational decisions made on individual cases, at the time those decisions are being made. See Section 3.4 for further explanation.

Box 2.

The context: digital government, openness and managing data as a resource

The growing use of administrative data for statistical purposes is one dimension of current efforts in many countries towards open government, increased use of digital technology and strengthened data management techniques.

- Data is being seen as a resource to be carefully and transparently managed, with data holdings made available to the public, consistent with privacy and confidentiality.
- A theme is that each individual item of data collected, each sliver of information in the system, should be fully used including in multiple applications in and out of government.

Progress has been slow until recently

The potential power of treating information as a shared resource has been discussed for decades, but progress has been slow.

- **Cultural factors.** Policy and programs have remained in traditional silos that create cultures and rules that are hostile to data sharing.
- **Some preconditions missing.** It is difficult to create the preconditions for effective data sharing and the management of data as a resource, for example:
 - ▶ Standard ways must be found to catalogue data holdings using consistent concepts and definitions and to provide user-friendly access to that data.
 - ▶ IT and data management solutions must be found for allowing the same data items to be used for multiple purposes, without infringing on privacy.
 - ▶ The digital literacy skills of those who collect, manipulate, and use the data must increase.
 - ▶ Trust must be built among the individuals and organizations that provide the data initially, including confidence there will be no negative repercussion for privacy and confidentiality.

Box 2.**The context: digital government, openness and managing data as a resource**

- **Modest payoffs.** Early efforts were often concentrated on open government themes – providing the public with access to data files that were designed to support existing internal operations:
 - ▶ One goal was greater transparency in government and a resulting increase in public oversight. Progress here has been slower than hoped for, partly because of the privacy and confidentiality restrictions placed on access to data – and because the data in question was not designed to support public accountability applications.
 - ▶ There have been some successes in the related goal of providing users with data that they could use in other applications. However, again, the resulting benefits have been far less than many had anticipated.

In reality, data sets designed for internal operational purposes are often not directly useful for other purposes and there has been a lack of capacity to make the adjustments that would be needed to increase their relevance.

Current initiatives address these weaknesses

Lessons have been learned. Recent initiatives do address the preconditions. They focus on internal data management – the flow of data within the whole system – as well as providing public access to existing data. In Canada, these initiatives include:

- The recent Data Strategy for the Federal Public Service.
- The creation of a Minister for Digital Governance and departmental Chief Data Officers.
- Open government initiatives that include all orders of government, including municipalities, with links to similar initiatives in other countries.

The role of the Statistical Office

Statistics Canada will play a pivotal role in next phases of digital government and open government reform.

- **In supporting the technical pre-conditions.** Statistics Canada has world-leading expertise to help build the capacity to:
 - ▶ Develop the common concepts and classification systems that describe, catalogue, and allow access to data.
 - ▶ Develop the techniques that allow data items from different sources to be used in combination for a range of purposes, while still protecting privacy.
 - ▶ Support increased digital literacy.
 - ▶ Using the data from multiple sources to make major improvements in the statistics available to support improved quality of life as described in this paper.
- **In building trust.** Statistics Canada is uniquely situated to provide a safe and trusted place where data from different sources can be combined using techniques and processes that avoid concerns about privacy and confidentiality. This trust has been strongly reinforced by recent action in Statistics Canada:
 - ▶ A world-leading framework has been developed (the Necessity and Proportionality Framework) for balancing use, privacy, and response burden.
 - ▶ Sophisticated tools now exist to ensure privacy protection, using techniques that split out individual identifiers from the data that describe those individuals. No master file of identifiable records is created.

A virtuous circle

The role played by Statistics Canada supports the sharing of data throughout the system. Increasing amounts of data from administrative files will flow into the statistical system along with data from survey respondents. This will result in more useful statistics both for the bodies that provided the administrative data and for the respondents to surveys. That, in turn will create a climate that will encourage even greater data sharing as the system evolves.

Box 3.

Comparison of statistical outputs in a multiple-source and in a traditional single-source model

Audience for statistical outputs

In the traditional single-source model, there are two main audiences for social statistics:

- A general audience with an interest in trends in the social domains of life where the goal was to provide a basic socio-economic and demographic profile of Canadians. These data provide a consistent common denominator for applications ranging from indicators of social wellbeing, to measures of the extent of social problems, and to the marketing activities of firms.
- Analysts and researchers dealing directly or indirectly with social policy and the determinants of social problems.

In a multi-source, quality-of-life model, there will be the same two audiences, with the following differences and additions:

- The policy applications will be more finely tuned to practical program agendas. They will be focussed on the effects of programs, and combinations of programs, on the lives of citizens in diverse situations and with diverse needs. The focus expands from trends to also include equality and sustainability considerations.
- A new audience will be served consisting of persons and service providers who will be provided with information on the probabilities of 'what is likely to work best' in making decisions in the main domains of social life.

Availability of tabulated outputs

- In the traditional model, users once had to rely on paper publications reporting on survey results, often produced with considerable time delays. More recently, users have much more flexible and timely access to data tables using Statistics Canada's website. However, the tables on that site are still organized by individual survey and census sources.

In the multi-source, quality-of-life model, there would be similar access via a single interactive web site. However, the site would also contain tabulations based on linked data files, would contain synthetic data, and would be increasingly structured around data content rather than data source.

Box 3.**Comparison of statistical outputs in a multiple-source and in a traditional single-source model****Availability of microdata**

- In the traditional model, there are a range of methods of access to single-source microdata files – with varying restrictions to access related to privacy protection. Microdata is mainly seen as an additional product that users must pay to receive.
- In the multi-source model, access restrictions related to privacy protection will be at least as strong. However, microdata will be seen as a core, free product of the system, with the statistical system itself providing standard anonymized micro databases drawn from different sources, including synthetic data of high quality. Many of these standard data bases would contain lifecourse and longitudinal data.

Indicators and dashboards

- In the traditional model, at least in Canada, systematic production of indicators and dashboards is mainly seen as the responsibility of users, not of the statistical agency itself.
- In the multi-source model, quality-of-life indicators and policy-relevant dashboards would be an integral feature of the website that provides access to aggregated and microdata.

Probability of success information

- In the traditional model this output did not exist.
- In the multi-source model,
 - ▶ Service providers, including those who refer people to particular kinds of services, would have access to a site (created in partnership with the service agency in question) that provides them with tailor-made information on which interventions would likely work best for a particular client, along with information about the ways of accessing the intervention in question.
 - ▶ Individuals seeking information – such as information related to job search or training options – would have access to similar interactive sites (again maintained in partnerships with responsible subject-matter agencies) that would provide information on the options that were open – and likely the success of each.

Box 4.

Comparison of inputs in a multiple source model with those in a traditional model

Vastly more input data would be collected:

- As a result of the growing demand for disaggregated, responsive statistics.
- As a result of the availability of much more detailed, relevant data from administrative and other sources.
- As a result of the computing power to manage vastly more data – without creating response burden and while maintaining privacy and confidentiality.

A change in the mix of input sources

- **Censuses and traditional surveys** such as the Labour Force Survey are likely to collect either about as much data as they have in the past, or possibly less as a result of response burden issues and the availability of alternative data from administrative sources and from the new, more responsive surveys described below.
- **More flexible surveys.** A suite of much more flexible survey instruments would be in place whose role to respond quickly to current demands and to provide information on topics that cannot be well covered by other sources such as administrative data.
- **Administrative data** sources would play a much larger role, particularly in the coming five to ten years. Statistics Canada will, partly in consequence, play a more active role in the overall management of data throughout the federal system and nationally.
- **Hybrid vehicles** that are designed from the outset to gather data from both surveys and administrative sources are already in place and could well become more common in the future, perhaps allowing designers to draw on the strength of both sources in ways that have not yet been fully explored.

Non-traditional sources such as social media or personal monitoring devices will almost certainly play a larger role in the longer-term, but not likely in the next 5 or so years since these sources are not representative or fully developed. It is difficult to predict when these new technologies will start playing a large role, but their effects are potentially transformative. Experimental and developmental activities are likely to dominate in the coming several years.

- **Feedback loops** from individuals participating in ‘What is likely to work best’ initiatives will play a central role in a mature system, providing a major source of data that can explore causal relationships in policy analysis. This will take years to fully develop, however.

Box 4.**Comparison of inputs in a multiple source model with those in a traditional model****Linked data would become the new norm**

The defining feature of a multiple-source system is its capacity to draw on data from multiple sources including combinations of data that were not on any one of the original data sources.

- Many of the data sets that are used in these applications contain overlapping information which allows the creation of synthetic data and provides quality controls.

Stronger emphasis on longitudinal data

More of the input data will be longitudinal (following the same individual over time), arising mainly from administrative data sources. The traditional model relies on longitudinal surveys.

Inclusion of data about individuals based on their geographic location

Data about geographic areas would be calculated by aggregating the characteristics of the individuals who live and work there. In addition, economic and environmental information about those areas would be obtained from exogenous sources.

- This would allow analysis of data from both sources to be carried out at the level of the geographic area, a form of analysis that is already in place and steadily improving.
- More important here, it would allow economic and environmental data to be imputed to individual records based on where the individual lived, enabling a potentially powerful new form of analysis that sets individuals in their temporal and spatial environment.

Consistent data about social programs and services

- In a mature system, some years in the future, highly disaggregated data about specific, micro-level program and service interventions would be collected and used in conjunction with the individual data collected about the participants in those programs. This would allow an examination of the outcomes of particular social programs and of the combined effects of the programs of different orders of government.
- This contrasts with traditional macro approaches that examine systems as whole, such as the income security, health, education, social services, and justice systems. Information is collected about the volumes of people served and costs – but, in the case of services, without a direct link between the services and individuals who were served. This capacity for linked micro analysis already exists for income transfer programs.

Box 5.**Processing and analysis in a multiple-source, quality-of-life microdata model****Conceptual Framework: standard definitions at the micro-level**

When input data are drawn from many sources, there must be an independent set of consistent concepts and definitions that describe the individual data items that are being measured.

In the social area, standard definitions are needed to describe, at a fine level of detail, the characteristics of:

- **Residences:** where people live, including the type and quality of the residences and their geographic locations. (Other forms of real estate are also be described, but in social statistics the main focus is on dwellings.)
- **Peoples' characteristics:** including basic demographic and cultural identifiers, and snapshots of status at different points in time, including skills status, employment status, health status, income and much more. These snapshots are also known as stocks.
- **Peoples' activities:** how they spend time in a day, including a full range of social and economic activities and economic transactions. These are also known as flows.
- **Organizations,** including businesses which are described using concepts derived from the System of National Accounts and social programs and social services which are based on an input-process-output model.
- **Geographic areas,** including environmental characteristics.

Standard ways of showing relationships

The concepts used must allow analysis of the important relationships among the topics above. Examples include:

- People living in the same residence, including household relationships.
- The activities (known as flows) that link people with each other and with organizations (including both economic transactions and flows of services and information).
- Geographic proximity of residences, organizations, and commuting patterns.
- Socio-economic characteristics of residents in different geographic spaces.
- Social capital defined in terms of network relationships.

Standard ways of showing and describing changes over time

There are four main ways of describing change over time:

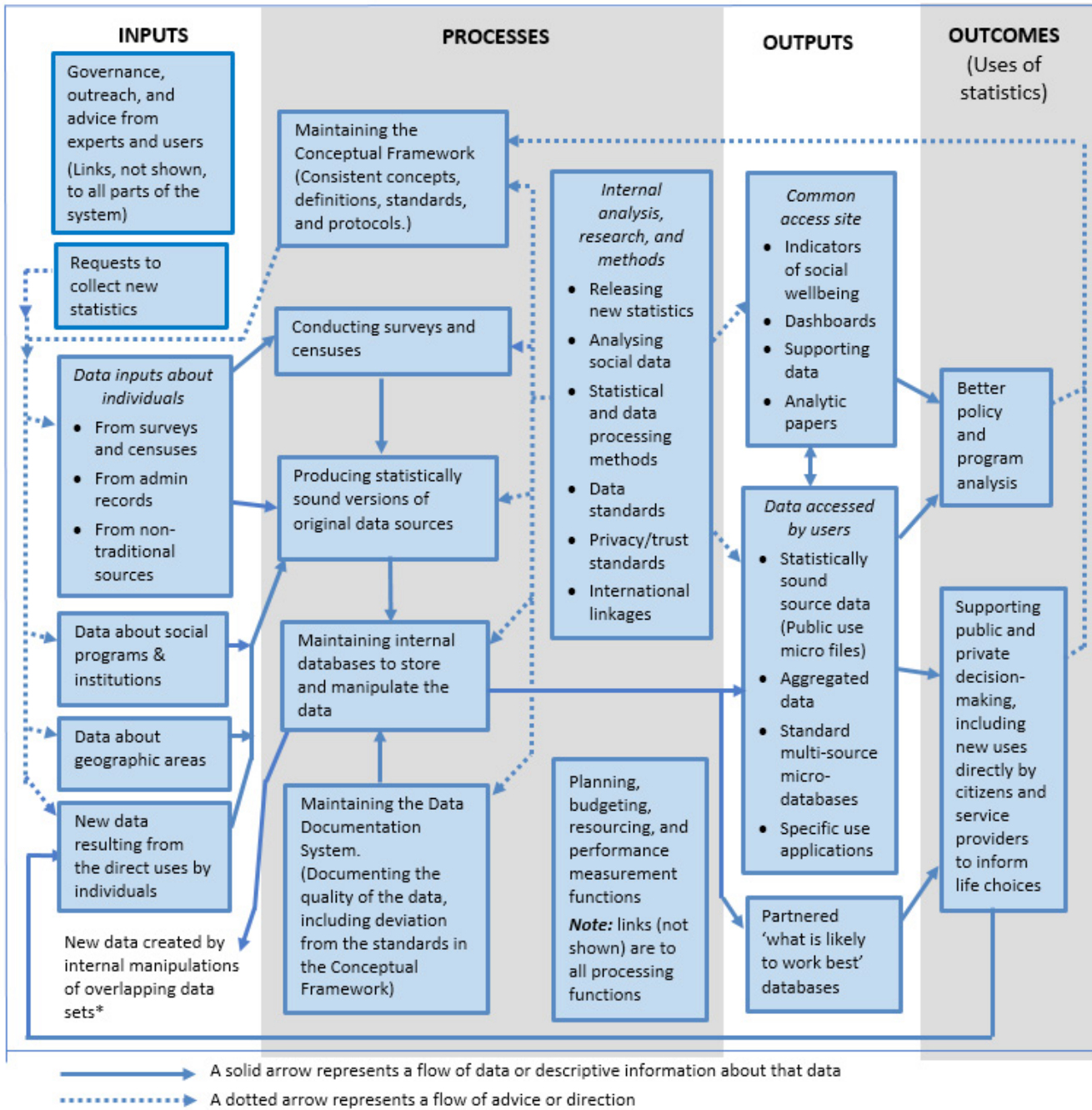
- **Longitudinal or recall data,** where information for the same individual is gathered at different points in time, or where the individual is asked to recall information about his or her status or activities at earlier periods.
- **Stock and flow data,** where stocks at a later period of time equal the stocks at an earlier time, plus or minus the net flows that took place between those periods.
- **Lifecourse concepts,** where the point-in-time flows are coded as events. This, in turn, allows analysis of the transitions and trajectories that make up the main domains of life, such as life in the family, in work, in learning, or in caregiving and care-receiving.
- **Hypothetical changes** which are the changes that might have occurred in aggregate systems based on 'what if' analysis using microsimulation or other analytic tools. When applied to the micro-level probabilities of success of individual decisions, 'what is likely to work best' predictive analytic tools are used.

Box 5.**Processing and analysis in a multiple-source, quality-of-life microdata model****Combining input data**

Incoming data from the multiple sources identified in Box 4 are combined to produce the new responsive, quality-of-life statistics using new principles and new tools.

- **The principles that guide the choice of data sources.** The principles used in decisions on which sources and combinations to use in a particular application is known as the **Necessity and Proportionality Framework**. It takes into account the importance to the nation of the statistical outputs to be produced, the need to minimize response burden and to protect privacy and confidentiality.
- **Tools to combine data, ensure its quality and protect privacy.** A rich set of technical tools is now available to link data from different sources when required, to create synthetic variables, to ensure quality, and to protect privacy, including:
 - ▶ Traditional statistical sampling tools.
 - ▶ Tools to correct errors and impute missing data.
 - ▶ Registry tools developed in Scandinavia and elsewhere.
 - ▶ Linkage tools that do not disclose personal identifiers and tools for ensuring that statistical outputs are anonymous.
 - ▶ Microsimulation and other tools to create synthetic data.
 - ▶ More generally, tools that make use of the multiple accounting devices (double entry bookkeeping) built into the standard micro-level concepts and definitions.
- Tools of **analysis**. A wide range of familiar and new analytic tools are used to analyse the rich microdata, including:
 - ▶ Static and dynamic microsimulation.
 - ▶ Geomatic tools for spatial analysis.
 - ▶ Artificial intelligence and machine learning tools.
 - ▶ Lifecourse and longitudinal analytic tools.
 - ▶ Mixed methods, including new opportunities for qualitative research.
 - ▶ Interactive analytic tools, including visualization tools.

Figure 1. A logic model of a mature multi-source system of national social statistics



* Note that the input "New data created by internal manipulation" is technically part of the processing function.

Note: The chart shows the logic model for the entire system of national quality-of-life statistics, not only those parts that are within the responsibility of the field responsible for social, health and labour statistics.

2. Outcomes: the purpose and scope of the evolving system of quality-of-life statistics

Section 2.1 discusses the first of the two outcomes identified in the logic chart in Figure 1 – supporting policies and programs.

Section 2.2 provides a similar description of the objective of supporting decision-making.

Section 2.3 proposes a possible statement of vision and mission.

2.1. Supporting policies and programs

Providing evidence to support policies and programs has always been the main goal of national social statistics. Canada has always had a reputation for world-leading excellence in statistics. When it comes to its main traditional goal of supporting social policies and programs, the agency can build on strength.

However, there is also a long way to go to fully respond to the policy demands described in Chapter 1. This can be seen in the sidebar which provides an expanded version of the outcome that is summarized in Figure 1. While this wording is illustrative only, it captures much recent thinking in Canada and internationally.

The extended statement suggests that the mature system should be designed to improve on existing arrangements in the following areas:

- **Monitoring.** No capacity has been set up to monitor the various dimensions of social wellbeing in a way that is current, comparative, and authoritative. Information about inequality and sustainability risks is not as developed as are indicators of trends and levels.
- **Cause and effect.** The system is improving but is still weak in measuring the interplay among the multiple factors that affect wellbeing in different population groups – such as combinations of poor jobs, health, skills, living alone and housing.

An extended statement of the outcome: supporting policies and programs

The national system of quality-of-life statistics will provide readily accessible, disaggregated and timely information on:

- Trends, levels, inequalities, and sustainability risks related to wellbeing and the quality of life, both overall and for different population groups – and for the various ways in which social policy responses are organized such as health, income, or learning.
- The causes of these levels, trends, inequalities and sustainability risks and the role of existing government policies, and of alternatives to existing arrangements, in improving wellbeing.

Implications for the design of the multi-source system

- A monitoring capacity that is timely and responsive, that covers all areas of social wellbeing, and that provides indicators of trends, levels, inequalities, and sustainability.
- A capacity for causal analysis based on detailed descriptors of individuals using microanalytic, longitudinal tools.
- Inclusion of consistent, detailed descriptors of the outputs of social programs and social services and of their beneficiaries.

- **Effects of programs.** Much analysis of the social effects of policies is weak since it ignores the obvious, namely that there is seldom a one-to-one relationship between a single government program (or the services provided by social institutions such as schools, hospitals, and community organizations) on an individual's wellbeing.
 - ▶ There is good data on the direct effects of programs that provide income.
 - ▶ There is inconsistent data on the effects of programs that provide services.
 - ▶ In consequence there is virtually no information on the combined effects of income and service programs from all levels of government on the lives of individuals.

The consequences for the design of the new multi-source system are shown in the second sidebar above.

2.2. Supporting public and private decision-making

The second purpose or outcome of the statistical system is to support better decisions by a wide range of actors, public and private.

Public decision-making

The improvement in public decision-making will occur because of:

- The policy and program uses discussed above, including the use of stronger 'what if scenarios'.
- The provision of 'what is likely to work best' information directly to those who deliver services within the public system. This is discussed below.

Private decision-making: by organizations

Currently, most direct uses by organizations relate to marketing. Demographic, socio-economic and health data often support the marketing functions of firms and of institutions that provide health, education, and other social services. National social statistics provide a kind of common denominator that describes the needs and characteristics of potential recipients of the products and services of those organizations. Projections of key demographic variables are particularly useful in planning initiatives in both public and private organizations.

- Often this type of information is provided by private market research firms who combine data from the national statistical system with data obtained from other sources.
- The shift to a multi-source model in the national system will provide even better 'denominator' statistics because of their greater granularity.

Private decision-making: by academics and researchers

Much academic research is, of course, directly or indirectly, related to the policy and program uses above. As well, a wide range of academic social research, especially research that examines the longitudinal dimensions of life, will benefit from the increased granularity and integration of the statistics produced by the multi-source system. The addition of lifecourse concepts to the micro-level conceptual framework will facilitate mixed qualitative and quantitative analysis. Many new tools of access and analysis are being developed to support external researchers (in the research data centres and through a range of virtual tools including the Data Analytics as a Service Platform initiative) and authorized researchers in other departments (e.g., Virtual Data Labs).

Box 6**Supporting decision-making by citizens and service providers****The rhetoric**

A familiar, long-standing theme in the social policy literature is the importance of moving away from systems that are designed to make life easier for those who deliver highly siloed programs and services and to move towards systems that are centred on the needs and circumstances of those who use the programs and services.

- In **health**, there have been decades-long calls for shift from disease-centred services to patient-centred services with an emphasis on population health and wellness. While progress has been limited, a recent [C. D. Howe article](#) shows how the massive COVID-19 shift to virtual care provides an opportunity to embrace a more patient-centric and cost-effective healthcare system.
- **Educators** wish they could move from one-size-fits all classrooms to student-centred learning. Again, the possibility exists that experience with COVID-driven virtual learning can provide an opportunity to make real progress in the years ahead.
- **Case managers** and counsellors aspire to coordinate flexible packages of social, health and learning interventions tailored to individual needs – even if the components of such packages were not designed to be compatible.
- **Police and social services** are asked to coordinate their activities at the community level but often without the resources to do so effectively.
- There have been endless calls to make the existing labyrinth of complex and overlapping **income security programs** more accessible – with the system itself taking the lead in putting together the most advantageous packages of benefits for recipients, rather than leaving potential recipients with the responsibility for determining eligibility.

The reality

Despite much effort and some successes – and the silver lining opportunities suggested by the pandemic – social programs and service providers still operate largely in separate silos organized in ways that are designed to support accountability and efficiency within each silo.

- The obstacles to moving forward in the area of income support programming are largely jurisdictional and cultural.
- In the service area, however, a fundamental obstacle in moving to a client-centred model is the lack of adequate information on which kinds and mixes of services and supports work best in improving quality-of-life at the level of individuals.

In the absence of such hard evidence, it is not possible to develop cross-cutting, client-centred delivery systems that are accountable, efficient, and effective.

Fundamental improvements are becoming possible

New data sources and predictive analytic techniques are opening the possibility of making real progress in providing the statistical evidence that is needed to support the shift to client-centred systems of health, education, and social services. It is now possible to foresee the day when tailor-made statistical information about ‘what is likely to work best’ can be provided directly to individuals and service providers:

- As they make choices in the social domains of life and at the time when decisions are being made.
- Based on probabilities of success calculated from what worked best in similar circumstances in the past.

Box 6

Supporting decision-making by citizens and service providers

In some areas, such as training and labour market programming, the needed multi-source microdata from has been available since the late 1990s although jurisdictional siloes have prevented its use except in evaluations.

- In other service areas, it will take years to develop the needed micro multi-source information in the detail required to calculate reasonable estimates of probable success.
- In all cases, success will depend on collecting the needed microdata and on launching demonstrations and pilots over the next several years to develop the tools for using this detailed information in a range of ‘what is likely to work best’ applications.

With high potential payoffs

The payoffs from such a gradual, experimental process will eventually be high:

- To the individuals themselves.
- To those who design and administer social programs who will also have access to micro-level information on what is likely to work best.
- To society, as the result of the improved operation of health, labour market, educational and other social systems, including reductions in the cost of government remedial programs.
- To the statistical system itself since more people will have direct gains from its operation with increased trust and support and with powerful new data feedback loops that will continually improve our capacity to learn what is working best.

Private decision-making by individuals

Currently, most uses by individual Canadians are indirect. Some individual Canadians may consult the Statistics Canada website to find background information related to personal decisions. However, such direct uses are almost certainly small when compared with indirect benefits from the policy-related statistics described above. These uses will result in better policies and programs, and a more informed media. This, in turn, will have positive important effects in improving people’s lives and in the operation of institutions.

Direct uses by individuals are likely to increase. The importance of directly supporting individuals has been well recognised. For example, as seen in the sidebar, Statistics Canada’s current Modernisation Initiative highlights the importance of supporting quality decision-making by citizens.

However, the reality is that a statistical system based on single-source statistics has, in practice, provided little information that is of direct and immediate use in helping individuals make decisions.

In a multi-source system, on the other hand, a wealth of information is available to support a wide range of individual decisions using predictive analytic tools. Although this will require much development work, over the long-term, these uses could become a central outcome of the system of social statistics.

As elaborated in Box 6, this shift to serving Canadians directly responds to a long-standing aspiration in most areas of social policy, but one where little progress has been possible because of a lack of supporting statistical data.

New emphasis on serving individuals

Two of goals of modernization are specifically cast in terms of better serving individual Canadians:

- Produce more timely and responsive statistics – ensure Canadians have the data they need when they need them.
- Develop and release more granular statistics to ensure Canadians have the detailed information they need to make the best possible decisions.

The largest long-run payoffs will be from ‘what is likely to work best’ data. The shift to a multi-source system will make it possible to provide individual citizens and service providers with real-time information about which type of social or health intervention is likely to work best, based on what has worked best in similar circumstances in the past.

2.3. A statement of vision and mission – and a new title

Developing a statement about the purposes and scope for the multi-source system of quality-of-life statistics could help coordinate implementation activities. It could be useful in building understanding among users and partners.

A draft of such a statement of vision and mission is shown in Box 7. The statement is illustrative only and intended for planning purposes. A final version would be based on consultations and expressed with greater elegance.

Below the statement, the box provides an explanation of the factors that were taken into account in its drafting, including thoughts on the scope of Statistics Canada’s mandate in the social statistics area and the rationale for replacing the phrase *social statistics* with *quality-of-life statistics*.

Box 7.

Draft statement of vision and mission for the National System of Quality-of-Life Statistics

Vision

To provide the statistical evidence that people, organizations, and governments need in order to work together in improving the quality of life in Canada.

Mission

The national system of quality-of-life statistics will support the development and evaluation of public social policies and programs by providing readily accessible and timely information on:

- Trends, levels, inequalities, and sustainability risks related to wellbeing and the quality-of-life, both overall and for different population groups — and for the various ways in which social policy responses are organized such as health, income, or learning.
- The causes of these levels, trends, inequalities and sustainability risks and the role of existing government policies, and of alternatives to existing arrangements, in improving quality of life.

It will support the decision-making of governments, businesses, social service providers and individual citizens by providing timely and accessible information on the social characteristics and activities of representative Canadians, and how these are shaped by social programs and interventions.

Considerations that shaped the drafting of the statement of vision and mission

Consistency with the larger Statistics Canada mission

The two outcomes of the proposed multi-source model are the same as the first objective of Statistics Canada as seen in the statement of mandate and objectives found on the agency's web site, namely:

- **To provide statistical information and analysis about Canada's economic and social structure to:**
 - ▶ Develop and evaluate public policies and programs.
 - ▶ Improve public and private decision making for the benefit of all Canadians.

Considerations in wording a Vision and Mission Statement

- Wording that is consistent with a gradual approach to providing direct 'what is likely to work best' statistics.
- Wording that highlights the importance of partnership and inter-sectoral horizontality in the statistical system and in the uses of statistics.
- Wording that does not suggest that the national system has a monopoly on social statistics.

Scope and limits

Related to this last bullet, it may be useful for the mission statement to signal the main roles that the national system will play, namely:

- Statistics related to nationally representative individuals that describe their quality-of-life and the factors that affect that quality-of-life.

The federally funded, national system should have its focus on statistics that apply across the country and should not arbitrarily favour any one province or locality.

- Statistics that describe social programs and social services.

The wording of the mission statement should, however, not preclude the following exceptions:

- **Demonstration projects or pilot studies** which are not nationally representative but will usually have the potential to be scaled up to the national level if resources and demand is there.
- In some cases, the needed representative data is simply not available and **non-representative proxies** must used, as in crowd-sourcing initiatives.
- **Other exceptions.** There will always be cases where only the national statistical system has the expertise and capacity to undertake important projects that are not intended to be national in scope or to address social issues that are not based on data about individuals. In these cases, decisions can be taken at a senior level to make exceptions and undertake such projects on a cost-recovery basis.

Considerations that shaped the drafting of the statement of vision and mission

Title – What we call ourselves

The use of the phrase ‘**quality-of-life**’ statistics, instead of the more familiar ‘**social statistics**’ does not signal any change in the scope of what is included in terms of subject-matter areas. Rather it conforms to language used in Budget 2021 and is intended to signal the new focus on a micro-level understanding of the determinants of the quality of life at the level of the individual.

The change might also resolve some current ambiguities in the use of the term ‘social’:

- The term ‘social statistics’ sometimes refers to all the domains in society that are not economic or environmental statistics – even though, logically, ‘society’ is the overall heading with the economy being only one dimension of society.
- Moreover, even within the traditional area of social statistics, the word ‘social’ is sometimes used only to cover residual elements. For example, the title of the Statistics Canada organization, the Social, Health and Labour Statistics Field, logically implies that health and labour statistics are not social statistics – even though the real meaning is ‘social statistics including those related to health and the labour market’.
- Additional ambiguity arises with the commonly used term ‘social and health statistics’ which sometimes means ‘social statistics, including population health statistics’ but which sometimes has a broader scope and refers to ‘social statistics including those related to population health and to the operation of the health care system which centres on diseases’.

3. Outputs: statistical products

This chapter deals with outputs – the statistical products of the proposed national system of quality-of-life statistics – that were identified in Figure 1.

- Section 3.1 describes the proposed interactive web site or portal, including indicators and dashboards. It discusses progress to date and the need for additional action.
- Section 3.2 describes progress and next steps related to access to aggregated data.
- Section 3.3 describes progress and further work related to access to microdata.
- Section 3.4 deals with the partnered databases that will need to be constructed to produce ‘what is likely to work best’ information tailored to individual decision-making.

3.1. Progress in building the interactive web site

Box 3 described the Quality-of-Life information on the Statistics Canada website:

- Where all the new social statistics will be available on the day of release.
- Which contained high level quality-of-life indicators and flexible dashboards of supporting indicators that would be cast at a level that is relevant to practical policy agendas.
- That would provide ways of drilling down to detailed aggregated and micro data that will be described in Sections 3.2 and 3.3.

Rationale for releasing on a common website. Continuing the present practice of using a common website for all releases of the latest social statistics, such as the latest Labour Force Survey data or the latest census tables, has several benefits:

- It allows those who draft the press releases and associated material to situate the current statistics in the context of other social changes that are taking place – including how the current data are related to the more general story being told by the social indicators.
- It allows the media and other readers easy access to related indicators, dashboards, and supporting tabular and micro data.
- Users accessing any statistics will always get the most recent figures.
- More generally, it reinforces the shift from the traditional model to the more responsive, multi-source model.

Rationale for including systematic social indicators and policy-relevant dashboards. Box 8 describes the continued interest in high level indicators that monitor progress, and point to possible problems, in achieving a high quality of life. In part, such systems are valued because they hold promise of compensating for the highly siloed nature of most current social statistics and of the policies they support. However, Box 8 indicates that indicators are not much used in practical policy-making and proposes a better approach based on the micro-level multi-source model.

Box 8.

A different role for social indicators and policy dashboards

Interest in social indicators remains high

For many decades, there have been efforts to compensate for the siloed nature of social statistics by developing social indicators that provide a multi-dimensional measure the quality-of-life in a society.

The 2009 Stiglitz-Sen-Fitoussi Report gave much impetus to the social indicator movement, including the development of the OECD's Better Life Index. Currently, some governments, including Canada, are exploring ways of casting their planning and budgeting process in a quality-of-life framework.

Content of most traditional social indicator initiatives

Typically, social indicator initiatives consist of a manageable number of indicators that attempt to summarize key changes in the different areas that are thought to be most related to quality-of-life – health, learning, income, inequality, sustainability, perceptions of happiness.

- Arranged in several tiers, with the most summary indicators in the top tier, with further breakouts at a second level and even lower-level tiers.
- Since there is no consensus on what constitutes wellbeing, much creativity and consultation goes into their creation.

Their weaknesses

Despite their theoretical appeal, traditional social indicators have had relatively little impact on social policy-making for the following reasons:

- The lack of consensus on what constitutes wellbeing.
- A lack of symmetry between the content of the indicators and the agendas under which policy discussions are framed. Policy decisions are seldom taken at the high level of generality found in typical social indicator tiers.
- Traditional indicators do not provide quick access to the supporting data to facilitate follow-up analysis.
- The often-long time delays before indicators are released. They are often conceived as annual paper reports rather than interactive, current web sites.

Box 8.**A different role for social indicators and policy dashboards****The proposed solution**

The indicators on the interactive web site would:

- Provide context for, and a quick gateway to, the much lower-level dashboards of indicators that are of actual use in policy contexts.
- They would **not** be cast as an official top-down way of defining the components of wellbeing or the constituents of a high quality-of-life. Rather they would be seen only as those that are chosen by the editors of the web page to be those of greatest current interest or that corresponded with clear policy goals such as meeting poverty reduction targets.
- Interested users would be encouraged to develop high-level sets of indicators that mirrored their own interests.
- These would not be seen as being in competition with the ‘correct’ official approach, but rather as a legitimate reflection of the different priorities seen by different groups.
- For example, media stories might compare the indicators sets chosen by business groups interested in efficiency with those chosen by groups interested in equality, or sustainability, or disadvantaged communities.

A core of common indicators

At any one time, a core set of indicators would be highlighted to cover topics of highest current interest from a pan-Canadian perspective.

- If, as suggested in Budget 2021, the Government of Canada were to adopt a standard set of quality-of-life indicators as part of the budget process, these would be highlighted.
- A Canadian version of the OECD’s Better Life indicators would be an alternative.
- The website would **not** publish a single overall measure of wellbeing. The emphasis is on moving downward to more policy-relevant dashboards, not upward into theoretical constructs.

Dashboards for policy uses

Lower-level dashboards of indicators would be created. Based on consultations, these would focus on a variety of current policy agendas – where decisions are actually being made.

- The different dashboards would not need to have the same format and they could overlap. For example, different dashboards could be created to support groups working on different aspects of a same policy issue, for example:
 - ▶ Dashboards designed for those working on an issue in a particular geographic area.
 - ▶ Dashboards that focussed on effects on different population groups.
 - ▶ More complex dashboards could be referred to as Hubs, as is the case on the present Statistics Canada website.
- These would be designed to provide a small handful of indicators to support analysis of the policy issue at hand, including interactive graphics where this makes sense.
- The site itself would provide quick access to a large number of standard dashboards that are in demand.
- As well, user would be invited to create their own dashboards. When they signed in the next time, users would be asked if they wished to see an updated version of those retained dashboards.

Why such flexibility in indicators and dashboards is possible

The proposed system can be flexible because consistent, integrated concepts are built into the underlying microdata. For example, when there is need to consider a change to a high-level indicator or a dashboard, it can simply be replaced by an indicator that comes complete with its own historic time series -- in a way that allows ready comparison of the old and replacement indicators.

Progress to date and the need for additional action

Many of the desired features of the proposed interactive web site already exist on the current Statistics Canada web site, which is becoming increasingly powerful.

- **Current releases of new data** are found in the Daily section of the web site. There is often useful commentary about the wider implications of the new information, although the focus remains on the information from the particular survey or other source of the statistics. Further work is needed to develop a capacity for releasing new data in the context of a broader set of quality-of-life indicators.
- **Indicators.** The list of key indicators on the Daily page of the website provides easy access to important single indicators as do the key indicators shown on other pages such as the hub related to gender, diversity, and inclusion statistics. However, future work is needed to develop a systematic set of quality-of-life indicators.
 - ▶ Development work, mandated by Ministerial mandate letters and referred to in Budget 2021, has been underway with the Department of Finance to explore the creation of such indicators in the context of policy formulation and budgeting.
 - ▶ No decision has yet been made on how, and if, to proceed. However, the collaboration that has taken place means that if such indicators are adopted, they will be measurable and can be included in the proposed interactive web site.
 - ▶ **Dashboards.** There are already a handful of dashboards on the current website that have been carefully designed to shed light on specific policy and program agendas along the lines proposed in Box 8. These include several dashboards associated with the COVID response and the Dimensions of Poverty Hub, with its 12 indicators. Future work will involve consultations that are designed to greatly increase the number of such dashboards.
 - ▶ **Interactivity and visualizations.** The site provides increasingly powerful tools for visualizing and mapping the data, including interactive tools, and more dynamic charts and thematic mapping tools are planned. In other words, the present rate of progress is excellent. Next steps should include a provision for users to retrieve their own tailor-made dashboards.

3.2. Access to tabulated data: progress to date and next steps

Recent improvements on the Statistics Canada website, with the shift from CANSIM to CODR were major steps in the direction of flexible access to aggregated data. In addition, a plan should be developed in consultation with users to add new tables that:

- Are based on data drawn from linked microdata files and including synthetic variables. For example, Statistics Canada's researchers produce many valuable one-time reports based on a variety of current topics using linked data. Some contain tables that would be of much interest if they were updated regularly and included on the web site.
- Complement existing tables, which are based on sources such as a particular survey, with new tables that include tabulated data related to the same subject matter but drawn from different sources, such as employment and related labour market data drawn from different surveys and administrative files. This would be one dimension of the work in developing dashboards discussed above.
- Develop ways of explaining the quality and source of linked and synthetic data and of limitations on their use. (For example, some synthetic variables or data about synthetic individuals may have been created by using methods that preclude their use in regression analysis.)
- Develop processes to ensure that new data tables do not allow the possibility of residual disclosure.

3.3. Access to microdata: progress to date and next steps

There has been much progress in recent years in using microdata for analytic purposes and in creating powerful linked data bases for research purposes as seen in Box 9. More are available in the Research Data Centres including those that use administrative data to extend the life of past longitudinal surveys.

Box 9. Examples of linked and linkable datafiles for research purposes

Longitudinal Worker File is used to measure the evolution of layoff rates. It consists of data drawn from:

- Record of Employment and T1 and T4 tax files
- Longitudinal Employment Analysis Program which provides employment-related information of businesses that, in turn is based on tax files and the Statistics Canada Business Register.

Longitudinal Immigration Database, which combines data from the Immigrant Landing File, the Non-Permanent Resident Permit File and T1 tax data.

Canadian Employer-Employee Dynamics Database. It is a set of linkable files that provide both longitudinal matched data between employees and employers including data from the following files:

- 7 income tax files, both personal and business.
- Record of Employment from Employment Insurance.
- Trade importer and exporter files
- National Accounts Longitudinal Microdata file.
- Longitudinal Immigration Database and the Temporary Residents File.

Education and Labour Market Longitudinal Linkage Platform. This is a particularly powerful set of linkable data bases that includes:

- Registered Apprenticeship Information System.
- Post-Secondary Information System which contains information pertaining to the programs and courses offered at an institution, as well as information regarding the students.
- T1 tax records.
- Canadian Education Savings Program.
- With plans to add more data including the Canada Student Loans Program, Record of employment and census data.

Today, microdata can be accessed from sites that are designed primarily to support research: The Research Data Centres, the Public Use Microdata File Collection, the Data Liberation Initiative. Microdata can be analysed at a distance through Real Time Remote Access or special requests. As noted, many tools are being developed to support researchers using microdata, including the Data Analytics as a Service Platform and the Virtual Data Labs.

However, the next step – using these rich sources to create standard multi-source databases that can be accessed through an interactive web site – is not on current planning agendas.

- Statistics Canada currently maintains one such output that is made public, the database of the *Social Policy Simulation Database and Model*, but it is only accessed in conjunction with its supporting microsimulation model.

A new product: standard multi-source microdata files

In a multi-source system based on microdata, an important new output will be a set of standard multi-source microdata files that can be used for variety of different purposes. Responsibility for creating these is currently seen as belonging to users of the statistical system, as opposed to being a major output of the statistical system itself.

In a future multi-source model:

- Statistics Canada would, based on consultations and often working in partnership, create and itself take ownership for a growing set of user-friendly microdata bases that would be designed to meet the needs of groups of users with common analytic interests. They would be anonymized and fully documented, including information about quality and sources.

- The inputs to these standard micro databases would be drawn from a variety of sources most appropriate to the topic at hand. Synthetic variables would be included if they met a specified minimum level of quality.
- Depending on the type of use that is envisaged, some of these databases could be based on real individuals using data masking techniques to protect privacy. Others would be based on synthetic individuals. Some might be longitudinal, others cross-sectional only.
- Only data that is consistent with existing, verifiable source data would be included. In other words, from the perspective of the user, the data would look like a traditional file from a survey or administrative source – only with a great many more variables.

A consultation strategy will need to be set up to determine the content of these standard databases. Early examples might draw on the linked research files in Box 9 or those in the Research Data Centres. The work could be done in partnership with the researchers involved.

3.4. Partnered ‘what is likely to work best’ databases

Box 6 identified the large potential payoffs from developing predictive tools that could assist in individual decision-making. This would be a new area for Statistics Canada, and it will be important to be clear on how this kind of tool relates to other more traditional evidence-based approaches such as ‘what if’ scenarios and ‘what works’ initiatives.

- **‘What if’ scenarios** also make projections based on microdata, sometimes using microsimulation modelling tools. However, these are designed for the examination of inputs and outputs at the level of programs and policies. They can show the effects of possible program design changes on groups of individuals with different characteristics. They can provide information on, for example, program costs that result from changes in the characteristics of program participants. Statistics Canada’s Social Policy Simulation Database and Model is a good example in the area of income support programming.
- **‘What works’ initiatives** are those that provide retrospective assessments of the effectiveness of a program or intervention such as an evaluation or a clinical trial that provides information on outcomes, as well as the relationship among inputs, outputs and outcomes. They are often rigorous, based on random control techniques. Each such assessment is, however, quite limited, in that it deals with only a single program intervention taken in isolation. Many often describe only average outcomes. Much progress has also been made in systematically examining many different assessments of this kind in order to draw conclusions about which kind of intervention is likely to work best in the future. Examples are the UK What Works Network, and the assessment made by non-governmental bodies such as the Cochrane Collaboration, the Campbell Collaboration and Arnold Ventures.

A skills shortage rationale

There is an inherent logic in the Statistical Agency itself taking responsibility for providing access to one of its most useful outputs. This is supported by considerations related to skill shortages.

- There are shortages of staff with the needed technical skills – both inside the statistical agency and in the user community – to create, link and use complex, multi-source microdata files.
- The needed skills must necessarily reside in Statistics Canada where the needed record linkage and metadata tools are developed and maintained.
- If the user community has main responsibility for constructing, documenting, and sharing linked databases, they too will need to draw on the same limited supply of skills. Progress will be slow, despite the important and welcomed improvements that are underway to provide new tools of microdata access and analysis to external researchers.
- If the statistical agency takes responsibility for their construction and documentation, working in partnership with external researchers, the total use of scarce resources among both the creators and users of the statistics will be minimized and the potential uses of the data will be maximized.
- Skill shortages are related to the stage of development of a technology. The final chapter discusses broader implications.

- **‘What is likely to work best’** initiatives that use individual-level predictive analytics of the sort being proposed here can also produce the system-level results found in the ‘What Works’ initiatives described above. However, they go much further by also providing real time information to inform individual decisions. This is done by providing information on which type of intervention has the highest probability of success for a specific individual, based on past experience in similar circumstances. It is based on far richer longitudinal data than that used in traditional ‘what works’ analysis and can, for example, take into account the combined impact of different programs on individuals over the course of their lives. The sidebar describes data developed by Employment and Social Development Canada (ESDC) which can be used for this purpose.

Implementation strategy

In the next several years, a realistic goal might be to:

- Identify a number of areas where ‘what is likely to work best’ information directed to individuals and services providers is likely to have early payoffs.
- Establish partnership agreements to conduct pilot studies in several of these areas.

Recent work in measuring the subsequent success of **post-secondary students** in specific academic programs suggests that early pilots in this area might be warranted.

The area of active **labour market programming** (training, job counselling, subsidized work experience, etc.) would be an obvious area to consider in establishing pilots if partnerships could be arranged with ESDC and one or two provinces:

- The ESDC predecessor department successfully piloted this approach some 20 years ago where information about the likely success of different kinds of interventions was provided to staff making referrals, at the time the referral was being made.
- The initiative was dropped when these programs were transferred to the provinces, but the techniques and data sources remain intact and used in evaluations.
- That ESDC data, described in the sidebar above, could be potentially housed in Statistics Canada, and quickly made operational in pilot settings while protecting privacy and confidentiality.

While predictive analytics are widely used in many applications, the pilots would examine their role in official statistics, particularly in calculating and explaining the level of uncertainty that is inevitable in estimating probabilities of success at the level of particular individuals.

- At minimum, where there was much uncertainty associated with the calculated probabilities, the technique could nevertheless still provide useful contextual information to the individual or counsellor in question such as descriptions of the interventions that were open to the individual and the average success of each.
- At maximum, a reliable estimate could be made of which intervention had the highest probability of success for that particular individual.

New ways would have to be found to communicate a range of such findings to the individual or counsellor in question, in simple language and in real time. Development work would also be needed to determine the degree of granularity in describing both the individual participants and the intervention in question. What is likely to work best on average may not apply to particular individuals with different needs and aspirations.

The data to support ‘what is likely to work best’ in Labour Market Program interventions

ESDC’s *Labour Market Program Data Platform* contains rich microdata from multiple sources about the socio-economic and labour market experience of the whole population and of those who participated in different types of employment and training interventions, both before and after the intervention – as well as information about the intervention itself. It can be used to measure the outcome associated with participation in different interventions, including the subsequent:

- Probability of employment.
- Employment earnings.
- Weeks in receipt of EI benefits.
- Amount of EI benefits received.
- Amount of Social Assistance benefits
- Proportion of EI use.
- Dependence on income support.

4. Processing: transforming inputs to outputs

This chapter deals with progress and next steps in the processing section of the logic model shown in Figure 1 – how the incoming data is processed into outputs. All of the processing boxes in Figure 1 will be touched, directly or indirectly, by the shift to the multi-source system. However, the main changes will be in three functions:

- Section 4.1 describes the conceptual framework that provides standard descriptions of data items, independent of the source of that data.
- Section 4.2 describes how the input data from multiple sources is managed and processed in a way that ensures quality and protects privacy – the Figure 1 box entitled *Maintaining internal databases to store and manipulate the data*.
- Section 4.3 describes the approach to documenting the quality of actual data in the system, as found in the Figure 1 box entitled *Maintaining the Data Documentation System*.

4.1. The conceptual framework

The goal is to develop a standard way of defining all the data items that are contained in the source files and that will be used in producing statistical outputs.

- In many cases, there will be information about the same characteristics from many input sources. For example, employment data is collected from the Census, from the Labour Force Survey, from tax files, from employers (e.g., the Record of Employment).
- The data from these different sources will often differ, if only slightly, for reason of reporting error, or differences in the wording of the original question that was asked, or because the administrative records were created to support somewhat different purposes.
- There is therefore a need to identify the ‘correct’ or standard data that is independent of the source of the data and that sets out the concept that is ideally needed for purposes of social analysis.
- The standard description of data items must be part of a consistent conceptual framework that supports analysis of the interrelations among all the data that are found in the database, including the factors that cause, or are associated with, changes in these characteristics over time.
- There are two such conceptual frameworks, one related to individuals and the other related to programs and institutions that provide social services and supports.

Progress to date and the need for additional action

The existing Statistics Canada policy on standards calls for:

- The use of conceptual frameworks, such as the System of National Accounts, that provide a basis for consolidating statistical information about certain sectors or dimensions of the Canadian scene.
- The use of standard names and definitions for populations, statistical units, concepts, variables, and classifications in statistical programs.
- The use of consistent collection and processing methods for the production of statistical data across surveys.

What is yet to be developed is:

- A comprehensive dictionary of standard definitions and concepts of variables that exists independent of particular data sources, definitions that are explicitly linked to the common conceptual framework.
- An explicit conceptual framework that encompasses existing definitions but that can be extended over time to create powerful new ways of integrating all the micro data in the quality-of-life statistical system.

Separate paper on the conceptual framework. A paper is being drafted that will describe the conceptual framework. It will show:

- How the characteristics of individuals will be described using stock, flow and lifecourse concepts. These are consistent with existing definitions but also allow new kinds of analysis of the interrelations among people in their social, spatial, and temporal contexts.
- How the characteristics of social programs and services will be described based on an input-process-output-outcome logic model. This will allow analysis of the effect of these programs and services on the lives of those who participated in them.

4.2. Managing and processing the data

The shift to a multi-source system requires a new approach to the management of data. Data become assets that have value in themselves, available to be used in a variety of applications, independent of their source.

- The standard definitions referred to above are an important part of the response.
- Other key dimensions are summarized in Box 5, including the sophisticated tools and mechanisms that will be needed to manage the huge amount of overlapping data in the system and to make full use of its potential in the statistical products that are produced.

Progress to date and the need for additional action

On technical matters, there will be many challenges ahead, but Statistics Canada is well positioned to take them on:

- **Data management strategy.** Statistics Canada has already developed a strategy for independent management of data, covering topics such as privacy and trust, IT strategies, metadata strategies and outreach. The strategy will, of course, evolve during the course of the transition to a full multi-source system, but the changes are unlikely to be fundamental.
- **Methodological and analytic tools.** Again, much development will work be needed to fully develop and document the technical tools that will be needed to collect and analyse data of high quality in a multi-source world. However, Statistics Canada is already a world leader in the development and use of many of these tools and the modernization initiative has further strengthened its capacity for leadership in these technical areas.
- **Data linkages.** The data linkage lies at the centre of the shift to a multi-use system and here as well, Statistics Canada is already well advanced. As noted earlier, it has developed:
 - ▶ A world-leading framework for balancing use, privacy, and response burden (the *Necessity and Proportionality Framework*).
 - ▶ Sophisticated tools for linkages that ensure privacy protection by techniques that split out individual identifiers from the data that describe those individuals (the *Social Data Linkage Environment*).

There will, of course, be a need for more development, including new tools for creating the linked data needed to analyze the course of people's lives in various social, spatial contexts. However, the need has been recognized, a solid start has been made, and future progress will likely follow pathways that have already been established.

The paper describing the conceptual framework will also:

- Outline the historic and intellectual background to the shift to a multi-source, micro-level framework.
- Show how it is consistent with other statistical frameworks, including the System of National Accounts and environmental indicators.
- Explain the relationship to the concepts and terminologies used in different disciplines, including the disciplines used by statistical methodologists, IT designers, and by the researchers in different scientific disciplines.
- Provide examples of its use, including in policy applications that involve system-level 'what works' applications and individual-level 'what is likely work best' applications.

4.3. Data documentation

Section 4.1 described the need for a standard set of concepts and definitions that is independent of data sources. There will also, of course, be a continuing need to document the actual data that comes into the system from various sources. That includes information on:

- The extent to which the actual data deviates from the standard definition.
- The source of the data and the methods used in its creation.
- Its quality, which Statistics Canada defines in terms of the multiple dimensions of the ‘fitness for use’ of the data in question.

Progress to date and the need for additional action

Existing systems of documenting data (known in statistical circles as metadata) are strong and being steadily improved.

- Statistics Canada is developing a new platform that will provide seamless access to all its internal data and related metadata descriptions.
- Sophisticated ways of accessing metadata have been developed to support the increasing use of microdata, including remote access.
- Work is ongoing on several fronts to develop standards and documentation tools that will apply to all the data holdings within the federal government – as part of the broader strategy to increase data flows across the whole system.
- Work is ongoing at the international level to build comparability in the approaches to metadata used by the statistical systems of different countries.

Statistics Canada is therefore well placed technically to meet the new documentation challenges that will arise in a multi-source statistical world.

- User-friendly ways will need to document sources and ‘fitness for use’ in tables or microdata files that were created from multiple sources. New approaches may be needed to provide simple documentation related to synthetic data. Traditional footnotes to tables or separate documentation packages may no longer be sufficient.
- At a later stage, when probabilistic ‘what is likely to work best’ statistics are produced for direct use by individual Canadians, new approaches may be needed to explain, in lay language, how users should interpret those probabilities in making their own decisions. As discussed in Section 3.4, there is always uncertainty associated with causal analysis and simple ways must be found to explain the extent of this uncertainty. This may require tailormade explanations that are generated during the course of each use.

5. Inputs: the multiple sources of data

The chapter deals with the ‘input’ set of boxes in Figure 1 that describe the incoming data. The *Governance*, *outreach* and *advice from experts* box is discussed in Chapter 6.

- Section 5.1 describes the increased responsiveness of a multi-source system to demands for new statistics.
- Section 5.2 describes response burden and the growth of hybrid collection vehicles.
- Section 5.3 describes approaches to accessing new administrative data.
- Section 5.4 describes ways of maximizing incoming data based on geography.
- Section 5.5 describes inputs that describe social programs and social services.

5.1. Responsiveness to requests for new statistics

In the past, requests for new data (for example to meet a new policy priority), would typically be cast in terms of requests for statistics from a particular collection vehicle. Often the request would wind up in the in-baskets of staff responsible for administering a particular survey or administrative data set. If the needed data did not

exist, and if the priority for obtaining it were high enough (often as demonstrated by the willingness to pay for the collection of new data), a new survey would be designed, or additional questions added to existing surveys such as the General Social Survey.

That process did work and produced important new information. However, it was typically slow and cumbersome, often taking years before the required data could be obtained.

In a mature, multi-source system, the interactive web site discussed in Chapter 3 will allow much easier access by users to existing statistical information regardless of source. Compared with today, many more requests will be able to be filled from existing data, including the expanded CODR tables and standard multi-source microdata files.

When the needed statistics are not found on the website, new approaches are being developed that will greatly improve responsiveness.

- Changes have already been made such that expertise in subject matter content and in the administration of particular collection vehicles will be located in separate organizational groups. The subject-matter content organizations are being designed to support more pro-active outreach – to better anticipate the need for new statistics and to work with clients in formulating those needs.
- Requests for data would be initially framed in terms of the content of the needed information, without reference to its source. Tools are being developed such that the first response to the request would be a rigorous search of existing data holdings, including information that could be obtained from the linkage of existing administrative and other data files.
- In order to respond to situations where new collection is still needed, new survey instruments are being put in place to maximize responsiveness. They are being designed to be flexible and fast:
 - ▶ Already in place are web panel surveys conducted as needed (currently every other month), to deal with hot topics with very quick responses. It is called **Canadian Perspectives Surveys** and asks a representative sample of individuals to voluntarily respond to online questions. It was piloted in 2020 and provided much important information related to COVID-19.
 - ▶ A new **Omnibus Survey** is under development. It will be a quarterly survey with 20,000 respondents. About a quarter of the questions will provide standard core content, including socio-economic variables. The other questions will vary depending on priorities. Some will cover social topics that are already collected in General Social Survey, but only at 5-to-7-year intervals.
- If the need cannot be met by data linkages or by these new survey vehicles, then less conventional sources will be explored. For example, Statistics Canada is currently piloting a **Well-Being App** where a sample of individual are asked to download a cell phone app and use it to report on their subjective feelings of well-being several times a day, along with information on what they were doing at those particular times. The results are linked to administrative tax data to provide socio-economic background information.
- If the need is truly urgent and the information cannot be collected by the representative survey instruments that are within Statistics Canada's normal mission, **non-representative means** such as crowdsourcing can be used, as it was to collect information related to cannabis use and the impact of COVID-19.

5.2. Response burden and hybrid collection vehicles

Non-response rates are increasing in surveys and response burden is a serious concern. This is one of the key factors behind the increasing use of administrative records whenever possible. Longitudinal surveys have been at special risk as non-response increases during successive waves of the survey.

Hybrid collection vehicles

Part of the solution has been to replace some questions that were once asked on surveys with data obtained from administrative sources.

- For example, the census and surveys such as the Survey of Household Spending now use tax records as the source of income data rather than asking respondents to provide this information.

- The Longitudinal and International Survey of Adults draws on tax, pension, and immigration files to complement the information provided directly by respondents.

This use of administrative data to replace survey questions both reduces response burden and increases quality, since often the administrative data is more accurate than that which respondents to surveys can recall.

It is therefore likely that future multi-source systems will make much greater use of what can be called hybrid vehicles that are designed from the outset to draw on both administrative and survey data. These hybrids can draw on the strengths of both sources and will often be much more powerful than either source taken individually. As indicated in the sidebar, this opens up the possibility of more flexible survey designs.

Do hybrid vehicles allow flexibility in survey design?

It is important in traditional surveys that exactly the same question be asked of all respondents. Does the existence of external control totals mean that questions (and introductory material) can be better tailored to the differing circumstances of respondents, with the same or better quality and reduced non-response? Can survey questions be formulated based on what we already know from admin sources?

Will response burden really diminish?

There may seem to be a conflict between the desire to reduce the burden on individual respondents to surveys and censuses and the objective of collecting much more detailed information. What is actually happening is more of a rebalancing:

- Data from individual respondents will shrink as a percentage of all data collected, given the much greater reliance on administrative data. This is particularly true for longitudinal surveys.
- As well, the data management techniques discussed in Box 2 mean that individuals will not always have to enter the same information many times in different administrative and survey applications, as seen in the previous discussion of hybrid vehicles.
- The ‘what is likely to work best’ applications, when these are developed, will require participants to provide considerably more information. However, this will be entirely voluntary and the participants will receive major direct benefit from participating. That is, response burden in these applications will not be an issue.
- On the other hand, new surveys will continue to be mounted as indicated in the preceding section and some of these will ask for quite detailed and often sensitive information. For the individuals selected to participate in these surveys, there could well be an increase in the amount of data collected. Whether this is seen as burden will depend on the respondent’s trust in the system, including an appreciation of the value of the resulting data to society. We return to this topic in Chapter 6.

5.3. Obtaining new administrative data

Major progress. In the traditional system, obtaining administrative data for statistical purposes has typically been difficult, even though the Statistics Act provides the agency with the power to access this information.

- From the data supplier’s perspective there were often technical and cultural barriers to sharing.
- From the perspective of those within the statistical system, obtaining access to the data was often seen as a sideline to the real work of conducting surveys, censuses and working with statistical summaries of administrative data that were compiled by the holders of those files.

As shown in Box 2 in the introductory chapter, we are entering a world where priority is placed on government-wide data management, on the use of digital technologies and on open government – a world where there will be easier access to administrative files for statistical purposes, including the opportunities to shape the content of those administrative files.

In practice, huge progress has already been made in the use of linked data from multiple sources as can be seen by the examples in Box 9.

Short-term action. Despite this success, we are only at the beginning of a journey.

- Much still needs to be done as the COVID-19 experience has illustrated. A lack of shared data proved to be a major problem that delayed responses in dealing with the pandemic.
- Even when agreement to share data exists, there is often a lengthy period before it arrives.

For those who are involved in the shift to a multi-source system of national quality-of-life statistics, future action must be taken in the context of the broader data management initiatives referred to in Box 2 and, perhaps, of political-level, post-pandemic discussions about better data sharing. Next steps to *supplement* these broader initiatives could include:

- **Building consensus on the priority gaps that could be filled by administrative data.** This would involve consultations with stakeholders and experts whose purpose would be to identify topics where the gains from accessing new administrative data would have high payoffs, especially when combined with data from other sources. Emphasis would be placed on how the resulting new statistical information would benefit those agencies that would provide the new administrative data.
- **Developing a longer-term implementation plan.** Based on these consultations, strategic plans could be developed for accessing the missing data, including timetables that would be realistic from the perspective of the suppliers of the administrative data. Existing governance arrangements in areas such as education or justice could oversee the development of these plans.

5.4. Incorporating geographic information in individual records

There has been much progress in Statistics Canada's use of geomatics and spatial visualization tools and further strengthening is underway. Basically, these tools allow statistical information to be shown on a map of a particular neighbourhood, municipality, region, health regions, or province. This becomes a particularly powerful method of showing interrelationships when data from different sources and different subjects are shown on the map in question.

The information will be also contained in the records of individuals. In a mature multi-source system, this rich new source of data will be available not only at the level of the geographic area but will also be fed back directly into the records of the individuals themselves at the level of the geographic location of the individual's residence. This will allow much richer analysis of the determinants of the individual's quality of life in areas such as access to services and employment opportunities, local costs of living, or quality of housing. This would be an important new source of data that would allow richer analysis of individuals in their economic and environmental contexts.

Statistical information that can be both mapped and incorporated in the micro records of individuals

- **Access.** The number of hospitals, restaurants, elementary schools, grocery stores, community centres in a geographic space.
- **Environment.** Average local temperature, number of parks and the extent of green areas, measures of pollution, health indicators based on sewage wastewater, walkability indices, the incidence of disease including the effects of pandemics.
- **Economy.** Local cost of living indicators, vacancies in the firms located in that area and other indicators of economic activity.
- **Society.** Data on the socio-economic composition of geographic areas created by adding up the characteristics of the residents in those areas – resulting in data on income inequality, extent of victimization, average educational attainment, ethnic diversity, health indicators, and housing density.

There are already many initiatives related to spatial analysis and there would appear to be potential gains from additional work that incorporates these elements into analysis at the level of individuals.

5.5. Inputs describing social programs and services

The shift to the mature multi-source model will eventually result in micro-level data about social programs and services being included in the statistical system as described in the box below.

Traditional versus micro descriptions of social programs and services

Traditional descriptions

Traditional statistical systems described social programs and services at a macro level using tabulated data, such as:

- The costs of health, education, and judicial systems, and of income support programs.
- The general type of service provided, such as treatments related to different diseases, or different types of education provided, or the amount of income transferred in a tax credit program or by social assistance.
- Characteristics of the recipients of those services.

Trends over time have been to produce increasingly more disaggregated information of this type – for smaller geographic areas and more granular descriptions of services and clients. Indeed, some health and educational data now exist for particular hospitals and schools.

Micro descriptions

For each interaction or intervention between the client and the program or service in question, there would be:

- Standard ways of describing what happened in the organization (in doctors' offices, in classrooms, in the rules for calculating the size of income transfers).
- Standard ways of describing the individuals who were served, including their detailed health, skills, and income status before and after the intervention – and later in life.
- Standard ways of describing the actual activity that took place during the interventions, including its duration and other descriptors that will be described in a forthcoming paper entitled "The System of National Quality-of-Life Statistics: Conceptual Framework".

Such microdata now exist for programs that provide income support and are contained in Statistics Canada's Social Policy Simulation Database and Model. Consistent, comprehensive data of this sort do not yet exist for most services.

Such micro-level data about social programs and services will allow analysis of the actual difference that particular social interventions made in the lives of individuals – the kind of information that is essential for:

- Shifting to the client-centric social programming described in Box 6.
- Introducing 'what is likely to work best' pilots described in Section 3.4.

As noted in the box, the needed micro data already exists in the case of income transfer programs. However little progress has been made for services:

- It does exist for Labour Market Programming as shown in Section 3.4.
- Much of the needed data exists in some areas of post-secondary education and apprenticeship.
- In health, some of preconditions are being put in place through means such as the digitization of health records. COVID-19 assessments may give a push to further action.

However, it will take time to develop a comprehensive approach to describing social services using the consistent micro-level descriptors that are discussed in the Annex.

This suggests a gradual approach to development in this area, similar to the approach for 'What is likely to work best' pilots. The first steps would be to identify areas of low-hanging fruit and find partners willing to develop and apply these descriptors in their area of expertise.

6. Managing constraints and risks

The future directions discussed in this paper will take Statistics Canada into some quite new territory, some of it world leading. This comes with some constraints and risks during the implementation process that must be carefully managed.

- Section 6.1 discusses resource constraints, particularly those related to skills shortages.
- Section 6.2 discusses ways of ensuring understanding and support in the user community.
- Section 6.3 discusses ways of maintaining and building trust and support among the public and those who provide the input data, including individuals who respond to surveys and the program administrators who provide administrative data.

6.1. Skill shortages

In terms of resource constraints, dollar funding can be largely managed by internal reallocations, many of which have already taken place, and by the new funds that were allocated in Budget 2021.

The agency's IT plan will need to be adjusted in light of the implementation plan. Some of the newer technologies require a lot of computing power that may require considerable lead time to put in place. However, the technology itself is not new and is being used in other applications.

The big resource challenge will therefore likely be a shortage of people with the needed skills. As already discussed, creating and using multi-source microdata for statistical purposes requires individuals with skills sets that are in short supply.

The literature on the typical life cycle of technologies, summarized in Box 10, suggests we are still at early stages of development where the technology is complex and requires high skills levels to operate. Especially in the early stages of the transition period, the best allocation of scarce resources would be for the same team to be involved in the creation and use of complex multi-source data bases (and in related projects involving complex techniques such as microsimulation and predictive analytics).

Otherwise, staff with about the same skill set will be needed in both the producer and user communities. Or, to be more realistic given the shortage of these skills, there will not be adequate skills available for either community to do a good job.

In practice, this means that:

- Statistics Canada should take the lead in most such projects as it will typically have the greatest expertise. Statistics Canada also provides a secure and trusted location in which data linkage can take place without threats to privacy. And, as suggested earlier, it makes common sense from an accountability perspective that Statistics Canada itself should take ownership for standard multi-source micro databases especially those that will become a key product of the new system of national quality-of-life statistics.

Box 10. The life cycle of technologies and skill shortages

Technologies, such as those used in developing and using large multi-source files, go through typical life cycles.

In the early stages of development, technological factors dominate:

- The technology will typically not be easy to use. Instability, difficulty in use and inelegant appearance will often be the norm. Documentation will often be weak.
- In the early stages, there is also a blurring of the lines among those who develop, maintain, and use the technology.

Problems associated with source data, with technological development, and with new kinds of analysis are best dealt with as an integrated package, by the same team of people.

In the middle stages and later stages, the technology moves to the background. The functions of development, maintenance and use become separated:

- Work on product innovation takes place in separate R&D groups.
- Maintenance becomes a separate function that includes setting standards, documentation, training materials, and marketing – making things easier for the user.
- Users will acquire the technology as a commodity, with emphasis on its speed, flexibility, ease of use, and low cost. The technology itself will be largely invisible, the accuracy of the results simply being taken for granted.

- Statistics Canada should bring in partners from the user community and academia at the outset of development. Skills would be shared during implementation, and skill base in the user community would grow as a result of their experience during the development period.

Such a model may require some adjustment to existing approaches to data linkage. As noted earlier, some existing plans assume that Statistics Canada's role is to provide users with the tools to develop high quality, anonymized multiple source statistics. The analysis here suggests that, in many cases, it should be Statistics Canada itself that plays the lead role, in partnership with others.

Box 10 suggests that, in addition to existing recruitment and training practices, there may be a need for a specific HR strategy to support a smooth transition to the skills that will be needed as the technology matures.

6.2. Building understanding and support in the user community

A main objective in shifting to a multi-source model is to better serve users. However, most users, partners and potential funders of social statistics will be unfamiliar with techniques that will be used and of the usefulness of the new data that will be created. Concepts such as multi-source microdata files, synthetic data, predictive analytics, and lifecourse analysis will be unfamiliar to many, even among experts. A risk therefore is that Statistics Canada could be seen as moving off in unknown directions, trying to meet new needs and to anticipate the needs of existing users rather than responding to filling existing gaps and being responsive to needs that have already been formulated.

That risk will be mitigated by the early development of the interactive web site which will provide speedier access to all existing information and by the new approach to responding to current requests for information, including the new suite of flexible survey instruments, that will produce results quickly. However, these steps should be supplemented by outreach mechanisms that will allow users to work in partnership with Statistics Canada in developing practical ways of applying the new statistical evidence that will be produced.

An emphasis on outreach. Much recent attention in Statistics Canada, including in the social area, has been devoted to the issues of horizontality, partnership, and outreach. New centres have been created with a view to increasing the capacity to respond quickly to changing demands for statistics and to anticipate new demands. This emphasis on outreach is important but needs to be carefully managed. To be effective, it should be guided by the following considerations:

- Outreach should not be based solely on traditional subject-matter policy silos such as health or education. This could result in cross-cutting topics being ignored, such as equality or wellbeing. Wellbeing in different population subgroups often results from combinations of interventions from different traditional policy silos.
- Many disciplines are involved in the development of the statistical system and in its uses – and they tend to use different conceptual frameworks and terminologies. One of the goals of outreach should be to use the neutral territory of statistical planning to help build bridges across these communities.
- Given the often-long time horizons in statistical planning, outreach should be with those that have long-term view as well as those concerned with current policy needs.
- The risk of fragmented outreach grows if there are many small specialized organizational groups each having independent outreach strategies.
- Outreach and consultation take much time and use skilled resources. They should never be treated as ends in themselves. They should always be directed to a particular purpose.

A balanced outreach strategy to support the evolution of the multi-source system might therefore include the following strands:

- **Dashboards and web-based products.** Outreach directed to developing the new dashboards that will be on the interactive website. This would be an obvious responsibility of the subject-matter divisions of Statistics Canada responsible for social statistics. It could be associated with consultations on the new content of the CODR tables and the standard multi-source microdata files.
- **Development of the conceptual framework** and standard definitions described in the forthcoming paper entitled “The System of National Quality-of-Life Statistics: Conceptual Framework” should be accompanied by a consultative strategy that engages many groups inside Statistics Canada and in the academic and policy communities. Discussions on the concepts to be used in measurement are close to discussions about what should be measured. That is, the statistical system can, implicitly, provide a neutral space that can foster cross-silo consensus-building around longer-term directions in the content of social policy.
- **Leading-edge analysis** where advice and partnership are sought on the conduct of research on high priority policy research topics that often cross sectoral boundaries and that use multi-source data in innovative ways. The Analytic Studies Branch already has a workplan organized along these lines.
- **Improving flows of incoming data**, particularly the longer-term strategy related to administrative data referred to in Section 5.3.

Governance

The multi-source system will be designed to provide integrated statistical information that cuts across existing policy silos. Existing governance structures in areas such health, education and justice statistics are based on those silos, reflecting the differing federal and provincial roles in various areas of social policy.

- The shift to a multi-sources system does not, however, necessitate any change in the complex, asymmetrical governance structures that have emerged in these areas over time.
- The parties in existing governance structures will gain important new information from a multi-source system. They will have at least as much a stake in its successful development as they do in the existing system.
- The eventual addition of ‘what is likely to work best’ applications and of the inclusion of micro descriptors of social programs will require strong partnerships with those who have responsibility for existing social and health policies and programs. If the existing governance structures did not exist, something like them would have to be invented in order to oversee these partnerships.

Coordinating outreach?

Different strands of outreach from different parts of the organisation are currently coordinated by informal means, such as telephone calls among colleagues. Will the new strands require a more formal means of coordination?

- **Parallel consultation strands with Indigenous peoples and with groups with distinctive needs.** This could build on existing outreach initiatives related to gender, diversity, and inclusion -- including with immigrants, people with disabilities, racialized minorities as well as people in remote areas. These would focus on the information needed in these communities.
- **An international outreach strand** could be considered since the challenge of developing a multi-source system of national quality-of-life statistics is of world-wide interest, with the potential for sharing experience

An emphasis on partnership. Many of the outreach activities above will involve close partnerships with users. In addition, as described earlier, it is proposed that particularly close partnership relationships be established in those areas that are at the leading edge from a technical perspective:

- Partnerships with researchers and analysts in other departments, in academia and think tanks in building the standard multi-source microdata bases.
- Partnerships with agencies responsible for service delivery in developing ‘what is likely to work’ pilots.

6.3. Ethical issues: building trust

Any system that makes use of data about individuals from different sources will, properly, raise alarm bells related to privacy, consent, and confidentiality. These issues are not new. The question of getting the right balance between these concerns and those related to the value of the uses of official statistics and the problems of response burden have been the subject of in-depth discussions for many years in many countries.

The privacy-related issues are part of a wider set of ethical concerns relating to the social acceptance of the boundaries of official statistics and the role of the agencies that produce them. These include the independence and neutrality of the statistical office, the transparency of its operation, trust in the quality and relevance of the statistics that are produced, the accessibility of statistical outputs, and the efficient managing and protection of information within the system.

Canada, and its statistical agency, are well-positioned on all these ethical fronts:

- The independence and neutrality of Statistics Canada, and the quality of its work, have never been in doubt. Its independence was reinforced in recent amendments to the Statistics Act.
- As noted earlier, Statistics Canada has already put in place a world-leading framework (called the Necessity and Proportionality Framework) for balancing use, privacy, and response burden.
- Again, as noted earlier, sophisticated tools have already been developed that ensure privacy protection when drawing on data from multiple sources, using techniques that split out individual identifiers from the data that describe those individuals. Linked files of identifiable individuals will not be created.
- The activities of the Statistical agency are seen as part of government-wide initiatives, described in Box 2, that emphasize the effective management of information resources including the reduction of response burden, the full use of all data collected as well as privacy and confidentiality.

The directions outlined in this paper are designed to further strengthen trust and acceptance in the population at large and among those respondents who provide data inputs:

- Most important will be the increasing usefulness of the statistics to those who also provide the input data.
 - ▶ This will be particularly important as ‘what is likely to work best’ evidence grows, where important statistical information will be directly provided to those who supply the input data, both program administrators and individual participants.
 - ▶ More generally, disaggregated statistics that recognise the diversity of the population will be seen to be more relevant and useful to those who provide the needed inputs, including survey participants.
- New emphasis will be placed on the role of Statistics Canada itself – a secure and neutral location – in producing statistics that are based on multiple sources of data, with built-in assurance of both quality and privacy protection, reducing past reliance on external users to carry out some these functions.

- The use of synthetic data will provide even greater anonymity of the resulting statistics, particularly in applications where the data are based on synthetic individuals created by statistical manipulation techniques.
- Implementation of initiatives where there are issues surrounding quality, such as the uncertainties associated with estimating probabilities of success at the level of individuals or the use of cloud-sourcing, will be conducted only on pilot or trial basis, until its quality is assured.

As implementation proceeds, a more immediate and transparent balance will be struck among response burden, usefulness and the protection of privacy – strengthening the implicit social contract that exists among those who provide, process and use data for statistical purposes.