SCHOOL OF
Policy Studies

Queen's
UNIVERSITY

A FRAMEWORK

FOR CITIZEN-CENTRED

SOCIAL STATISTICS AND

ANALYSIS

Peter Hicks

ABSTRACT

Economic policy-making has long been more integrated than social policy-making in part because the statistics and much of the analysis that supports economic policy are based on a common conceptual framework – the system of national accounts. People interested in economic analysis and economic policy share a common language of communication, one that includes both concepts and numbers.

This paper examines early attempts to develop a system of social statistics that would mirror the system of national accounts, particular the work on the development of social accounts that took place mainly in the 60s and 70s. It explores the reasons why these early initiatives failed but argues that the preconditions now exist to develop a new conceptual framework to support integrated social statistics – and hence a more coherent, effective social policy.

Optimism is warranted for two reasons. First, we can make use of the radical transformation that has taken place in information technology both in processing data and in providing wide access to the knowledge that can flow from the data. Second, the conditions exist to begin to shift away from the straight jacket of government-centric social statistics, with its implicit assumption that governments must be the primary actors in finding solutions to social problems. By supporting the decision-making of all the players (particularly individual citizens) who affect social trends and outcomes, we can start to move beyond the sterile, ideological discussions that have dominated much social discourse in the past and begin to build social systems and structures that evolve, almost automatically, based on empirical evidence of 'what works best for whom'.

The paper describes a Canadian approach to developing a framework, or common language, to support the evolution of an integrated, citizen-centric system of social statistics and social analysis. This language supports the traditional social policy that we have today; nothing is lost. However, it also supports a quite different social policy world, one where individual citizens and families (not governments) are seen as the central players – a more empirically-driven world that we have referred to as the 'enabling society'.

A FRAMEWORK FOR CITIZEN-CENTRED SOCIAL STATISTICS AND ANALYSIS

Peter Hicks, July 2012

1 – INTRODUCTION

This paper describes the way in which data should be structured and integrated in the official system of social statistics. In the 70s, social accounting was perhaps the main approach to developing the needed conceptual framework to support such an integrated system of social statistics. Social accounting was seen as a counterpart to the system of national accounts that had proved to be such a success in structuring and integrating economic statistics. The paper reviews the history of social accounting and describes an updated version of an integrating framework – one that builds on historical precedents but that also reflects the radically changed circumstances that now exist.

As explained in Box 1, the paper is a slightly revised version of a paper produced under a 2011 contract with Statistics Canada to provide that agency, and HRSDC, with advice that could be used in long-run thinking about the future of the system of social statistics.

The audience to whom the paper is addressed

While many players in both of the original sponsoring organizations are well versed in the topics under discussion (and, indeed, some are world leaders in areas such social indicators, micro-simulation modelling and other current techniques of meta-analysis and information management), this paper is intended to provide a simple introduction to the topic in ordinary language.

The paper attempts to answer two tough questions about a radically changed statistical system

The terms of reference for the original project as set out in Box 1 are ambitious, to use an understatement. This paper will be structured in a way that sets the stage for answering two particularly tough questions that are raised, or implied, in the terms of reference.

Tough question 1: what constitutes a conceptual and measurement framework in a world where priority is placed on social data to meet the future informational needs of Canadians, businesses and institutions and that covers multiple academic fields?

This definition of the audience for official statistics goes far beyond the traditional scope of official statistics which, in the past, have mainly been seen as supports to government policies and programs and which have relied on data that is almost entirely based on only one academic field – economics. In this new definition of the audience, the primary user of the data is the individual citizen or institution. An example of a main use of the future statistics system would be information that individual citizens would receive directly from the internet that would enable them to learn directly from the past experiences of others in making their own decisions about learning, jobs and health. This information would be tailored-made to their individual circumstances and would be available instantly. This is far away from today's system of official statistics.

Box 1. The paper was first commissioned by Statistic Canada, with the support of HRSDC

The original paper was commission as part of a larger 2011 project sponsored by Statistics Canada with the support of Human Resources and Skills Development Canada (HRSDC). The larger terms of reference call for the development of

> …*a conceptual and measurement framework for renewing Canada's social data system to ensure that future social data can meet informational needs of Canadians, businesses and institutions and that the data infrastructure is efficient and flexible. The 'renewed' social data system should also support the development of data that inform policy developments.* … *and should cover multiple fields (e.g. economics, sociology, psychology).*

The larger project also encompassed two other papers: (1) a more detailed paper on the role of life-course perspectives in the future structure of social statistics and (2) a paper dealt with the related, but quite different, topic of social indicators – particularly the capabilities and human development approach associated with Sen. We refer to many of the insights from both of these papers later in the present paper.

The author is grateful to the sponsoring organizations for allowing him to prepare this paper. It a great testimony to their commitment to openness that they commissioned papers that take a quite different perspective on longer-term directions in social statistics. This includes the proposal in this paper for a radical change in the system of social statistics – one that is citizen-centric, not government-centric.

Indeed, the definition of the audience for official statistics outlined in the terms of reference described above goes far beyond the traditional scope of official statistics which, in the past, have mainly been seen as supports to government policies and programs. In this new definition of the audience, the primary user of the data is the individual citizen or institution. As well, our current social data is almost entirely based on one academic field only – economics. If seriously meant, the terms of reference call for a revolution in official social statistics.

At the outset of the project, the author sought guidance on whether the implied radical change in the audience for social statistics should be interpreted as a mainly a rhetorical, preambular statement, or whether it was to be taken seriously in developing the paper. Assurances were given that it was seriously meant – that the two departments were truly interested in exploring a citizen-focussed approach to social statistics.

In these circumstances, it is especially important to underline the traditional warning that the somewhat controversial views expressed in this paper are only those of the author, and not those Statistics Canada or HRSDC. In particular, the last Q&A at the end of the paper is new to this edition and was never seen by Statistics Canada or HRSDC officials.

The author wishes, in particular, to thank Marie Drolet at Statistics Canada for organizing the project and for her work in commenting on earlier drafts.

Tough Question 2. How does the enormously strengthened power of current technology, and the even more powerful technology that will soon be widely available, reshape an <u>efficient and flexible</u> data infrastructure?

Today's data infrastructure is mainly organized around particular collection vehicles such as surveys, censuses, experiments and admin files. This structure was dictated by the information processing capacities of the pre-computer age. Shifting to the social accounting approach suggested in this paper makes use of today's database and web-based computing and dissemination techniques, a shift that will result in a very different approach to planning, financing, organizing and disseminating social statistics.

As will be explained in detail in the paper, today's system of social statistics is not only pre-computer age in its structure and processes, but it also reflects pre-computer thinking in its content and uses. Shifting to current technology will be easy at one level; almost all the new techniques that will be required have been thoroughly developed and tested in other domains.

However, the shift, when it comes, will be radical in its impact on organizational and job structures, as well as on the human resource skills and competencies that will be needed. While not specifically part of the mandate of this paper, it does therefore briefly touch on some practical implementation issues.

Outline of the argument

Ideally we would have liked to tell a story that described: the changes in the technologies of data collection and storage that are ahead of us; the even more radical changes in knowledge dissemination technologies; the transformative shift to the new citizen-based audience for social statistics; the consequential shift away from an almost exclusive reliance on surveys and censuses to far greater use of cheaper and more relevant administrative data; and the content of a new conceptual framework that corresponds to today's, and tomorrow's, needs for a system of social statistics.

We would have liked to tell that quite complex story in an integrated way and from an historical perspective. However, that task proved to be well beyond our story-telling skills. Accordingly, we have settled for dividing the paper into more manageable and traditional chunks:

- Section 2 provides background material on the evolution of social accounting.

- Section 3 describes the desired structure of social data today, and the need for a supporting integrating framework.

- Section 4 describes the Olivia framework – a contemporary conceptual framework that provides a common language for policy-related social measurement and analysis.

- Section 5 describes the radical implications of putting the emphasis on uses by citizens, firms and groups and of using new real-time dissemination technologies. It provides an example of how the new system would work in practice and its implications for consultation strategies.

- Section 6 sets out some conclusions and provides some Q &As which act as summary of the more important parts of the paper.

2 – THE EVOLUTION OF INTEGRATING FRAMEWORKS

This section is structured as follows:

- Section 2.1 shows where integrating frameworks fit into the recent evolution of social statistics.

- Section 2.2 summarizes the main debates in what we call the golden age of social accounting, in the late 60s and 70s.

- Section 2.3 reviews the state of social accounting, including its future directions, as it was seen at the end of the golden age.

- Things did not evolve as foreseen. Section 2.4 explains the very different, much more pragmatic, history of the period since the 80s.

2.1 Context

Social accounting is one dimension of the search for a comprehensive, integrated approach to measuring social progress. We might, arbitrarily, distinguish three main trends that have had a major influence on the planning of official social statistics over the past 50 years:

- In the practical collection of new data, there has been reasonably steady growth, but a growth that has been centred on discrete collection vehicles with little system-wide planning. Starting in the 70's, there was rapid growth in the numbers of cross-sectional sample surveys covering many domains of social life. In the more recent period, there has been significant growth in the area of longitudinal surveys and greater statistical exploitation of administrative records.

- In the area of the tools of analysis, transformative changes have been taking place as a result of new computing power. However they have mainly been used to undertake traditional, pre-computer analysis better, faster and cheaper. There has also been some modest development of new information-age analytic techniques. These include the increasing power of tools that allow us exploit microdata more effectively, with microsimulation modelling being the most obvious example. The potential of new web-based tools of dissemination have barely been tapped.

- In the area of conceptualizing statistical systems, progress has been quite uneven. There was much activity in the 60s and 70s, followed by forty years in the wilderness, followed by renewed interest in the last few years. (Conceptualization of statistical systems includes such things as the reasons for collecting data, the theories of society that determine what is important to measure, and how the data should be structured and integrated.)

The social accounting movement belongs to the last of these categories, where it was originally tied closely to the social indicators movement. Unlike today, where social indicators are usually treated as a distinct, highly structured and selective component of social statistics, in earlier

periods social indicators and social change were sometimes seen to be almost synonyms – except that social indicators referred to those social changes that were of greatest interest to policy. The two became seen as being separate, but related, during the golden age debates.

2.2 Social Accounting: Its Golden Age

...Its rise

Social accounting flourished during the relatively short period from the late 1960s until the late 1970s, mainly in the United States. Characteristics of this period include:

- It was a period of relatively high economic prosperity and economic optimism. (There had been a similar period of interest in the measurement of social progress in the late 1920's, a period with a similar economic climate. The Great Depression ended that period of interest in measuring social progress.)

- It was also a period of rapid social change and rapid change in social policy. These were the years when the post-war welfare state architecture was being completed in most western democracies.

- Rational systems thinking was being introduced in many fields. This included the general use by governments of a rigorous planning, programming and budgeting system (PPBS), built around costed activities within a hierarchy of outputs and outcomes. The PPBS framework was first introduced in the mid-60s in the Defense Department in the US – and spread to Canada where it still, in modified form, underlies the planning and budgeting architecture of the federal government.

- The System of National Accounts had just been extended into the form that we know today. The social accounts movement was an attempt to further extend national accounting principles on the social side – or, in other cases, it was an attempt to emulate national accounting principles using other means. Indeed, key figures involved in establishing the 1960s national accounting framework, Richard Stone and Richard and Nancy Ruggles, were key figures in the social accounting movement.

- The power of computers to revolutionize social statistics collection and analysis was starting to become clear, but the computers that could be used in practice were still primitive by current standards. Visionaries could foresee developments such as microsimulation modelling or the creation of new data by combining micro data from various survey and administrative sources, but applications were very costly and technically challenging given the computing power then available.

...and fall

The factors above help explain the existence of this brief golden age of social accounting. It quickly faded away in the 80s for analogous reasons:

- Economic times became tougher and there was less optimism on the social front. The task was to refine the new welfare state architecture and control its costs. There

was little appetite for innovative thinking about new directions in social policy, let alone its theoretical structure.

- Expectations created about the benefits of a rational, systematic approach to planning were far higher than could be realized. Practical, operational systems did not exist for the data collection and analysis that were contemplated by the social accountants in this period. It is only recently that the needed computing power has become available on a wide-spread basis.

- More generally, the impossibly high expectations created by rational planning systems such as PPBS were never realized. They turned into paperwork monsters. People became leery of initiatives where the main effort was devoted to feeding an elaborate system, without reference to what was really important. This worry spilled over to concerns about elaborate statistical accounting systems.

- The payoffs from practical, incremental improvements to existing data proved to be large, as were the payoffs from new surveys that were planned more or less in isolation, free of any larger integrating framework. As noted, the growing computing power of those years was more suited to (greatly) improving the efficiency and effectiveness of traditional approaches to social statistics.

Accordingly, interest in social accounting faded quickly.

- Key figures turned to other subjects. For example Thomas Juster, perhaps the key thinker that shifted social accounting to a micro basis, moved on to more practical pursuits – taking charge of the leading-edge longitudinal Health and Retirement survey in the US.

- Work continued sporadically in some areas such as the extensions of the national accounts through use of SAM's, but these were mainly seen as ways of broadening the scope of economic statistics a little – not as a new direction for social statistics itself (and, indeed, were largely unknown to people interested in social statistics).

- Interest in social indicators did not die, but its focus shifted away from indicators based on highly structured theories of society to more practical compendia that were developed with existing data limitations very much in mind – and without any attempt to provide an integrating framework for the whole of social statistics.

- Interest in new analytic techniques such as microsimulation continued, but with low profile. For example, while Guy Orcutt introduced the microsimulation approach in 1957, it took another 30 odd years before computing technology had advanced to a point where static models were feasible in many applications – and even now they are often considered as quite specialized tools. Dynamic models of the sort that were envisaged in the golden age still remain mainly in the developmental stage, with limited use for general policy analysis. It is only in the last ten or so years that the technology has become available that would allow their more widespread use.

The golden age dialogues

The main themes of the social accounting dialogue of the late 60s and 70s are, happily, easily found. A major workshop took place very near the end of the period in question, in Washington in March 1980, where all the big names in the field discussed the state of the art as it had developed to that point and contemplated future directions. The results were published in a superb book, *Social Accounting Systems: Essays on the State of the Art*[1] that includes papers from the main players involved and overall assessments by key figures.

While the source material is easy to find, summarizing it is more difficult. The contents of the various papers are rich, detailed and thoughtful – covering both technical matters and matters relating to the uses of statistical knowledge. (It is a surprising, welcome contrast to the comparatively narrow content of many current workshops and symposia on similar topics.) However, it is not a quick read.

Accordingly we will summarise at a very high level of generality and will draw heavily from the assessments of the workshop (by James House and, especially, by Richard Ruggles) which conclude the book, even paraphrasing their words in places.

...Areas of general agreement

There was general agreement that social accounting should address two main faults in the social statistics that then existed:

- Existing statistics were descriptive (cross-sectional, in today's language), rather than dynamic and analytical. They were helpful in understanding where we were in terms of social progress, but not in how we got there or what the future might hold. That is, they were weak in helping us understand cause and effect.

- Existing statistics were not easily integrated. There was little ability to see how changes in different aspects of society related to each other. For example, there were data on both subjective well-being and on objective characteristics, but they could not be used together to get an integrated understanding of the sources of well-being.

There was also agreement that social indicators should be seen as outputs of a statistical accounting system, not as part of that system.

The general solution was to create accounting systems that would: (a) show how social data were related to each other and with economic data and (b) that were based on flows, which enable a better understanding of inter-relationships and that provide a better handle on understanding cause and effect.

...Time-based accounts versus demographic accounts

Papers were presented on a range of approaches, but the main focus was on two types of general-purpose accounting systems:

---

[1] Edited by F Thomas Juster and Kenneth C Land, Academic Press,1981, ISBN 0-12-392550-9

- Demographic accounts as proposed by Richard Stone with a related version by Land and McMillan.

- Time-based accounts, as proposed by Thomas Juster and colleagues.

Demographic-based accounts are based on national accounting principles. The basic idea is that one can examine tables (matrices) that show flows (or states and flows[2]) among different population groups (i.e., not individuals) with different characteristics at different points in time.

- This approach results in an integrated approach to examining changes in status including many social dimensions – education, health, marital status, income, wealth or even subjective well-being status.

- The demographic dimension of the accounts allows life expectancies to be divided up into expected time in a whole series of states.

- Economic links enable costs or benefits to be associated with the relevant states.

Time-based accounts use two types of accounting principles: one based on national accounting principles and the other on time use. This approach is based on micro (individual level) data. It provides an integrated approach to subjective and objective measures of well-being.

- Time is allocated to market production (working for money) and to a range of non-market activities. These activities produce a range of outputs (earnings which can buy goods and services, a clean house, improved health, etc.). The activities add up to the 24 hours that are available in a day, providing a powerful accounting tool. The satisfaction obtained from the outputs can be measured – as can sustainability (the capacity to continue on in life).

- The outputs produced can be associated with the capital stocks associated with time use, including physical capital, financial capital and human capital.

- There are three levels of accounts: a household output account, a capital account and social output account. The capital account provides the link to economic statistics and the SNA, while the social production account deals with matters related to social well-being.

The discussion of the two approaches was polite and constructive – to the effect that these were both useful approaches that should, and could, be treated as being complementary not rivals. They were reconcilable, but development work was not yet sufficiently mature to allow a full synthesis. Work on parallel tracks was still needed.

---

[2] In Richard Stone's version, the state of the population at a given point in time is viewed as a stock, while changes in status (births, deaths, and other changes of state) are viewed as flows that alter the state of the population. The Olivia framework follows the Stone usage.

The Land version is conceptually the same but is expressed in terms of the inflows of population into a given time period (births, immigration and population surviving from the earlier time period) as being exactly equal to the outflows of population from that time period (deaths, emigration and population surviving the to the next time period).

and the winner was …

This constructive tone did not mean that commentators felt the two approaches were of equal merit. It is quite clear by the end of book that time-based accounts were seen as playing a far more dominant role in future development than demographic-based accounts – although both had a role.

- Two main problems were identified with using demographic accounts as the main approach. The most important was that these accounts were not based on micro data and, in consequence, were far less flexible. They could not be easily used to analyse many variables at the same time. As well, their reliance on large matrices – even when only a few variables were at play – raised the worry that far more data would needed to be collected than was really needed. That is, much effort would be devoted solving technical accounting-type problems as opposed to with problems that have a high substantive priority.

- The micro-data approach of time-based accounts was thought to have the potential for far more efficient manipulation, collection and storage of data (including from multiple sources) and allowed far more integration (e.g., between objective and subjective data). It allowed use of powerful microsimulation tools that had been proposed earlier by Guy Orcutt. As well, the micro approach could readily take account of macro data – such as the control totals needed to ensure consistency at the level of the economy as whole. Macro data also have the potential to take account of micro findings but this would not be easy to accomplish in practice.

- On the other hand, much of the data needed to get full advantage of the time-based micro approach simply did not then exist – with data from sample surveys, in particular, being fragmentary and often conflicting. But even here, this seeming weakness was really a backhand complement. The time-based micro system could work with whatever data was available; it did not require full-scale implementation to be analytically useful.

2.3 Looking to the future – from the perspective of 1980

Ruggles, in his concluding remarks, identified the following characteristics that should be sought in future work on social accounts:

- The system should be general purpose in nature – serving a broad spectrum of demographic, social and economic research, and creating a comprehensive common data base.

- It should integrate micro and macro data. The macro accounts should be conceptually derivable from aggregations of micro accounts and should provide a framework for the whole system.

- The system should not require the estimation or recording of trivia that arise purely because of accounting design.

- The system should be open-ended, so that analysts interested in specific problems or particular kinds of data can build on the existing system without having to begin anew.

These principles, in turn, led to the following R&D agenda:

1. More work is still needed on conceptual problems of integrating micro and macro data. In addition to micro accounts for households, firms and governments, we may need separate accounts for other reporting units including spatial ones (neighbourhoods, environments etc.).

2. Work is needed on techniques for matching and merging data from a variety of different data sets.

3. Attention needs to be placed on resolving inconsistencies in the classifications and definitions used in different parts of the system.

4. New kinds of information will need to be collected. In particular, time use information and subjective indicators should be given increased attention.

5. There is a need to establish the principles to be used in generating general purpose micro data sets, and to align them with the constructs in the macro accounts.[3]

6. The resulting accounts need to be tested in a wide variety of types of analysis.

These principles and the R&D projects were not followed up under the heading of 'social accounting' per se, which fell out of favour shortly after the workshop. However, they remain relevant and have been followed up in other contexts. Their influence can be clearly seen in the current Olivia framework, the subject of Section 4.

2.4 From the 80s to today: exploiting the revolution in information technology

The explicit social accounting schemes discussed in the golden age never were implemented and look very dated today. Even the term 'social accounting' in the sense used in this paper has largely disappeared[4]. Huge technological advances have meant that we can shift to a more comprehensive approach – collecting as much data as possible that might be of potential use and storing/retrieving that data at the micro level, i.e., at the level of the individual (or firm, etc.) being observed. Consultations among many users with different needs and theories, as well as common sense, respondent burden and resource limitations will determine what will be collected – not theoretical hypotheses.

The new approach also reflects the potentially huge increases in the amount of data that can by 'collected' at very little cost (and with no response burden) through increased exploitation of existing sources such as admin data, and through combining data from multiple sources in order to create synthetic data.

In other words, technology is allowing a democratization of the statistical system. Official statistics can be used by many more people for many more purposes, without being filtered through the lens of pre-existing academic theories or restricted to pre-tabulated files. That will,

---

[3] He is likely referring here to the construction of micro-data sets based on synthetic people (avoiding privacy issues), with aggregations matching control totals taken from macro analysis.

[4] 'Social accounting' has a completely different meaning in current usage. It now refers to the process of reporting on the environmental and social effects of economic actions.

of course, put a premium on finding quite new ways of allowing users to ensure that the data is meaningful and of high quality.

Simplification is still essential; no one could find meaning in the vast amount of potential data that will be available without some kind of theory, or lens, that tells us what is most likely to be important. However, this simplification will take place, not at the stage of data collection, and but rather at the later stage of data analysis, using powerful analytic tools that have only recently become available.

...Wolfson's criteria

By the mid-1990s, a shift away from social accounting approaches to microdata frameworks was largely complete – at least among the small number of world experts who addressed the topic. One of these experts is Michael Wolfson from Statistics Canada who set out the criteria for a modern framework in 1995. His approach placed an emphasis on the development of micro-simulation tools and operationalized some aspirations expressed by the earlier generation of framework builders, especially Stone, by applying computing power that simply did not exist a couple of decades earlier. Wolfson's criteria for a framework are shown Figure 1. The points that he listed some 15 years ago still very much represent current thinking.

Stiglitz Commission in 2008

Big picture thinking about social data returned, at least a little, in 2008 when the President of France commissioned Joseph Stiglitz and Amartya Sen to organize a commission to report on measures of economic performance and social progress. It was this report that created a wave of interest in approaches to the social data, including at the OECD.

The basic goal of the Stiglitz commission was to see how one could move beyond GDP as a measure of progress. The commission dealt with possibilities associated with the System of National Accounts itself, with other measures of the quality of life, and with measures of sustainability. The Stiglitz report was, however, not about social accounting as discussed in this paper. Rather it was about social indicators – and the emphasis here was less on big theoretical systems of indicators, but on specific questions and specific approaches.

We return briefly to the Stiglitz report in Section 4.

Figure 1. Wolfson in 1995 on the key elements of a framework for socioeconomic statistics

At any point in time, the population is best represented by a sample of individuals, each of whom is characterized by a set of attributes and relationships.

- Attributes include income, educational attainment, consumption, various aspects of health status, and time use patterns of activity.
- Relationships include conventional kinship ties as well as cohabitation (i.e. in database or graph-theoretic terms, such relationships can be represented by various kinds of pointers to other individuals – each of whom is also in the database).
- Relationships also include interactions with the major institutions of society – school, work and government programs. These contacts, relationships or transactions between individuals and major institutions can also be considered part of the set of individual attributes. They can take the form of pointers to descriptions of the institutions – schools, workplaces, and government programs – with which the individuals were interacting.
- This individual database can then easily be viewed as comprised of a hierarchy of various types of units – e.g. individuals, nuclear families, extended families, and households.
- Each unit (individual, family or household) can be described by any one of a number of summary attributes such as disposable income, leisure time, or self-reported satisfaction.
- Measures of variety can then be defined by summary statistics over this multivariate joint distribution of units (e.g. Gini coefficients, quantile shares).
- Over time, the population is best represented by a series of individual biographies, the equivalent of a broad and deep longitudinal panel survey.
- Given this longitudinal representation, a coherent family of summary indicators can be constructed from generalizations of the notion of life expectancy – including partitions of life expectancy into cumulative sojourn times in various life states.

In essence, this socio-economic framework would contain a complete longitudinal micro data sample, a microcosm of the actual population and its relationships to major social and economic institutions.

From this microcosm, a wide variety of statistical indicators could be readily constructed – effectively with no more effort than pressing the ubiquitous <Enter> on a computer keyboard to launch the appropriate software algorithm and have it pass through the microcosm data.

By construction, all such summary indicators would be coherent because they would be derived from the identical underlying micro data base. The summary indicators would not obscure the population's variety and heterogeneity, because the underlying micro data base would always be open (at the click of a mouse button, say, in terms of contemporary informatics functionality) for detailed inspection.

The main question is from where would this microcosm come? For the very practical reasons of cost, respondent burden, and concerns for individuals' privacy, it could not come from an omnibus longitudinal household survey. Moreover, there is not time to wait half a century or more for such a longitudinal survey to be substantially completed, by which time many things will have likely changed dramatically. The unavoidable conclusion is that the microcosm will have to be synthesized.

Such synthesis would be an extension of the synthesis of a population cohort already implicit in indicators such as life expectancy. It would differ methodologically, because the semi-aggregate or cell-based approach inherent in the underlying life table is incompatible with explicit micro data foundations. Instead, microsimulation is required.

Implications for data collection systems in a national statistical agency

- …These implications may not be that costly (relative to primary data collection costs), and most are relatively straightforward:
- Data collection processes cannot exist as "stovepipe" systems, in isolation from one another.
- One kind of coordination across data collection processes is use of common concepts and Definitions (e.g. identical definitions and methods for eliciting educational attainment);
- The other kind of coordination is assuring appropriate overlap in content – basically to anticipate the need for synthetic statistical matching (or equivalent methodologies).
- Microanalytic uses of raw data are far more demanding of data quality than aggregative uses.
- In effect, this means that data collection systems must be jointly planned, and that micro level data quality standards must be more stringent.

Michael C. Wolfson, *Socio-Economic Statistics And Public Policy: A New Role For Microsimulation Modeling*
ISBN: 0-662-21734-9. A Paper Prepared for the 50th Session of the International Statistical Institute, Beijing, 21-29 August 1995

The importance of the shift

The shift from golden age 'social accounting' to a descriptive 'integrating framework for micro data' represents a fundamental break in social thinking:

From	To
A world where the goal was to use theories of society in order to determine what is important to measure.	A world where data priorities are based on open consultations with many users, reflecting their practical needs and their multiple ways of seeing the world.
Frameworks which were conceptually interesting but partial and fragile – since we really do not have an agreed theory of society that can act as the equivalent of the System of National Accounts.	A new 'integrating framework' that is boring but stable, being basically descriptive in character. Its main purpose is to describe, and enable the reconciliation of, data sets that deliberately overlap and duplicate each other.
A world where analysis was based on point-in-time cross-sections – including time-series based on those cross-sections, and transitions across successive cross-sectional readings.	An analytic world that can deal not only with cross-sections but also that can deal with dynamic, longitudinal data that explores how things change over time at the micro level of individuals.
Analysis that relies on pre-tabulated, pre-packaged data.	Analysis that is based directly on the microdata and that allows the creation of new and missing data using tools of imputation and reconciliation.
An explicit approach to simplifying theories in both analysis and data planning.	A descriptive framework for use in data planning, with flexibility for user choice in applying different theoretical frameworks at the level of analysis.

While the new approach does not require any single theory of society, such theories may well be implicit in the views of the people who are consulted about their priorities for new data collection[5]. Rather, what the framework does is describe and systematize things at the level of the micro data itself, thereby allowing the data to be used in multiple applications, including applications that are based on quite different conceptual frameworks. It is much easier to gain agreement on priorities for creating particular pieces of micro data than it is to find consensus on big theories of society.

Section 4 will return to the most complete example of a contemporary 'social accounting' framework that has been developed to date. It is intended to provide a useful language for planning social statistics, especially in Canada.

---

[5] In principle, one could argue that some theoretical structure is at play in the choice of people or groups to consult, or in the importance we attach to the views of different groups when these are at variance. However, we are suggesting in Section 5 that consultations be broad and open. And experience has suggested that there is large overlap in the kinds of data that are needed, when these are cast at the micro-level.

Do explicit accounting type frameworks still have role to play in the development of new data?

The shift away from grand theories as the basis for planning our data collections systems does not mean that accounting-type principles will not be used within parts of the system – both for analysis and for reconciliation. As we shall see, accounting frameworks will have a central role in the future.

Indeed, the new warehousing approach to collecting, storing and accessing social data will require a contemporary version of a 'social accounting framework'. These new techniques will be unfamiliar to many readers. Section 3 will attempt to make things clearer.

3 – THE FUTURE STRUCTURE OF SOCIAL DATA

An understanding of the future role of social accounting, and of integrating frameworks more generally, can only be understood in light of the coming transformation in the way in which we plan, structure and use social statistics.

The social statistics system of the future will consist of:

- Data inputs, structured in a manner that is described in Section 3.1.

- As set of integrated processes for transforming these data inputs into useful information and knowledge in the hands of users. The integrating frameworks are described in Section 3.2.

- Information outputs, which involves a radical change in the uses of official social statistics, as discussed in Section 5.

Section 3.3 returns to the present and discusses the role of frameworks in developing the new approach.

3.1 The structure of the data

We are moving towards a world where social data will not be structured according to collection instruments, such as surveys or censuses or administrative files. Rather there will be a huge set of integrated but overlapping, duplicative data bases.

The honeycomb analogy

The easiest place to start is by imagining a single integrated social data base, one that contains all social data of conceivable interest. Think of an analogy with a gigantic honeycomb consisting of billions of cells. Bees have already filled some of these cells with honey and could potentially fill many more. Still other cells (i.e., potential holders of honey) will always remain empty. In the social data world, the corresponding cells are empty 'data holders' that we could potentially fill with tiny bits of microdata. The whole system can be thought of as a warehouse of these data-holding cells.

In fact, when the system is mature, many years in the future, there are likely to be an integrated set of overlapping honeycombs/warehouses – with one holding data about individual, another holding data about firms, and still others holding data about government programs and other units of analysis.

In the world of official social statistics, the most important of the warehouses relates to data about individuals. Priority needs to be placed on its development and it will be the main subject of this paper. However, the same principles will apply to all.

In the individual warehouse, the cells/data-holders would be structured into main clusters and within that, sub-clusters:

- There are over 30 million main clusters of cells (corresponding to the number of people in Canada).

- Within each cluster there is a secondary structure of some 35,000 sub-clusters of cells (corresponding to the number of days in a person's life). That means there would be something like a billion sub-clusters.

- There could be thousands of cells within of these sub-clusters to hold the actual data.

Following are some examples of data that could be potentially stored in these empty data holders:

- Data on the characteristics of the individual on that day (age, sex, ethnicity, language status, self-perceived well-being, hundreds of bio-medical markers, and anything else we might want to know about the individual from a social perspective).

- Data on what the individual did during that day at home, at school, at work, in stores and in a range of social networks. The activities would include interactions with other people, with institutions and other groups (including government programs), with those interactions being defined in terms of resources flows such as monetary transactions and non-monetary flows of goods, services (including care-giving of family members and child-raising) and information. The data would capture the duration of such activities and where they took place spatially.

- Data about the characteristics of the individual's dwelling, which provides links to data about the towns, labour markets, macro-economies and eco-systems in which they live.

In the honeycomb analogy, there is only one huge warehouse with billions of data holding cells. In the real world, with advances in information processing technology, such gigantic data warehouses[6] are becoming a real possibility and it might well prove to be useful to have a complete warehouse for purposes of backroom development and deep research.

However, for most everyday uses in the more immediate future, we will need to produce more specific data bases tailored to the application at hand – and to do so in a simple, user-friendly manner. The tools for doing this are discussed below.

From honeycombs to social accounting: the power of reconciling data overlaps

In the honeycomb analogy, there is only one holding cell for each item of data. In the actual social data base of the future, there will redundancy. Conceptually similar data will, in principle, be located in multiple holding locations in the data base. Social 'accounting' in our usage refers to the techniques for reconciling conceptually equivalent data that are located in holding cells

---

[6] We are imagining a huge warehouse with cells that can hold comprehensive data for every day of the life of the individual, with many of the daily cells being filled with imputed data that was actually collected on a weekly, monthly, annually or periodic basis. It may well be that this is technically inefficient and that we might, in fact, have a separate but linked warehouse for the limited data that does actually describe data on activities within a day. That is a separate technical discussion. For initial planning purposes, it less cumbersome to think of a single, huge warehouse from which any number of user-friendly bases can be created.

in different parts of the overall data base. It includes techniques such as requiring data to add up to exogenous control totals – and the use of double-entry bookkeeping along the lines of the national accounts.

Many such overlaps, redundancies and exogenous controls will deliberately be built into the system where they will perform important functions that will be described later. Examples are shown in Figure 2.

The goal is to develop a database that allows analysis of the inter-relationships among the data. This analysis of relationships is important for three reasons:

- Efficient data capture, with reduced cost and eventual reductions in respondent burden.

- Data integration and quality control.

- Creation of new data – without new data collection, through imputation and simulation.

...Efficiency

The data warehousing approach suggested here reflects a fundamental principle of data management in the information age. It is that *data is entered into the system only once and, once captured and verified for quality, is then used freely in multiple applications*.

Figure 3 shows an example of efficiency that applies in the constructions of the social data base itself. One item of data can be used to fill multiple data holding cells.

A key aspect of efficiency is that *the data warehouse system works no matter how many of the data holding cells are actually filled*. That is, there is a need to collect only data that is useful substantively. There is no need for collecting data whose only purpose is to feed the framework. Again this benefit was clearly seen in the golden age of social accounting discussed in the last section.

In other words, the system will work with only the data that is presently collected. However, the addition of every new data set, particularly ones that complement and partially overlap existing data, will greatly increase the power of the data base.

Reduced response burden is another efficiency gain. As noted, there is a deliberate intent to collect similar, overlapping data in order to produce new insights and to control quality. This is desirable, and essential. However, today's system often collects the same data in different vehicles for reasons related simply to the inefficiencies in the design of the collection instruments and the lack of capacity to share data among instruments – or even to share data within the same collection instrument at different points in time. In the new world of data warehousing, this kind of duplication will no longer be needed, with potential major savings in response burden and cost.

Figure 2. Examples of deliberate overlaps, duplication and exogenous control totals

Stocks and flows: Separate data holding cells will exist for data on both stocks and flows. The two are	**Comparably overlapping data about individuals and government programs.** In exactly the same way,

completely reconcilable, conceptually.

- For example, if we take one's stock of financial assets in one period of time and add to it net financial flows (income minus expenditures), the result equals the stock of financial assets at the next point in time.
- Similarly, our human capital (measured in terms of skills) at one period of time is the same that at the next period of time, if one adds the skills acquired in the intervening period and subtracts the skills lost during that same period.
- As will be seen in the next section, the stock and flow notions can be built into most aspects social data – even including data on values and perceived well-being.

Micro and macro data: Some data will be collected on an aggregated basis – say data from schools about average class sizes. Some will be micro data – say about the individual students. There will be many cases where sets of micro data, when aggregated, will be equal to data collected at the level of institutions. This point is expanded in the bullets below.

Individuals and institutions: As noted, in a fully mature system, there will be separate data bases for families, firms, associations, and other social groups. These will overlap the database that holds individual data.

- That is, while some characteristics of organizations (such as sales or profits in the case of firms) can only be described at the level of the institution taken as a whole, other elements are the simple the aggregation of the characteristics of the individuals that comprise those groupings (such as its employees) – with the data about the latter conceptually overlapping the data in the individual database.
- As well as a conceptual overlap, there are often overlapping sources of actual data. For example, data about various aspects of a person's job are provided by the individuals themselves (e.g., from the Census or Labour Force Survey), and by the employers (e.g., from the survey of Employment Payrolls and Hours – and are also provided to the tax system.
- The social data base of the future will have separate data-holding cells for all these overlapping sources – as well as the means for showing how they can be reconciled.

government programs can be described, in part, by aggregating up the characteristics of program beneficiaries and contributors (obtained through individual information on the administrative records of the program – or from evaluative surveys). In part, however, programs can be described only at the level of the program as a whole (such as measures related to the efficiency of delivery or the extent to which program objectives were being met).

Overlaps between individual and spatial data. The same principle applies to spatial analysis, but in an even richer manner.

- The characteristics of a neighbourhood, or city or province or country can be determined, in part, by adding up the characteristics of the individuals who live in those spaces.
- Still other spatial characteristics can come from aggregating up the characteristics of the institutions that comprise those spaces (e.g., economic data from firms, data on patients from doctors and hospital).
- Still others (such as crime rates, pollution, and accessibility of services) can only be determined at the level of the geographic area taken as a whole.

National Accounts overlap. As will be seen in Section 4, our proposed framework includes a way of capturing economic transactions (and approaches to economic stocks and flows) that is identical to that used in the national accounts – enabling full linkages to the double entry bookkeeping tools of reconciliation and analysis that are the strength of the modern accounts.

Time-based accounting: An even more powerful tool of reconciliation and analysis is time-based accounting where activities must add to the 24 hours a day. Recall that this approach, because it is based on micro data, was seen in the golden age as having even more potential than national accounting tools. The framework discussed in Section 4 captures all of the benefits of 24-hour accounting – and adds to them by also incorporating life-course analysis where the total of all life trajectories must add up to an individual's expected life span.

Instrument-based overlaps: often arise as a result of the design of data collection instruments, where it is often necessary or efficient to collect identical or overlapping information. For example, basic demographic information is collected in the census, in many surveys and in many administrative files. The social data base of the future will have holding cells for all of these.

Figure 3. Collected data can automatically fill many data holding cells

A bee fills the empty cells in a honeycomb with honey one a time. However, statisticians will often be able to automatically fill multiple cells in the new social data warehouse using the same data.

- For example, we usually need to fill the empty gender holding cells with information only once in a person's lifetime (i.e., filling some 35,000 cells at

In time. His or her marital, disability, employment or neighbourhood status in the period between these events can be filled in automatically.

- Equally, data about the state of the national economy at a particular point in history can be used to fill millions of cells at one pass (since it applies to everyone living in Canada at that time), while data about the adequacy of

---

- one go, without allowing for sex change operations).
- For many variables, we need only record events where changes in status occurred. The cells in the intervening period between these events can be filled in automatically.
  - For example, a person got married (or employed, or became disabled or moved to a different house) at one point in time, and then got divorced (or lost his job, etc.) at later point

the social services or the pollution in a particular urban neighbourhood could be used to fill all the appropriate cells for people living in the neighbourhood.

- Often data is only available on a monthly, annual or periodic basis. For example, an annual total would be automatically applied to the 365 cells representing the days in that year. (This could, of course, result in a misrepresentation of the reality in any one actual day, a problem that is resolved by the use of the framework's maps, as will be described in Section 3.2.)

…Integration and quality control

The new social data warehouse requires that the quality of the individual data items be high and, especially, that the inter-relationships among the separate data items be known. That is, reconciliation and integration are built right into the system, and not left as an afterthought or an annoyance, as is often now the case.

For example, we now collect conceptually similar data from about numbers of employees and their pay from the Labour Force Survey, the Census, establishment surveys (SEPH) and administrative records (especially tax files). Income data is similarly collected from surveys and tax records. This is not improper duplication since we tend to ask about different aspects of employment in the different surveys. (As well, the use of tax records reduces the need for surveys and associated costs and response burden). However, in those data that are intended to measure the same thing, there will invariably be differences because of variations in the wording of questions, sampling and other non-sampling errors, etc.

These variations, if they are not large, have traditionally been seen as an annoyance, not a major problem, since most analysis today is based on data that comes from a single collection instrument. Typically reconciliation studies are undertaken on a periodic basis that attempt to explain and, ideally, quantify the reasons for differences. However, these seldom produce results that allow users to reconcile the numbers on an ongoing basis, or to know the 'correct' numbers that should be used in different applications[7].

Once the survey-by-survey approach is replaced by a comprehensive data base approach, these seemingly minor differences can cause major problems for integrated analysis. Issues of this sort have arisen in recent years in application such as micro-simulation analysis which does draw on information from multiple sources and where decisions need to be taken on what is the right source for that particular use. Indeed, modelling is increasingly being seen as tool for controlling the quality of the whole data system, as well as useful analytic tool in its own right.

The solution in the new system, once it is mature, will be that a *data item must be in a format that identifies its quality, and the nature of its relationship with other data in the system, at the time when it first enters the system.* That is, quality and reconcilability must be built in. The necessary tools to accomplish this are discussed below.

---

[7] Often, there is one big primary data collection source, such as the census, which is taken as providing the 'correct' answer because of its scope and internal consistency.

…Creation of new data, without cost or response burden (magic at play?)

There always will be much unevenness in the data that will be actually collected to fill the data holding cells in the new social statistics data base. That is, the data holding cells are comprehensive and include all the data that we would ideally like to have – detailed data on both flows and stocks, highly detailed information about neighbourhoods, firms and families – collected continuously. Response burden and costs will mean that the number of empty data holding cells will, of course, always be many thousands of times larger than the number that are filled.

In the new way of doing business, we can use the inter-relationships among the partial, but overlapping, data that is collected in order create new or missing data.

- For missing data, resulting from non-response or obvious errors in answering questions, it has always been possible to use imputation techniques to fill many of the gaps. In recent decades, there has been explosion of new data created by applying new technologies to clean up messy administrative data sources. In the new social data base, the possibilities for imputing missing data increases by many orders of magnitude; there are many more sources of related and overlapping data on which to create the missing data.

- Of much greater importance is the use of new computing power which can examine the inter-relationships among similar or overlapping data sets in order to create new data that was not there in any of the original collection instruments.

Following are some current examples of new data that has been created. These are major breakthroughs, but still only hint at the potential power of the new way of doing business.

- In recent years, Geographic Information Systems have provided vastly richer data bases of local area data, using new computing techniques that can draw on different data sources.

- In Canada, data on from the Adult Literacy sample survey has recently been imputed onto the individual records of the 2006 long-form census, to provide a rich source of completely new data on skills by industry, occupation and geographic areas. It has even been possible to impute mismatches in the supply and demand for skills by developing data on employers' demand for skills (using the fine occupational coding on the census micro-records along with exogenous essential skills profiles developed by HRSDC).

- Microsimulation tools, particular the LifePath model, have perhaps gone furthest in creating highly detailed life histories of (synthetic) Canadians, again drawing on inter-related data from multiple data sources.

The potential of powerful new computing used in conjunction with large data bases of inter-related data is extraordinarily large. As with any statistical technology (the introduction of sampling technologies is a good analogy), there are limits to, and possible errors associated with, the use of new technology for this purpose. However, given a realistic period of R&D,

the payoffs in terms of richer data, at minimal cost and no response burden seem extraordinarily high.

3.2 – The new social data warehouse requires an organizing framework and new tools for entering and accessing data

The need for an integrating, descriptive framework will be obvious from the preceding section:

- The new social data warehouse will require data to be of known quality and to be reconcilable with related data – topics that can be highly technical in nature. Planning for the new warehouse will require detailed, comparable descriptions of possible future data as well of existing data.

- The warehouse will be fed from a large number of quite disparate sources. It will be planned and used by many people with quite different needs and backgrounds, including non-technical backgrounds. Analysis will be undertaken by people from many different disciplines with different ways of understanding society.

The only way in which the two bullets above can be reconciled is to have excellent maps for locating and understanding the data – and excellent, user-friendly tools to adapt the huge underlying database to the needs of specific users.

We lump the framework and tools together since both are necessary and are closely inter-related in the new world of data creation and analysis. Much of the new data will have been created using analytic tools that combine existing data from multiple sources. Many applications will require the use of analytic tools to translate the billions of possible data elements in the warehouse into something that is meaningful.

In other words, two functions that were once separable will need to be managed in an integrated fashion, namely:

- Planning to determine the contents of the holding cells in the data warehouse.

- Planning of the analytic tools (such as micro-simulation tools or real-time internet access to the data) needed to create/acquire the data and to allow user-friendly access to the data.

The framework

The accounting framework consists of maps to the social data warehouse:

- One map describes the *data holding cells* in the warehouse (including the very large number of empty cells). It describes the precise definition of the desired contents of the cell as well as how it relates to other cells – i.e., how the whole system is integrated. It describes what the ideal data elements would be.

- The second map *describes the data in the filled cells* – its source (e.g., whether it came from surveys, admin data, qualitative studies or were created synthetically), its quality (e.g., sampling errors and a range of non-sampling errors including any discrepancies between the ideal concept associated with the data holder cell and the data chosen to fill the cell), and other practical descriptors.

It is the first of these maps that is the successor to the social accounting frameworks described in Section 2. The Olivia framework, described in Section 4, is this kind of map.

...Does the first map need to be explicit?

It has sometimes been argued that we do not need an explicit social accounting framework, the first of the two types of map. All that is really needed, it is argued, is the second type of map – that provides definitions for the data that is actually in the warehouse.

Such a view made some sense when data were organized around specific collection instruments. It even makes some sense when one is talking about a single model, such as microsimulation model, that is fed from multiple data sources; such models are developed by a small, unified group of experts who can talk with each to resolve issues about the relationship among data from different sources. Indeed, the Wolfson principles in Figure 1, in the last section, were intended to support a microsimulation model, not a freestanding conceptual framework.

However, an explicit map of the holding cells (filled and empty) and of their inter-relationships is essential for clear communications in planning for the new social database, especially longer-run planning.

- It would provide a needed common language for the large numbers of players that need to be consulted in planning the future of the database, including setting priorities for the collection or creation of data that currently do not exist (and hence are not included in Type Two maps). These players – including ordinary citizens, those who finance surveys, as well as analysts and researchers from many disciplines – often see things in quite different ways and use different vocabularies.

- It would greatly facilitate the tactical task of determining which data should be collected using which collection vehicle, especially in maximizing the potential of the overlaps and control totals discussed above. For example, resources limits and response burden will mean that only some data on some flows and on some stocks will be collected – and the selected data can come from different collection instruments. Decisions are needed on the mix that will result in the most useful information when the database is looked at as a whole.

The associated tools

The tools in question are those that are needed to enter data into the data base – and to provide user-friendly access.

In terms of entering data, the data must be validated, edited, documented and its relationship to other data must be defined. In some cases, where new data is being simulated or imputed, the entry tools are complex mathematical models.

In terms of accessing data, arrangements will look very different from those in play today. An analogy would be the evolution of Statistics Canada's web site so that it would replace all existing data publications and reports, answer all specific data requests in real time (including

tables, regression results, etc.) and would create tailor-made micro data sets to meet the needs of specific users. Figure 4 provides some examples.

Figure 4. Examples of the kind of outputs that will need to be produced in real time

- Data that answers questions posed by many kinds of analytic questions, both simple and sophisticated, including
  - Time series and social indicators, (i.e., cross-sectional variables at selected points in time.)
  - Distributional analysis, often for different units of analysis such as individuals, census families, extended families, households, social networks, firms etc.
  - Longitudinal data for selected variables, for both individuals and cohorts.
- Data that will provide real-time answers to practical questions posed be individual citizens over the internet – for example, about the kind of employment decisions that are likely to work best for individual citizens. Practical examples are found in Section 5. The argument there expands on this new, and primary, use of social statistics.
- Data to create traditional publications and reports, as well as descriptive information on internet sites.

- Tailor-made micro-data sets for users involved in more sophisticated research related to topics such as:
  - Different geographic spaces – neighbourhoods, local labour markets, provinces, etc.
  - The characteristics and dynamics of different social groupings or institutions, such as firms (the source of most economic data), schools, extended families, community groups, doctors' offices, informal networks, etc.
  - The inputs, processes and outcomes of specific government programs.
- Databases to support statistical planning and development, including:
  - Data that was created from specific collection instruments, such as the 2011 census or the longitudinal tax file (i.e., the currently normal structure – data bases that will, in the future, be mainly of interest for only for those analysts working on the development of those vehicles.)
  - Unfilled cells, for data planning purposes.

… real and synthetic people

Privacy considerations place an obvious limitation in producing tailor-made data for users and in allowing access to micro data. Data about real people would clearly not be accessible via an internet site. However, when the system is mature, most analysis will be based on data about synthetic people who have same characteristics as real Canadians. Using synthetic people eliminates the privacy issue and allows far richer micro-analysis – along the line now possible with dynamic micro-simulation models such as LifePaths.

Approaches involving synthetic people will almost certainly grow quickly in importance in the coming years, obviously so for data bases containing new data created by synthetic tools. As will be described in Section 5, the most important uses of social data in the future will be by ordinary citizens and groups using the data to help make practical life decisions. These will not require access to data about real individuals.

However, there will always be a need to maintain an additional capacity based on the original records to support quality control and more specialised analysis.

3.3 Moving towards the mature system of the future

To this point, the section has described a data warehouse approach that might exist when the system becomes mature – some years in the future. There will be many challenges in shifting to this quite different way of doing business.

The extent of the remaining challenges should not be underestimated. For the most part, our social data is still created, structured and used in much the same way as it was before computers were introduced. The world of official statistics has yet to fully enter the information age.

Many participants in the system are not even fully aware of the radical potential for a transformation in the amount of potentially useful knowledge that can become available, with little cost implications and reduced respondent burden. As well, in some of those areas where progress has been made, we are still stuck in the development stage – with tools that are still not user-friendly and that sometimes deliver less than they promise.

Figure 5 provides two analogies may help situate our present state.

The analogies in Figure 5 suggest that the shift to the new ways of planning and organizing social statistics will not be simple but, once it starts, change will occur quickly and dramatically. Three factors will likely be needed before the period of rapid change begins

- There will be a need to get agreement that the change is needed and agreement on who is responsible for driving that change.

- There will need to be a more broadly-based common understanding of the structure and potential of the new approach. We lack a shared vocabulary to describe what we would like to build and how work together to get there.

- The creation of some dramatic success stories that would demystify, and demonstrate the power of, the new approach.

With respect to the first point, a big step was taken with the decision inside Statistics Canada and HRSDC to shift social data planning to the level integrated 'data platforms' (groups of related collection instruments), i.e., not be based on separate collection instruments. The present contract and the related activities in the two departments are important follow-up steps.

Figure 5. What may lie ahead: two analogies

The introduction of personal computers In one analogy, we might compare our present state of statistical warehousing to the 1970s when computers had reached the point where their potential to transform many areas of life could be foreseen, but where actual computers were huge, expensive black boxes that could not be used in most operational systems within government –	in the post-war period, particularly with the introduction of the Labour Force Survey (LFS) to give a Canadian example. However, it was not until the LFS had a proven track record of several decades before sample surveys became common place in official statistics – and when competing data series drawn from (so-called 'real') administrative records were dropped as alternative

except to carry out pre-computer processes (much) faster and (somewhat) cheaper.

Especially in the past decade, the arrival of powerful PCs, user-friendly software and the internet changed all that. The transformations once dreamed of are taking place at an accelerating pace.

The introduction of sampling in official statistics

The other analogy refers to an even earlier time – the introduction of sampling in the world of official social statistics.

Modern surveys based on random sampling techniques had many historic predecessors, but only started playing a major role in official statistics

measures of unemployment.

Sampling represented a very different way of doing business, and much concern was initially felt, including by many official statisticians and analysts, about the quality of the data[*] they produced and about potential misuses.

The analogy is with the similar concerns that are often heard today about the use of simulated data and synthetic people. Both sampling and simulation require faith in the capability of complex, sophisticated mathematical tools to produce outcomes that can be seen to be almost magical and that cannot be independently verified by the researcher himself or herself – namely, the production of accurate information about a whole population based on only partial observations, often a quite small number of observations.

*The infamous example of the failure of even large samples to predict the results of the 1936 US election was still in people's minds

We suggest that the conceptual framework set out in Section 4 could well be an important step in providing the needed common language, referred to in the second of the bullets above.

With respect to bullet three, the citizen-based applications described in Section 5 could, similarly, provide the needed illustration of the power of the new approach.

4 – THE OLIVIA FRAMEWORK: A CURRENT APPROACH TO 'SOCIAL ACCOUNTING'

Work in Canada began several years ago on developing a new social accounting framework – a map of the holding cells of the new social data warehouse. It was called the Olivia framework[8] and, among other things, it provides a tool for helping answer the question before us, namely: what social data ought we to collect and how should it be structured. More generally, it holds the potential to be the new common language of policy-oriented social data and social analysis.

The section is organized as follows:

- Section 4.1 provides some background about the development of the framework.

- Section 4.2 examines the heart of the framework: integrated descriptions of individuals, institutions and of the flow of resources among individuals and institutions.

- Section 4.3 takes a closer look at resource flows and shows why they are at the centre of an integrated system of social statistics.

- Section 4.4 examines descriptors of individuals and institutions in more detail.

- Section 4.5 concludes.

4.1 The evolution of the Olivia framework

In 2003, the Policy Research Initiative began developing a descriptive framework to support policy-relevant social analysis, including supporting longer-term planning for data collection. The goal was simple but ambitious: to develop a consistent, quantifiable terminology for describing those aspects of society that were of potential interest for social policy purposes:

- The language could be used either in practical data analysis or in strategic planning exercises related to the collection and creation of new kinds of data.

- It could describe the data uses and requirement of individual citizens as well as the more familiar policy-related uses, a topic we will return to in Section 5.

- The language would encompass the concepts used in a range of social sciences – including but not restricted to economics.

- It could apply to both quantitative and qualitative analysis.

The framework provided an integrated, finely-grained, way of describing:

- Point-in-time, resource-flow perspectives, including the economic transactions that underlie traditional social policy analysis.

- Newer life-course perspectives, including transitions/trajectories analysis drawn from sociology and several other social sciences.

---

[8] The features of the framework were illustrated using the example of single hypothetical person, Olivia. Stéphanie Gaudet, a co-author of the original PRI version, invented Olivia as a way of explaining life-course concepts to analysts familiar only with traditional economic perspectives.

The framework was subsequently developed by Social Development Canada and HRSDC and Peter Hicks has added some further refinements to it[9]. Figure 6 is a brief summary of the framework as it had been developed up to 2007.

Figure 6. The contents of the Olivia framework as it existed in 2007

An information framework is a comprehensive set of consistently-defined words, and associated numbers, that describe the activities, needs and aspirations of individuals and institutions as they relate to society and the labour market. The prime focus is on the individual, with emphasis on the individual's relationships and spatial attachments, including ties to families, institutions and communities as well as participation in government programming and public institutions.

In its current form, the Olivia framework consists of four modules – each describing individuals and organizations from a different perspective, but using consistent, measurable concepts throughout.

Module One is about resource flows and is a simple extension of the traditional, point-in-time micro-economic analysis that is the basis of much of today's social policy analysis. It sets out consistent concepts for describing the interactions or transactions that take place among individuals and institutions. (In our language, an institution is defined very broadly to include government programs and informal social networks, as well as formal organisations such as firms and non-governmental bodies.)

These concepts encompass the monetary flows that are the basis of our economy, as well as non-monetary flows of goods, services, information and time. As well, Module One provides standard ways of describing the characteristics of both individuals and institutions. It is our starting point.

Module Two starts with the point-in-time resource flows and transactions of Module One, but uses them as the basis for a consistent approach for describing how people and social institutions change and develop over time. The module provides two quite different, but complementary lenses by which we can understand social change and development:

o Life-course analysis based on transitions in the main domains of life. Transitions are major changes or discontinuities in the resource flows of Module One.

This analysis allows us to describe the main compartments and changes of life.

o Stock and flow analysis that looks at, for example, financial capital, housing, human and social capital – and how these result from the point-in-time flows in Module One. We build up assets from flows at one stage of life, and convert them back into other flows at subsequent stages of life. This analysis allows us to describe the continuities that cross the various transitions and stages in the lives of people and institutions.

Module Three provides a set of concepts for describing the physical locations where individuals and social institutions are situated – neighbourhoods, cities, labour markets, provinces, the country as a whole, and international groupings such as OECD countries. This allows us to anchor people in real space and in real history. It allows us to examine community assets, natural resources and sustainable development.

And it also allows us to use the micro-analysis of Modules One and Two in conjunction with macro-level analysis – whether of the effects of business cycles, wars or climate change. Macro analysis applies to a geographic area such as a city, or province or the country as a whole. These macro characteristics can then be attributed to the individuals living in those spaces.

Module Four provides a consistent set of concepts that can be used to describe the purposes or goals of individuals, institutions and societies.

o For institutions, we use the familiar input-process-output-outcome model.

o For individuals, the purpose is well-being. We approach this through concepts related to values, satisfactions and expectations – with these concepts rooted in the finely-grained descriptions of the first three modules.

o For societies, purpose is shown through the development of social indicators.

The current version of the framework

Section 3 described the need for two kinds of maps to the new social data. The version of the Olivia framework presented here has been adapted (considerably in places) to provide the

---

[9] Prior to the present report, the most recent published version of the framework can be found at www.queensu.ca/sps/publications/working_papers/45-Hicks.pdf. It is an edited version of a paper that Hicks presented at an OECD seminar in Paris in 2007.

structure for the first of these two maps – a guide to the data in the holding cells of the new social data warehouse – both data that already exists and also data that we would like to collect or create.

4.2. The framework starts with individuals, institutions and resources flows

The heart of the framework lies in its description of the interactions and transactions among people and social groups. Suppose that we have a world with three individuals and three institutions (institutions are our name for social organizations). Figure 7 shows the basic flows of resources among them. In this simplified world, our language would comprise a set of consistent words and numbers that describe:

Figure 7. Basic resource flows model

- The characteristics of individuals 1, 2 and 3 and of institutions 4, 5 and 6 at any point in time.

- The resources – money, information, goods and services – that flow among the individuals and institutions at a point in time, as represented by the two-headed arrows in the diagram.
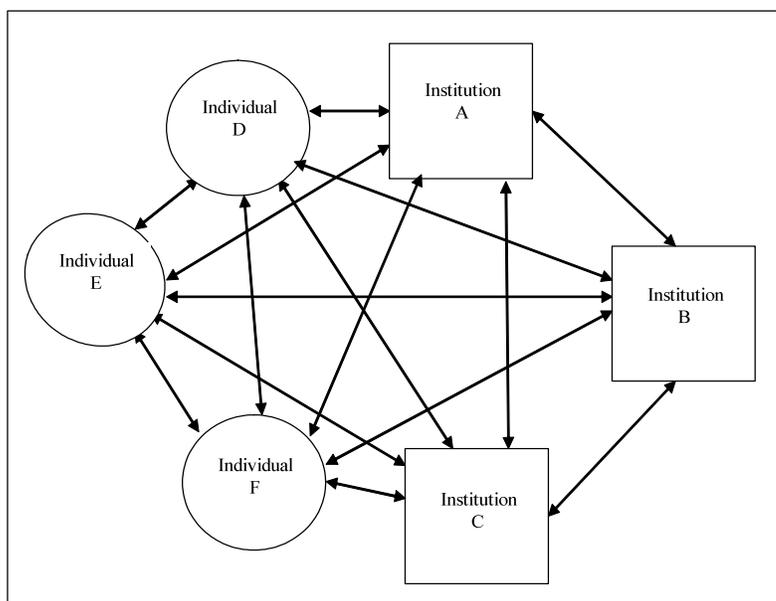
In this revised version of Olivia[10] four resource flows are identified:

- Money.

- Goods, defined broadly to include gifts, food cooked at home, as well as goods purchased in the market.

- Services, including the application of people's skills.

- Information[11] defined broadly to include reading, listening to music, chatting, watching TV and playing with children.

---

[10] In earlier versions of the framework, time was also considered as a flow. While this concept is used in some of the literature, time is best thought of as way of measuring the duration of flows and activities, not as a flow itself. As well, goods and services were combined in earlier versions of Olivia, but are broken out here for ease of explanation. For example, in the labour market there is a flow of services (the use of people's skills on the job) in return for money (wages).
[11] Information can be logically thought of as a subset of services. However, it is simpler and more consistent with ordinary usage to keep them separate. For example, we usually do not see chatting with neighbours or dinner-time discussion with one's children as 'services'. Publishing a book is a service, but reading a book is not usually thought of as consuming a service.

Flows are always two-way, as indicated in the double-headed arrows in Figure 7. However, they are not always equal in any one time period. In monetary transactions, 'equal' refers to the market value of the good or service being produced. For non-monetary transactions, it refers to the duration of the interaction – the number of minutes or hours spent. This is illustrated in Figure 8 where we use Olivia as example.

Subsection 4.3 will describe resource flows in more detail, while section 4.4 will describe how the framework describes stock, i.e., point-in-time descriptors of individuals and institutions.

4.3 Resource flows: the basis for statistical integration

4.3.1 Resource flows bring integration to the heart of the framework

The resource flows (money, goods, service, and information) are categorized as follows:

- Transactions, which are the market flows that are invariably monetary in character – buying and selling of goods and services, labour market and capital market transactions. Transactions are assumed to take place at a point in time. That is, they have no duration.

- Non-monetary interactions that take place in the family, with colleagues in the workplace, in community groups and social networks. These are almost always flows of services and information, but can also include goods that are made at home and given as gifts.

- Events[12] that change relationships among individuals and institutions – getting married or divorced, going to a new school or graduating, having a baby, having a spouse die, obtaining a job, being fired or retired, etc. Events, like transactions, are assumed to have no duration.

Non-monetary interactions are assumed to have duration. For example, providing services to an employer often involves spending eight hours a day. Providing house-work and child-care services at home with family might have a duration of perhaps 6 hours. The total duration of all non-monetary flows equals 24 hours a day. That is, all time is accounted for by non-monetary interactions. In this sense, a non-monetary interaction is similar to the concept of an 'activity' with a discrete beginning and end – a concept that is currently used, for example, in time use surveys.

Figure 8. Resource flows are always two-way, but not necessarily equal in any one time period

On any one day,	• *Olivia and her taxes and benefits* (flows of money). As
• *Olivia and her daughter*. Olivia gave her daughter Marie her allowance (flow of money), made her breakfast (flow of goods and services), read her a bedtime story (flow of information), and told her where she could find her winter coat (flow of information). Marie, on the other hand, told her mother about her day at school (flow of	a consequence of her day's work, Olivia also paid the Government of Canada the income taxes associated with the pay she received for that work – via a payroll deduction, with the actual calculation reflected on her pay stub every two weeks.
	o Her earnings were similarly used to pay for EI, CPP, and provincial income taxes. In return, she

---

[12] Events are technically not 'resource' flows, although they change the nature of subsequent flows. However, we wished to avoid the too cumbersome heading of 'resource flows and relationship-changing events'.

information). Marie, by definition, spent the same amount of time with her mother as her mother did with her. There was no flow of money or goods and services from Marie to her mother that day.

- *Olivia and her neighbour*. They commuted to work in Olivia's car, as they did every weekday (flow of services). There was an implicit transfer from Olivia to the neighbour in the form of the gas used, the depreciation on the car and Olivia's services as a driver. By mutual agreement, this was compensated for by the neighbour paying for the gas once a week.

- *Olivia and her employer*. Olivia provided J&C Insurance Operations with eight of hours of work (flow of services) and received $175.00 in pay in return (with the actual pay cheque of $1750 (175 x 10) being deposited in her bank every two weeks. Her day's work also contributed to building up her store of sick leave, annual leave, company pension and the like.

- *Olivia's investments and savings*. There is a one-way flow of money from Olivia to her RRSP savings account that day. This is in anticipation of a future flow of pension income later in life. More generally, the dollar value of her purchases that day was greater than the value of her earnings – thereby increasing her debt slightly (an increased flow of money out in anticipation of reduced flow out in the following week).

received a child tax credit (which would arrive in a lump sum in two months' time).

o As well, yesterday she made use of a wide range of government services, such as the road she was driving on, the water and sewage at her home. She heard about Statistics Canada's latest unemployment figures on her car radio (flow of information from the government to Olivia, via an intermediary institution).

- *Olivia at the community centre*. She pays a fee and, in return, spends 45 minutes on their exercise machines three mornings a week (flows of money and services). This activity is important to Olivia's social capital. She formed friendships there that extend into other dimensions of life. For example, a person in the group of friends who exercise together became unable to drive her car for health reasons and is being helped by the others. In the evening of the day in question, it was Olivia's turn to drive her to the grocery store (two hour flow of services).

- *Olivia alone*. After Marie goes to bed, Olivia watches TV by herself (a flow of services from the broadcaster, a flow of information to Olivia, with a dollar flow from Olivia to the cable provider.)

- *Olivia asleep*. In ordinary language, personal activities such as sleeping and eating are not seen as resource flows. However they are kept track of in the framework since they result in a flow back in the form of self-sustainability, health and well-being and because this allows us to create a useful social accounting tool – the allocation of time over a full 24 hour day.

These categories ensure broad integration:

- Transactions enable integration with economic statistics via the system of national accounts

- Non-monetary flows enable integration with time-based accounts

- Events enable integration with life-course data on life's main trajectories and transitions.

Using resource flows is not the only possible starting point for a micro-based framework for social statistics. Two alternatives were considered in the golden age as discussed in Section 2:

- Integration using market-based equivalents and demographic transitions.

- An activity approach where a series of activities are identified that account for the 24 hours of a day.

The resource flow approach of Olivia is similar the activity approach, but even stronger. As Figure 9 shows, it captures all the advantages of the older alternatives while being considerable more powerful.

Figure 9. Resource flows versus activities

The resource flow approach is similar to, and can be easily converted to, the activity approach used in the time-based social accounts discussed in Section 2 – and that is found in time use surveys. In both cases, activities (non-monetary flows) have durations that add up to 24 hours a day.

There are excellent reasons for thinking in terms of resource flows, rather than activities. For example, the resource flow model:

- Defines individuals in terms of their social context, in their dealings with others.
- Embeds the fundamental notion that flows are two-way, that things are always done for a purpose or have a consequence (a return flow) – but that this return flow might be of different kind and might be at a later time. This sets up analysis of both flows and stocks – of investments, of mutual obligations, of social capital.

- Invites comparative analysis of monetary transactions and non-monetary interactions.
- Reconciles two types of dynamic analysis: stock and flow analysis, and transitions/trajectories analysis.

In a few areas, however, the activity approach is closer to everyday language.

For example, in the Olivia system, it is necessary to include a number of passive activities, such as sleep whose function is internal to the individual, as 'resource flows' – even though nothing obvious is flowing. This is needed because such activities are significant (personal biological maintenance in this case) and to ensure that the total of all non-monetary resource flows adds up to 24 hours a day.

However, these problems are minor compared with the powerful strengths of the resource flow concept.

4.3.2 Describing resource flows

Resource flows are described and classified in the usual, obvious ways:

- The individuals and institutions involved in the flow are identified, as is the type of flow (money, goods, services, information).

- Dollar values associated with transactions are identified.

- Characteristics of the relationship-changing events are identified – e.g., whether a formal or common law marriage, any credentials associated with school graduation, reason for lay-off from a job, etc.

- As with time use surveys, there will be information about the timing and sequencing of interactions, as well as their durations, in order to distinguish resource flows that overlap each in time (such as simultaneously listening to the radio, cooking dinner and helping children with homework – or business lunches where we both eat and make deals) – so that we can still add up to the needed 24 hours.

In addition, as will be discussed in the remainder of this sub-section, resource flows are categorized in the following ways:

- Section 4.3.3 will discuss descriptors of the intensity of resource flows. This is an important topic related to the measurement of human and social capital and to providing practical measures of human capabilities.

- Section 4.3.4 will discuss ways of describing the purpose of resource flows.

- Section 4.3.5 will discuss descriptors of events, which set up life course analysis.

- Section 4.3.6 will mention several technical topics where resource flow analysis needs to be supplemented in various ways to enable full integration, especially with pre-existing economic statistics.

4.3.3 The intensity of resource flows – understanding the creation of human capital and social capital

Human capital is a main subject of social analysis. It deals with investments in developing skills and good health, and the subsequent returns from those skills and from that good health.

- 'Skills' is shorthand for analysis of the full range of peoples' skills, knowledge, aptitudes and abilities. It includes how these were acquired both through experience and in more formal learning settings – and how they are lost through lack of use or a deterioration of one's mental and physical health. It also includes how they were used, with most analysis concentrating on labour market uses (i.e., providing services to employers in return for wages).

- Health refers to analysis based on the WHO positive definition of health – including the determinants of good health and how health status helps or constrains subsequent resource flows later in life.

Social capital has received much attention in the academic literature in recent years, particularly in the field of sociology, with Canadians playing an important international role[13]. However it is less developed in the area of official statistics, with no fully developed consensus on either its role in official government statistics or on best approaches to practical, routine measurement.

Social capital deals with interactions within networks and with the subsequent benefits that are realised from investments in those networks. These investments are usually in the form of flows of information, (whether informal chatting with neighbours, on-line social networks, or more formal 'what worked best for me' discussions among people with similar interests and problems). However they could also include services (e.g., shovelling snow of a sick neighbour or more formal volunteering) or money (e.g., charitable donations).

---

[13] The Policy Research Initiative, the Social Research and Demonstration Corporation and HRSDC (in its various past manifestations) have played important roles in exploring the uses of social capital concepts and data in policy applications. In the academic world, Barry Wellman at the University of Toronto has played a key role in areas such as network analysis including the roles of the internet, social relations and social structures. An anonymous critic of an earlier draft of this paper usefully pointed out the fruitful intersection of human capital and social capital approaches in a number of topics of policy interest, including ways in which new immigrants integrate into Canadian society.

The actual measures of the levels of human and social capital – and of human capabilities – are stock measures that will be discussed in Section 4.4.

In this sub-section we are concerned with the resource flows that create this stock of human and social capital – or that diminish it. As will described in Section 4.3.4, the Olivia framework supports traditional analysis which typically divides life into categories such as life in school, in work, in the family, in sickness, in retirement, etc. However, it also supports newer forms of analysis that allow us to see the interactions across these spheres of life. It does this by categorizing resource flows (regardless of whether they are work- or school-related, or family-related, or community-related), by their:

- *Learning / skills intensity* – i.e., the extent to which different resource flows require the use of skills and, as well, the potential for learning associated with those flows. For example:

  o In challenging jobs or difficult academic courses, for example, we might expect that there is both a high level of skills being applied and a high level of skills being learned.

  o In more passive learning (say, in watching a TV documentary or reading a business report), there might be only a moderate application of skills but considerable potential for acquisition – learning new things.

- *Healthiness intensity* – where we look at the extent to which various resource flows are likely to contribute to subsequent good health and the extent to which health affected the performance of the resource flow. Take, for example, the resource flow of providing services to one's employer in the course of a particular day.

  o In terms of the physical and mental health needed to do one's job, we would have information related to sickness or disabilities that reduced or constrained one's contribution. In other words, we would incorporate the growing body of literature related to measurement of limitations on one's capacity to perform Activities of Daily Living (ADL analysis).

  o On the other side of the coin (i.e., the potential of the job to affect health), we could have information on the amount of physical exercise that occurs in carrying out ones daily job, as well the stress that exists, or the existence of pollutants in the workplace.

- *Networking intensity* of interactions with other people. As with the two intensity measures above, this is simply place-holding phrase suggesting future directions for the routine inclusion in the data warehouse of statistics relating to social capital. However, even more so than in the cases above, the actual operational content has not yet been developed.

  o The basic idea is that not all social interactions are important in building significant social networks and useful social capital.

  o The goal would be to develop a standard way of identifying and describing those that do. The explosion of on-line social networks provides a powerful new impetus to get on with this kind of R&D.

4.3.4 Purposes – how day-to-day life is related to high order social objectives

The typology for describing the purposes of resource flows is based on the familiar systems model of inputs, processes, outputs and higher level outcomes (often a hierarchy of such outcomes).

For institutions, the output of each tiny resource flow fits into a complex hierarchy of outputs and outcomes. The outputs at a higher level of the system can be the intermediate level outcomes for those at lower levels of the system. The highest levels of outcomes for firms are things like profits or market share. For government programs they are the ultimate objectives, the reasons for creating the program in the first place. (The formal accountability system of the Government of Canada uses an activity structure based on this hierarchical approach.)

Higher level outcomes are often explicit, or readily inferred, for institutions. It is not that simple for individuals or for societies as a whole. There is no consensus on what constitutes individual or societal well-being. The resource flow approach, however, enables exploration of many different approaches:

- It replicates existing approaches to material well-being such as income and consumption levels for individuals and GDP for societies.

- If deemed important, one could add measures of the expressed satisfaction of individuals with the particular resource flow in question (or their views on constraints that were encountered) – resulting in, among other products, a range of quality of service indicators.

- It is possible to explore cause and effect relationship by comparing the flow data, and the expressed satisfaction with those flows (i.e., process well-being) in conjunction with the point-in-time readings of stocks, discussed in Section 4.4. These stock measures include readings on an individuals' overall sense of well-being, their trust in institutions, their felt constraints, their human and social capital, their capabilities etc.

- The framework opens up the possibility of subsequently linking the social and economic data to more sophisticated ways of understanding behaviour, motivation and personality that that flow out of the discipline of psychology[14]. There are also strong links to human development theory[15].

4.3.5 Supporting integrated analysis of life-course transitions and trajectories

As a reminder, the Olivia framework is based on micro data, both cross-sectional and longitudinal. It provides two consistent tools for supporting integrated analysis of the dynamic, changing aspects of life:

---

[14] The work of the Canadian psychologist, Brian R Little (Harvard, McGill, Carleton) on personal projects and personality seems particularly compatible with the Olivia framework.

[15] The earlier Queens' version of Olivia discusses how the framework could handle cumulative, pathway and latency mechanisms.

- The first of these is stock and flow analysis, where people and institutions build up, use, and run down their various assets over the course of their lives (including health, skills and knowledge, physical assets, financial capital and social capital).

- The second, which is discussed in this sub-section, is a consistent approach to analyzing trajectories and transitions over the course of life. It describes the main compartments social life, including their continuities and discontinuities.

Life-course analysis divides an individual's life into various domains such as school, family, and work. Most people, when they are thinking of their lives and their experiences, tend to group those experiences in this way – their lives at home, at school and at work. We refer to these as trajectories. A trajectory consists of transitions and intervening states. For example:

- A *state* (or stage of life) might consist of holding a particular job or being married to a particular person, or going to elementary school.

- A *transition* might be losing that job, becoming retired, getting a divorce, or moving on to secondary school. Transitions are defined as large changes (discontinuities) in the resource flows shown in Figure 5, particularly the relationship-altering events that we have considered to be flow without duration.

- A *trajectory* consists of sequences of transitions and states in the main domains of life. As an example, a work trajectory would include a person's lifetime experience in the labour market, including various transitions (losing, finding, changing jobs) and states (holding jobs of different tenures, various periods of job search or training to find new jobs).

Typically young people will engage in at least two major trajectories: life in learning at school, and life in household with the family. A middle-aged person will also typically have two main trajectories: in work and in the household, but may also have a learning or community participation trajectory as well. Many retired people may retreat to a single household trajectory but some may also be part of a community or a care-giving trajectory that provides definition to their lives.

… trajectories are associated with social policy instruments related to rights and rules

The trajectories can be distinguished from each other by the type of rules that govern them. For example, the work trajectory is marked by workplace regulations relating to, for example, health and safety or working hours. The household trajectory is governed by family law. For schools, governments set mandatory ages of attendance and set rules about what is taught. In some trajectories, the non-legal rules of society (such as those related to gifts, the obligations associated with receipt of gifts, trust and reciprocity) may play a larger role than those that are entrenched in law.

In other words, the trajectories offer a new and potentially interesting link to a dimension of social policy that is traditionally handled on a completely different track. This is the rule-making dimension of social policy. It includes a wide range of topics: human rights, social

rights, regulations, crime and incarceration, the setting of standards and expectations, moral suasion and the like.

Standard trajectories will help integrated analysis

The new social data base will allow individual analysts to create their own trajectories, since a transition (which precedes a new stage in life) is simply a resource flow that has been identified as being large for purpose of the analysis at hand. For example, in some forms of analysis, we might like to treat life in the extended family as trajectory, or one's religious life, or one's life in some consuming hobby or sport or volunteering work. All this is possible; the analyst can decide what he or she needs.

However, in order to enable consistent, integrated analysis, it is also useful to establish common definitions for a range of trajectories (and standard transitions and stages within those trajectories) that are often encountered in analysis – such as life in school and learning, life in the family, life at work, etc. This is particularly the case for trajectories associated with important rules and rights policy instruments referred to above.

That is the reason for the inclusion of 'life-changing' events to the list of resource flows. These provide a standard means of marking the beginning and ending of these commonly used trajectories, as well as major change points within a trajectory, such as changing jobs or the birth of a baby.

Operational definitions have not yet been developed for these trajectory markers:

- For individuals, the task should be reasonably straightforward (e.g., they would include things like marriage, living together, separation, death, changes of cities or jobs, changes in disability status, school graduation or birth/adoption of children.

- For institutions, the task will be more complex[16] and less urgent (since we do not yet have much longitudinal data for institutions, apart from families). A good starting point may be the life-course of government programs, since capturing good program data will be an important medium-term challenge for the new social statistics system.

Which individual trajectories? – the proposal of Martin Cooke

An excellent paper by Martin Cooke on lifecourse topics, which was also prepared in conjunction with this contract[17], suggests – that in addition to the family, work and learning trajectories that were described in earlier versions of the Olivia framework – we should include also include standard definitions and concepts related to trajectories for health and crime/victimization. His paper provides practical examples of early priorities for measuring

---

[16] Institutions also evolve over time and build up and deplete their assets. They also go through transition points in their institutional lives: changes in ownership or control, changes in lines of business, introduction of new technologies in firms and new curricula in schools, and conversions to new products or services. Paralleling our description of individuals, it may be instructive to think of these transitions taking place in different domains or trajectories in the life of the institution – the legal/financing trajectory (e.g., incorporation, mergers, bankruptcy), the workplace trajectory (changes in human resources policies, changes in geographic location), the technology trajectory (changes in processes) or in the business line trajectory (changes in outputs or clients).

[17] Martin Cooke, *Integrating a Lifecourse Perspective with the Canadian Social Data System*, 2011

both events and states in these areas (as well as levels, which measure the changes in the characteristics of the main elements that comprise a trajectory).

…Integrated analysis across the standard trajectories

It is important that the framework support traditional analysis of life in school, in work, in retirement, in the family, etc. However the real strength of the framework is that it allows integrated analysis across the traditional trajectories – how life in the parental family affects success in school, how life in school affects life in work, how life in work interacts with life in the family, how both affect health – and vice versa. This is possible, because the framework is based on capturing resources flows in a consistent way, regardless of the trajectory in which they occur and by coding them in a similarly consistent fashion as to their intensity, as described in the preceding sub-section. As noted, it also allows users to construct their own trajectories and compare these with analysis from the more familiar trajectories.

4.3.5 Supplementing the flows to enable reconciliation

In a number of areas, it will necessary to supplement the flow concepts above in order to allow integration and analysis with other data sets, and to allow use of accounting concepts where things add up to outside control numbers.

- As already noted, one such supplementation will be to include a number of passive activities whose function is internal to the individual (e.g., sleep), even though nothing obvious is flowing.

- There will also have to a number of operational decisions about the conventions to be used in capturing resource flows that do not have an easily defined beginning and end. For example, the TV is on in background all day, but is only watched sporadically – or where three or four flows take place at the same time (e.g., saying hello to a neighbour, when getting in the car to go shopping).

- Another example is reconciliation with economic statistics. The SNA and household surveys have used somewhat different concepts in areas such a home ownership and employer contributions to pension plans. Also, there are important economic transactions taking place that affect the individual but without the individual's direct participation – such as the rise or fall in the value of stocks that are held. The conventions to be used in these cases were discussed in documents associated with the golden age. A recent Canadian paper[18] has made many of the needed adjustments in the context of developing a Canadian Social Accounting Matrix (SAM).

---

[18] Yussuf Siddiqi and Meir Salem, A Social Accounting Matrix for Canada, 2006 (posted at http://www.iariw.org)

4.4 Describing individuals and institutions – from flows to stocks

4.4.1 How stocks are described

The last sub-section described the flows among individuals and institutions. In this sub-section, we quickly examine the much more familiar descriptors of individuals and institutions themselves – the stock picture at different points in time.

In an ideal statistical world, there would be close to 100% overlap between flow data and corresponding point-in-time stocks. The stock picture at one point in time is exactly the same as the stock picture at an earlier point in time, plus and minus that flows that have happened in the interim. In real world, there will be many items of data that are missing on either the stock or flow side. However, if there is a large enough overlap between the stock and flow data, we can impute some of the missing data.

In other words, we should design our surveys and administrative records with the goal of creating as many relevant point-in-time descriptors of individuals and institutions as possible including those where the intervening flow data are missing or incomplete.

The full list of descriptors would therefore be lengthy. Figure 9 provides some examples.

Figure 9. Examples of descriptors of individuals and institutions

	Descriptors of individuals	*Descriptors of institutions*
Descriptors related to their resource flows	*(See section 4.3)*	*(See section 4.3)*
Descriptors related to their status and assets		
Descriptors related to the birth of the individual or institution	Characteristics that individuals were born with, or that arose out of their environment when they were born, including characteristics of the individuals' parents – their education, income, wealth, ethnicity, race, etc.	Organisational structure, size, legal status, main product lines, etc. Separate information for each workplace place. For the programs of public institutions, standard descriptions of inputs, processes and outputs at the time the program was first introduced.
Descriptors related to subsequent status – a trajectories perspective	Descriptors of the status of the individual as it relates to the trajectory in question (e.g., occupation, educational attainment, family status, disability status) Links to the descriptors of the institution to which the individual was associated during the course of the various	Various characteristics of the program, firm, or school during, i.e., numbers of employees, payroll costs, business lines, working conditions, etc. – with an emphasis on characteristics that affect individuals. For social programs and taxes, this includes information about the status of its inputs, processes, outputs and outcomes – as these change over

	trajectories.	time.
Descriptors related to subsequent status – a resource perspective that provides readings on the stocks (assets) associated with the resource flows above. *(overlaps the heading above)*	Data on physical and financial capital, on skills/knowledge levels, on health status and social capital at different times in the life of the individual. See Section 4.4.4 for examples	(This is primarily in the domain of economic statistics and it is the economic system that will determine contents here.)
Descriptors of attitudes, objectives, values and opinion.	See Section 4.4.5 for examples	The highest priority relates to the objectives of government programs
Descriptors related to the temporal and spatial context.	Based on a description of the characteristics of an individual's dwelling at a point in time – and of the neighbourhood in which the dwelling is located – rolling up to economic, social and environmental information about provinces and country as a whole. See Section 4.4.3 for examples	As with individuals, but based on individual workplaces.

Which institutions are to be described?

It is easy to understand the individual as a unit of observation. The situation becomes more complicated when it comes to institutions. Figure 10 provides a standard typology of institutions.

Figure 10. Typology of institutions

Institutions with employees	Programs of public institutions
(Where the workplace – defined in terms of the people who work there – is the unit of analysis)	*(Where the program is the unit of analysis – defined in terms of the people who are contributors and beneficiaries*
Firms – market institutions where the basic unit of analysis is the workplace – the location where people work. Workplaces can be added up into larger groupings such as establishments, firms, enterprises and sectors. The self-employed are also included here for purposes of market-related analysis.	Governments (and organisations that deliver services on behalf of government) can be defined as workplaces for some analytic purposes, as in column one. However, most social analysis is based on the programs of government. These are described in terms of their:
Non-governmental (third sector, or voluntary) organizations, where again the basic unit of analysis is the physical location of the place where the employees and volunteers work.	• **Inputs:** the Figure 5 inflows from individuals and other institutions (including the services provided by individual employees, as well as taxes and contributions).
Public institutions such as schools or hospitals, again where the basic unit is a specific school or hospital or prison in a particular geographic place.	• **Processes:** a hugely neglect part of the statistical system when one takes a citizen's perspective as described in Section 5.
Government proper (when we are looking at the government in its role as employer). We use the workplace as the unit of observation for some applications, such as in assessments of service	• **Outputs:** the Figure 5 outflows to individuals or other institutions. • **Outcomes:** the purposes of the program, as described below. Programs can be classified in the following ways based on outputs:

delivery or in comparisons with other employers. However, the usual unit of analysis is the program, as described in the column opposite.

Other institutions and networks

Households are a key institution, a physical location comprising the people who live there. It is key because geographic locations are attached to institutions and dwellings. The dwelling is the main geographic location used in much social analysis.

Families, where the most common unit of observation is the economic family (a group of two or more persons who live in the same dwelling and are related to each other by blood, marriage, common-law or adoption). We also include individuals living by themselves in this category. The framework is flexible and also allows for analysis of other family groupings including the nuclear and extended families.

Social networks are informal networks of contacts where, in most applications today, it is the individual that is the unit of analysis. That is, we tend to look at an individual's network of contacts rather than at the characteristics of the network itself – although this approach may change as social capital analysis becomes more mature.

- **By type of instrument.**
  - *Instruments based on income transfers/taxes to individuals and institutions.*
  - *Instruments based on rules and their enforcement.* Here governments set the rules that govern key aspects of the operations of markets, families and communities.
  - *Instruments based on the provision of information* such as descriptions of labour market and learning opportunities, or national statistics.
  - *Instruments based on the provision of services*, such as education and training, counselling, health care or active labour market programming.
  - *Instruments based on income transfers to other orders of government or to public institutions* such as hospitals or schools – so that the other party has the fiscal capacity to provide income transfers to individuals, services, rules and/or information.
- **By their purpose**. In the case of the Government of Canada, this would be based on Program Activity Architecture
- **By geographic location** of the workplace where this is relevant (e.g., in-person service delivery operations).
- **By organizational location**. In the case of the Government of Canada, this is known as the responsibility centre coding.

Particular attention is paid to ways of describing the social programs of government, more so than would be found in traditional social statistics typologies. As will be discussed in Section 5, as the main users of the data are seen to be individual citizens and groups (and as 'what works best' uses become more prevalent in consequence), new priority will be placed, paradoxically, on collecting data on programs – their inputs, processes, outputs and outcomes.

4.4.2 Geographic location – integrating macro knowledge into a micro framework

A key descriptor of individuals is the fine geographical coding of the dwelling where the person lives (i.e., of the location of the household of which the individual is a part).

- Similarly, in analyses related to institutions, the location of the workplace is used.

- In more specialized analysis, such as those that deal with commuting patterns or geographic mobility, both can be used.

This fine level geographic classification is absolutely essential to the whole framework:

- It anchors individuals and institutions in geographic space and historic time.

- That is, it allows the macro data associated with those larger spaces (the neighbourhood, the city, the local labour market, the nation, or the state of the world

generally) to be incorporated into the data files for the specific individual or institution.

- And, in a consequential manner, it provides the overlaps discussed in Section 3 that are so important in the new world of integrated, reconcilable social statistics. For example, it allows data on the collective characteristics of individuals to be included in data bases that relate to neighbourhoods or local labour markets or health districts.

Figure 11 provides examples of the importance of spatial information.

4.4.4 Human capital – and other assets

The flow data of Section 4.3 was about the resource flows that created and used skills, knowledge, aptitudes and abilities. It was also about the flows that changed health status for good or ill – often measured in terms of limitations placed on activities of daily living.

The associated stock data is, in practice, far richer then the flow data at present. It consists of periodic measures of skills and health status. Often these involve proxies, such as educational attainment, field of study and occupation as measures of skill levels. However, direct readings are becoming increasingly possible, such as those from PISA and the adult literacy surveys. Direct measures of health status are similarly becoming more common.

Figure 11. Examples of the role played by spatial information about an individual

Taking Olivia as an example, spatial coding would allow us to understand important exogenous information about the context in which she lived at different stages in her life:

- The dwellings where Olivia resided: their size, number of rooms, durables contained in those dwellings, market value, state of repair, ownership, etc.
- The neighbourhoods where those dwellings were located would be described in terms of the extent to which they were deteriorating or becoming rejuvenated, the crime rate, and environmental data such as pollution, weather, etc.
- Local labour market areas in which she lived would have descriptors related to employment and unemployment, pay and vacancies.
- At higher geographic aggregations (e.g., province, region, country as whole), we would also add macro-economic data – where we were in the economic cycle, labour shortages and surpluses, etc.

- Similarly on the social side, we would be able to include information about the prevailing public opinion at the time, the state of the income security and social services, and so on

This kind of spatially-driven macro data related to Olivia's dwelling can be used in an integrated way with spatial information about the institutions with which she is associated. For example:

- The quality of particular workplaces where she worked could be described including health, safety and working conditions information, in addition descriptors related to business lines, profitability, occupations, productivity and innovation.
- The schools that she, or her daughter, attended could be described in terms of class size, Pisa scores of achievement, curricula and teacher qualifications for schools – with comparable information on other social services such as the quality of community recreational facilities and hospitals.

Other forms of capital including environmental capital and sustainability

The Olivia framework can, obviously, easily handle more familiar forms of capital – financial and physical capital. Indeed, earlier versions of the framework included interesting[19] charts that compared the patterns of different forms of capital (financial, physical, human and social) over the whole course of Olivia's life.

Social capital concepts, as noted earlier, are well developed in the academic literature, but there is not yet a mature consensus on how the official statistical system can routinely describe the institutions that are involved, namely formal and informal social networks of various sorts – bridging networks, bonding networks etc. If and when a priority is placed on the development of this aspect of the framework, there are many new and quite powerful tools of computer-based network analysis that can be explored, as well as an associated theoretical literature. The Olivia framework, with its emphasis on resource flows and social interactions at the level of particular individuals and institutions, would seem to provide an ideal basis for the eventual incorporation of network analysis into official statistics.

In some usages, social capital (or cultural capital) refers to the community infrastructure – health care facilities, park, schools, movie houses, libraries, etc. This is readily handled using the spatial aspect of the Olivia framework.

Environmental capital has received much attention in recent years in conjunction with concerns about sustainability. The Olivia framework is useful for exploring the issues of social sustainability. In the social area, sustainable development can mean two quite different things.

- It can refer to *generational* dimensions – where the intent is to leave future generations no worse off than existing generations in matters such as income resource flows, access to various assets, including the capacity to develop one's own human capital, or to exercise real choice in the allocation of time over life. Because the focus of the Olivia framework is on the individual and on linked lives (including among parents, children and grandchildren), it can be used to explore intergenerational relationships directly.

- Or, it can refer to *environmental sustainability* whose objective is to ensure that one generation does not deplete natural or community resources, or leave a climate that will harm future generations. Here, the Olivia framework uses it spatial dimension to include descriptions of, say climate change, or of the stock of natural resources.

In other words, the framework opens up the possibility of interesting integrated analysis of many dimensions of social sustainability.

Human capabilities and the report of Jean-Yves Duclos

Human capital is sometimes confused with human capabilities. Human capabilities have played a large role in recent thinking about social progress, particularly in developing indicators that go beyond traditional economic measures and particularly in developing measures that allow

---

[19] It found that many forms of capital are 'U-shaped' over the course of life, being strongest when one is a child (and still in the parental home) and when one is old.

meaningful comparisons across countries, especially in developing countries. The work is usually associated with the UNDP human development report and with the social thinking of Amartya Sen[20]. However, there are many strands of this work and it has not reached the kind of mature consensus that could lead directly to practical systems of measurement. In an excellent paper that was also sponsored under this contract, Jean-Yves Duclos[21] sorts out the main themes and conclusions of this quite theoretical topic.

Capabilities analysis, and other work on social indicators, is best thought of as an application of social statistics, not a way of structuring social statistics. However, as noted in the introduction, an excellent test of the adequacy of a social accounting framework is its ability to support all forms of current analysis, including indicator work based on the concept of human capacities.

Accordingly, we examined *all* of the empirical suggestion for measurement in the Duclos paper and in the much more widely-ranging discussion of social indicators in the Stiglitz Commission report. Not all the data currently exists, of course, but our analysis did not find any suggestion for possible data collection that could not be readily handled by a data base structured along the lines of the Olivia framework.

4.4.5 Stocks and their relation to social values, goals and perceptions of happiness

The framework treats people's values, feelings of trust, and self-perceived well-being as point-in-time stock measures. In principle, it would be useful to develop a standard set of such measures which could be expanded on an ad hoc basis. While there may not yet be sufficient consensus to proceed quickly in building a comprehensive set of standard questions, it might be possible to develop some selective modules that can be used in different contexts. The question in the World Values Survey may be an example.

The stock data in the social warehouse – on self-perceived well-being (whether based on ad hoc or standard measures) and on assets such human and social capital – can be used in conjunction with the flow data discussed earlier (i.e., satisfaction with flows, perceptions of the limitations placed on the flows, etc.) to get a more rounded picture of quality of life and of social progress.

4.5 Tough operational challenges ahead

The discussion of the Olivia framework in this section has presented many tough practical challenges in translating the general concepts described here into operational conventions and classification systems. Still others will arise as new topics are considered.

---

[20] As already noted, the Nobel prize-winner Amartya Sen was also a key figure in the 2008 Stiglitz Commission's report to the French President on the *Measurement of Economic Performance and Social Progress*, a document that helped generate the current interest in social indicators and social accounts. That report does discuss the capabilities approach as one avenue for measuring the quality of life, but does not give it as much attention as might be thought (given Sen's involvement). It may be that the capabilities approach is not yet a strong base for a system of practical measurement, as Duclos has pointed out.

[21] Jean-Yves Duclos, *A report on a framework for possible use of capabilities in Canada's social data system*. Duclos explains that human development, in this theory, is seen as being an enlargement of choices and freedoms. Capabilities are defined in terms are the real choices or freedoms that an individual has, e.g., associated with health, skills, basic human rights, etc. 'Functionings' refer to the actual levels of achievement in the areas of capability, while resources are the means by which people gain those achievements.

It is important to remember that these challenges do not arise because of the framework. Rather, they arise as a consequence of taking on the ambitious task of transforming the system of social statistics, along the lines outlined in the terms of reference for the present contract.

The challenges are part of the normal, difficult work of statisticians when they develop new ways of measuring the real world – developing new systems of classification, finding measurable proxies for highly abstract concepts such unemployment, or creating new data through simulation.

In other words, the challenges – while large – lie in familiar territory when taken one a time. The real difficulty here is one of scale. There are multiple challenges in many areas. One of the purposes of an integrating framework such as Olivia is to make it easier to plan and coordinate the work that needs to be done on so many fronts.

5 – RETURNING TO THE TOUGH ISSUES: USES AND USERS, EFFICIENCY AND FLEXIBILITY

The drafting of Sections 3 and 4 has implicitly assumed that the uses of official statistics will remain the same in the future as they have in the past, namely the government-centric uses described in Section 1. Similarly, the preceding sections have mainly dealt with applications of technology that are of greatest interest for these two categories of traditional use, namely the technologies used for collecting and storing data.

However, as described in Section 1, the terms of reference for this paper raise two issues that go beyond these traditional approaches: the call for a radical departure in the way we conceive of the uses of official social statistics, and the need to situate matters of efficiency and flexibility in the context of the overall effectiveness of the statistical system in serving these new uses and users.

In terms of the uses of statistics, the prime focus in the future is to be on uses by individual citizens (and firms and other organizations) – with policy uses and other uses by governments being important, but secondary to these uses by individuals.

In terms of efficiency and flexibility, a radical shift in the users and uses of official statistics means that we must deal with the question of efficiency and flexibility for what end. That is, we must also take account of the effectiveness of expenditures on official statistics in meeting a whole range of users, including those of the new primary audience.

While, these topics are central, they are addressed late in this paper in order to facilitate what we hope is a clearer exposition. However, they were, of course, built into thinking that underlies Sections 3 and 4 of the paper. Earlier versions of the Olivia framework had been designed not only with official statistics in mind, but also to meet all the requirements for national social statistics, including a focus on direct uses by citizens and on academic research. It therefore should nicely meet the needs of a reconceptualised official statistical system that has citizens (and firms and organizations) as primary users.

- Section 5.1 uses a practical example to illustrate the combined effects of the new citizen-focus and the new web-based dissemination technology.
- These are radical, transformative effects, as discussed in Section 5.2
- Section 5.3 quickly summarizes issues related to efficiency, flexibility and effectiveness.

5.1 The citizen as the prime audience for official social statistics

The citizen[22] has historically been seen as the indirect beneficiary of good statistics. Statistics are there mainly to improve government programs and the individual will, in turn benefits

---

[22] In the remainder of this section we will mainly refer to citizens for purposes of simplicity. However, the conclusions also apply to firms and the other groups.

from better programs and policies. It is a fundamental shift in perspective to put the individual citizen as the prime, direct user of data.

Some citizens will be interested in the same kind of traditional macro and meso information that is of prime interest to governments and their advisors and critics. Is the current economic situation getting better or worse? How high is unemployment? Are indicators of poverty, inequality, crime or health getting better or worse? How do we stack up on these topics when compared with other cities, provinces or other countries?

These traditional uses are important, but they have already been discussed in the preceding sections. Canada generally does a good job in these areas and Statistics Canada is continually improving public access to data via its publications and internet site.

The new, and much more important, use of official statistics by individuals will be in the use of social statistics to:

- Inform and improve *individual* decision-making in the social domains of people's lives, particularly in social areas touched by social policies such as labour market decisions, retirement decisions, health decisions, or learning decisions.

- Provide the means by which the needs and aspirations of citizens can be *directly* linked to the responses provided by government, i.e., through processes of consultation and engagement.

Such uses were not possible on a routine basis using yesterday's technology. We could only deal with averages and broad groupings, not with information directly related to the individual citizen and his and her particular situation. The internet is beginning to change all that.

Consider an analogy with other current internet uses. Amazon, Expedia and Google allow me to buy books or plan my travels or search for information, in real time – by providing me with information that reflects my own preferences, that allows me to make links to the preferences of others and that, in some cases, reflects the messages that a variety of sellers wish to target to me. Personal finance packages allow us to pay our income taxes using secure information that we have provided in association with the tax rules provided by the government. We might even find a spouse by using an on-line dating service to compare our personal information with that provided by others.

These applications are possible through the real time use of very large bases of individual data, along the lines of the huge data warehouses described in Section 3, plus data that the individual citizen voluntarily provides (ideally on a protected basis so that privacy is completely ensured) about his or her own situation. That allows individuals, when confronted with a social choice, such as the choice between active job search or enrolling in a skills training program, to examine the outcomes of decisions taken in the past by similar individuals in similar circumstances. Figure 12 provides an example. The training example was chosen since the

technology for undertaking this kind of 'what is likely to work best for me' application has already been largely developed and piloted[23].

The example of labour market and learning information

Until recently, the labour market information services of government used computer technology to deliver essentially pre-computer products:

- Bulletin Board information such as descriptions of job openings and available training courses.

- The dissemination of traditional aggregated information drawn from surveys and projections models – in areas such as skills, training, occupational shortages and the like.

The radically different approach shown in Figure 12 is based on information that is tailored to the needs of specific individuals. In particular,

- It draws on vastly richer sources of data that can be accessed in real-time, with much emphasis on data obtained from the administrative records of governments and service providers – as opposed to the traditional reliance on census and survey sources. That, in turn, will redefine the relationship between Statistics Canada and operational departments such as HRSDC.

- Permits two-way flows of information such that individuals can directly access government information on, for example, what course of action has worked best for which people in the past and to apply that information to their own situation in making real-time decisions.

5.2 Broader implications

Figure 12 suggests a deep, transformative change in official social statistics. It may therefore be useful to take a step back and to review the overall context.

- Section 5.2.1 shows how a citizen-focus can be seen as the leading edge of a broader transformation of the roles of citizen and of the state.

- Section 5.2.2 explains how this transformation deeply impacts user consultations about the contents of the statistical system.

- Section 5.2.3 examines the changed roles of Statistics Canada and program departments such as HRSDC, as they relate to official statistics.

Figure 12. The example of labour market information to support decision-making by citizens

We are envisaging web-sites to provide tailor-made, real time information that would assist citizens in making sound life choices in the main social domains	but separate sites, both public and private) that provided information on what to do next – including descriptions of training options open in the city in

---

[23]Indeed HRSDC was piloting this technology in the mid-90s, although then it was designed for use by employment counsellors, not directly to individuals over the internet.

of life, including:

- o Choices in education.

- o Choices in jobs and adult training.

- o Choices in allocating time over life – such as retirement planning, or taking various forms of educational or parental leave.

- o Choices in the care-giving of family members (e.g., in early childhood or caring for the elderly).

These web sites could be public or private. If private, much of the needed data would still come from government. Government web sites could be provided by any order of government, but in all cases needed data must come from the Government of Canada.

The case of an adult interested in improving his or her employment prospects is an example.

The individual in question would be interested in knowing the likely probability of success in the labour market (including expected measures of future job stability and earnings) of different choices such as taking a training course, going back to school or changing jobs.

- o The individual seeking assistance would provide information to the site about his or her skills, employment history and employment aspirations.

- o The site would take this information and, using an underlying data base, calculate the expected success rates attached to different options. These calculations would be based on:

  - o The subsequent labour market experience of people with similar characteristics and in similar circumstances who had made comparable choices in the past.

  - o Projections of future labour market trends and information, where available, about recent success rates of specific interventions.

- o As well as projecting the likelihood of success associated with different choices, the site would refer people to information (typically on related,

question, job matching sites, etc.

All on this would happen in real time – similar to the kind of instant feedback received from Amazon.ca or Expedia.ca or their competitors when choosing a book or a vacation trip.

The underlying data base would, of course, contain information about all Canadians who had, in the past, undertaken specified kinds of training, or who had returned to school or who had changed jobs or occupations. It would also contain information about their employment and learning history before and after the training or job change.

> *(Current technology allows us to build new information from multiple sources and to do so in ways that completely avoid the privacy problems that loomed so large in the past.)*

Data on individuals' satisfaction with the service provided and suggestions for improvement could be gather as part of the transaction. Deeper interactive consultations could be developed for samples of people with different experiences (success stories and failures) during the years subsequent to the intervention.

Virtuous circle – an evolving learning organization. Of course, the same information would be available to those who funded, designed, administered and marketed the training or adult learning in question. This information would enable them to make continuous improvement in the training and to better target their clientele.

That is, the 'Enabling Society' information would be the basis of a quite new type of learning organization in the public sector, mirroring the similar transformations that have affected some private sector companies. A 'learning program' is one that automatically evolves over time based on evidence of what is working best.

Feasibility. This labour market example has already been largely tested and could be introduced quickly. Applications in other areas, where existing administrative records are not as good, would take longer to develop. However this approach will almost certainly become the normal way of doing government social policy business over the medium-term.

Section 5.3 continues this discussion of broader impacts by examining issues related to efficiency, flexibility and effectiveness.

5.2.1 A broader transformation of the roles of citizen and government

We are suggestion that, in the information age, official statistics might play the rather unexpected role as the leading edge of a quite new relationship between the citizen and the government. It will certainly transform the relationship between the official statistician and the citizen.

In past work, including the earlier development of the Olivia framework, I have used the phrase 'enabling society' to describe a society in which the main policy focus is on supporting informed choices by individual citizens (and firms and other groups) – and not on the needs of, and services provided by, government.

The concept of an enabling society goes beyond statistics, obviously. It is part of an overall shift in perspective where individuals and their families are seen as primarily responsible for their own life choices in both the social and economic domains of life – and where the link between citizen and government is much more direct, with less reliance on intermediaries. Figure 13 outlines the role of government in an enabling society and explains the link to official statistics.

… Why statistics can play a leading edge role

Official statistics provide an excellent way of moving into the enabling society:

- The needed technology has already been developed and the internet is widely accessible

- The content is non-controversial. No large philosophical issues are at play. Privacy issues which loomed so large a decade or two ago, have long been resolved. (The data about individuals would not be accessible; they are only used to develop anonymous, real-time calculations of probable outcomes.)

- The pay-offs from the perspectives of government, society and the individual in question are all positive, and large (as will be discussed in Section 5.3)

- Implementation can be gradual, with minimal risks – and can supplement not replace existing initiatives. (For example, the provinces and territories would provide some of the internet content in the example in Figure 12. Implementation could be on a pilot basis with those provinces or territories that wished to move ahead quickly. There could be much variety in the role played by provinces and territories; pressures for directing future evolution would come from citizens, not governments.

Figure 13. Official statistics and the enabling society

The role of government in an enabling society	
In an enabling society, the functions of government include: o The education, health care and justice systems	are as non-intrusive as possible – supporting not replacing citizen-decision making. Target tax credits are a typical current vehicle.

that provide the skills, safety and physical well-being that enable informed and reasonably unconstrained choices in life.

o Financial support that supplements market mechanisms where needed, in order to enable individuals to better manage risk and allocate their resources over the course of their lives (e.g., EI, pensions, workers compensation, maternity leave, or registered savings accounts).

o Providing information that individuals (as well as firms and other organizations) need to make rational decisions for themselves. Government information products play a relatively large role in an enabling society.

o When direct, often remedial, government interventions are needed beyond these basics, they are targeted to areas of greatest need and

Within government programs and policies, there is a focus on:

o Coherence in the use of information at all stages of the policy cycle, including continuous improvement of efficiency and effectiveness based on of evidence of what is working well at the level of individuals.

o High quality in service delivery, consistent with the central focus on respect for the citizen.

o Accountability based on measures of results achieved, defined primarily in terms of results at the level of the final user.

o A strong role for citizen engagement in the policy process – with priority on direct engagement with individual citizens as opposed to via intermediaries.

o Direct two-way information flows between the citizen and the government both at the level of individual transactions and in consultations.

The role of official statistics in the enabling society

The objective of statistical information in an enabling society is to:

o Support the needs of citizens as they relate to the information that they require in making sound choices in the social domains of life. The implications for data content and access are illustrated in the example in Figure 12.

o Support direct communications between citizen and state on perceived satisfaction with the information that was received and information on better meeting future needs.

o Support accountability, results, efficiency and service quality within government in providing this information – at all stages of the government policy cycle, and across different programs and policy areas.

In what sense is this new?

Most of the elements of an 'enabling society' are certainly not new, at least in terms of the aspirations they express. Most elements shown in the figure are already part of the social fabric in contemporary democratic market societies.

What is new is the power of current technology to actually deliver on these aspirations in ways that were simply not possible before the computer age.

For example, without current information processing technology, it would be impossible to have responsive, refundable tax credits. Two-way real time interactions with citizens in program delivery simply were not possible.

Radical improvements in internet-based information dissemination and communications are already affecting many domains of life, but have not yet touched social policy and social statistics. That will happen soon, with dramatic results.

5.2.2 Consultations and communications

It seems likely that a two pronged strategy will be needed to obtain the views of users on their priorities for the collection of new data:

- One strand would centre on consultations with traditional interested groups: federal departments, provinces, municipalities, business organizations, academic researchers,

non-governmental bodies that represent the interests of the beneficiaries of social programs, think tanks etc. The big challenge here will be to shift the focus from instrument-by-instrument consultations[24] to broader discussions about the priorities for the data itself and to extent the consultations to a broader set of players.

- The second strand would centre on consultations with the primary users: the individual Canadian, or the individual firm or group. The big challenge here will be to build user feed-backs directly into the way that citizens (and firm and other groups) will use the data, typically in on-line applications.

The paper has argued that the project is highly ambitious, will take considerable time and effort in the early stages of development and then pick up momentum and progress rapidly – almost automatically, after we reach a kind of tipping point.

In this environment, initial expectations should not be set too high for the early phases. High expectations could create scepticism, not cooperation among important partners in the traditional consultation stream.

- With respect to traditional audiences, we should avoid the temptation to be overly ambitious. For example, we should *not* cast the exercise as Canada putting itself in a world-leading position in implementing a new approach to social statistics, in line with the OECD and the Stiglitz commission. Nor should consultations be cast around the implementation of ambitious new approaches to social indicators – or around the Olivia framework itself. Approaches at such a high level of generality could simply lead to costly suggestions for new data that cannot be met, at least in the short-run – and would risk creating the scepticism referred to above.

- The suggestion here is to be much more practical and to cast the consultation strategy with traditional users around issues of efficiency. The task would be to explain to those consulted how the proposed data warehouse will reduce overall costs and response burdens in the long-run – but at the same time increase the amount of accessible data through the techniques discussed in the paper that involve much greater use of data obtained from multiple sources. As well, many of these techniques can be cast as ways of managing the risks associated with the recent decision to shift the mandatory long form census to a voluntary survey. The consultations would be cast in terms of how the needs of those consulted could be best taken into account during the period of transformation.

The consultation strategy related to individual citizens (or firms or groups) should be highly pragmatic and tied closely to the new interactive tools for accessing the data that were illustrated in Figure 12. That is, we would not ask citizens or firms about their priorities for official statistics. Rather we would obtain information about how the interactive site was actually used and how it could be improved. That is, user's needs for data can be determined

---

[24] Statistics Canada and HRSDC are already considering consultations about the content of various data platforms that encompass more than one instrument – such as rationalizing the data content of cross-sectional surveys, across longitudinal surveys, or across all data about departmental programs. This approach is a distinct step forward. However, it should be considered as a transitional step since it is still instrument-based – and misses the potential power of creating data from overlapping data bases (e.g., from overlaps between surveys and administrative data).

through analysis of actual use – supplemented by surveys of user satisfaction, and periodic focus groups/surveys that explore needs in more depth.

5.2.3 Changing roles of Statistics Canada and policy departments

Making the citizen the prime audience for social statistics means a radical shift in the content of official statistics and the way they are disseminated. In particular, it involves bringing together in the same web application:

- Administrative data related to the application in question, active labour market programming funded by the Government of Canada and administered by the provinces and territories, in the example used in Figure 12.

- Traditional survey data from Statistics Canada and administrative data that Statistics Canada has obtained (often tax data) for statistical purposes.

- Data from the individual user, similar to data that they would have provided to the employment counsellor, to continue following the example of Figure 12.

- Bulletin Board information on available jobs, available training, information and promotional information etc., often already accessible on other web sites – the task being to make appropriate linkages these sites.

The first three of these data sets need to be closely integrated in order to provide 'what has worked best in the past' calculations of expected success.

This, in turn, requires a quite new kind of partnership between Statistics Canada and line departments such HRSDC in producing this integrated data. For example, it involves incorporating administrative data about program operations into the system of national statistics. There has long been a world-wide trend towards greater use of administrative records in official statistics. This will accelerate greatly as we move towards a citizen-centric approach.

Questions of data ownership, per se, may not arise, since the policy department could continue to 'own' the original administrative data. Statistics Canada could 'own' the version of that data which was used in official statistics (after making needed adjustments and reconciliations needed for integration with other data sets). Service Canada and Statistics Canada might, for example, 'own' the still more refined version of the administrative data that was used on the web site itself, in conjunction with the personal data provided by the individual client. While each department would have an independent, separate role, the exercise obviously requires close co-ordination and would change the way in which departments have traditionally worked together.

In these circumstances, pilot projects and experimental methods are needed to test out new ways of working together. In the initial phase it might be simplest to work with just two departments: HRDSC and Statistics Canada. Leadership would be needed at the top of both departments, likely using a formal project management structure reporting at a senior level. Figure 14 suggests a number of areas where such leadership may be needed.

Figure 14. Some possibilities for a co-ordination check list

Following are examples of the kinds of monitoring that would need to be undertaken to ensure that the needed co-ordination was taking place:

o Is progress in implementation being monitored in the planning and accountability systems of the departments concerned?

o Have plans for internal or externally-sponsored R&D throughout the departments been adjusted to support the new project, where this is appropriate?

o Are key partners aware of the project and their potential role in it?

o Has a high corporate priority been placed on funding closely related tools – and on ensuring that they are specifically aligned with the goals of the project?

   o Examples would be Geographic Information Systems, dynamic microsimulation analysis, and informatics support.

   o In HRSDC, this would include a high priority being placed on developing labour market and learning information systems along the lines discussed in Figure 12 – particularly since HRSDC is a world leader in the technical development of such systems.

5.3 Efficiency, effectiveness and flexibility

Most traditional data and analysis deals with averages, crude measures of distribution and, more generally, deal with the effects of policy on representative people, not real people. It is good at assessing first order effects of programs (costs, numbers of winners and losers, etc.), but not how the programs affect people over the longer-term and how the effects of one program interact with other programs, including those of different orders of government. There is rarely any capacity to examine how programs, especially combinations of taxes and programs, play out in the lives of specific individuals.

All that changes with an integrated system based on microdata and linked data warehouses, including data provided individual users and detailed data about the programs in which they participate. Here we can actually measure program effects and assess the efficiency and effectiveness of government programs that provide public services. We can see the characteristics of people before and after an intervention, and measure which kinds of interventions made the biggest difference at least cost.

When this technology is combined with a citizen-centric focus, we see a quite radical transformation in the role of official statistical data. A key message from Figure 12 is that putting the citizen first results in better performance across the whole system.

- The nature of the relationship between the citizen and government is transformed in a way that results in direct, empirically-based, two-way, real-time relationships, whose purpose is defined by the citizen, not the state.

- The data that is needed by the citizen can also be used to create 'learning' programs that automatically improve over time as result of feed-back loops that based on empirically evidence of what is working and what is not.

- True consultations based on the actual needs and aspirations of citizens are built right into the operation system.

- Costs to governments fall (information is a relatively inexpensive product line) and societal as well as individual outcomes improve (through improved labour market efficiency in this example, as well as improved efficiency in the training market).

The term of reference for this project calls for an examination of efficiency and flexibility gains. These are large and important, but they are swamped by these even greater gains on the effectiveness front. To summarize Section 3, the efficiency and flexibility gains arise from:

- Data being entered only once, and once captured and verified for quality, is then used freely in multiple applications.

- A system that works no how much data is collected (no need to collect data that is only there to feed the system) and that works better with each new data set that is added.

- Data which are collected at the smallest micro level possible, with huge consequent gains in flexibility in subsequent use in a wide variety of applications, including those that cannot be currently foreseen.

- Comprehensive capture of micro data from multiple sources (including overlaps, intended duplication and exogenous control totals) will provide quality controls and allow the creation of quite new information without the need for new collection – and at very low unit cost.

- Uses of the data by citizens to improve their decisions, such as those outlined in Figure 12, automatically increase the richness of the data for future use – as we can learn from the actual subsequent experience of the individual, compared with probabilities of success that were calculated before the decision was made.

6 – Conclusions and Q&As

We will conclude by referring back to the terms of reference for this paper, including the content of initial discussions held with the sponsors to clarify expectations.

First, we were asked to provide a description of social accounting and how it was relevant to long-term planning for the collection of social data.

- The paper described the history of social accounting, its ups and downs and lessons learned.

- Perhaps the main lesson was that an accounting framework that describes the content and structure of social data needs to be descriptive in character (not normative as is the case with social indicators) and that it should be comprehensive, based on micro data about individuals, including the longitudinal and spatial aspects of people's lives. It needs to encompass both economic transactions and life-course dimensions.

- Perhaps going beyond the terms of reference, the paper also provided a fleshed out version of a contemporary social accounting framework that was adapted for use in the long-term planning of official social statistics in Canada.

Second, we were asked to explore the implications of the framework in light of a radical shift in the audience for official social statistics, with direct uses by citizens, organizations and firms taking precedence over the traditional policy uses and users.

- We did this in conjunction with a discussion of the transforming power of new internet-based technology.

- We discussed how official statistics, perhaps surprisingly, promised to be a good starting point for a more major shift in governance which we labelled as the shift to a new 'enabling society'.

- We described the consequences for the roles of the major players in the statistical system and for consultations strategies. Perhaps going beyond the terms of reference, we made some practical suggestions about implementation strategies.

Third, we were asked to discuss a data infrastructure that was efficient and flexible.

- A good deal of the middle the paper, Section 3, was devoted to describing the efficiencies and flexibility that will arise when we shift away from planning and structuring our data on a collection-instrument basis. Using the data warehousing approach and the associated social accounting (Olivia) framework will produce huge gains in efficiency and flexibility – including the production of quite new data at practically no cost by exploiting deliberate overlaps, duplications and exogenous control totals.

- Section 5 discussed gains in efficiency and flexibility that will also be possible with use of web-based technologies in the dissemination of data, particularly for the new primary citizen-centric audience. These will also be very large. However, they pale

beside the gains in effectiveness that will occur – both in reducing the costs of existing government programs and improving the operation of social systems of policy importance, such as employment programming.

Fourth, perhaps again going beyond the explicit terms of references, the paper attempted assist in subsequent analysis by showing how the findings of the other two papers that were part of this contract (the Cooke and Duclos papers) related to our social accounting paper.

- The Olivia framework includes life-course perspectives as well as traditional economic perspectives in its very foundations. The Cooke paper, which was specifically on life-course perspectives, made some valuable additions to the literature that had not been included in earlier versions of the Olivia framework. These were added to the version described in Section 4.

- The Duclos paper on social indicators related to people's capabilities argued that this approach was not sufficiently mature to be considered as the basis for an official statistical measurement framework. This is consistent with the argument of the present paper that we are very far away from any consensus on a theory of social progress that could be used as a framework for official social statistics. However, we did examine all of practical measures referred to in both the Duclos paper and in much broader Stiglitz Commission Report to see if all could be readily described and place within the proposed Olivia social accounting framework. All could.

Fifth, the sponsor asked for some specific recommendations on what data should be collected and with what frequency. A social accounting framework, of course, describes the content of social statistics, and does not deal with priorities for new data collection. It can point in some useful general directions but, by definition, cannot be specific. Following are some of the main new directions that are indicated for new data collection and for closely-related developmental work:

- Data needs to be collected about the *characteristics of government social programs* (including taxes and transfers), with routine updates to keep up with changes in those programs. This is done now for taxes and transfers, but needs to be supplemented with descriptions of the processes involved in service programs such as training. This, in turn, puts a priority on developing a standard classification system for describing the content of these programs – whether delivered directly by an order of government or by third parties.

- The development of an *interactive labour market and learning information web site* described in Section 5 should have high priority. Most of the development work has already been undertaken by HRSDC. It would be low cost (and, indeed, would soon produce significant savings) and would afford many win-win opportunities. It is consistent with gradual, experimentally-driven implementation – working with interested partners at the provincial level. It would be both a pilot for the new ways of doing business and a demonstration of its power.

- Consistent measures of *perceived well-being* have long been seen as a critical gap in official statistics. It was highlighted in both the 1980 report referred to in Section 2

57

and in the 2008 Stiglitz commission report. Measures of perceived well-being also play a major role in the Olivia framework's approach to assets and their relation to the flows that produce well-being. While a consensus likely does not exist for the development of a comprehensive approach to these indicators, it should be possible to develop one or more standard modules of questions covering some aspects of values and well-being that could be used in various collection instruments as appropriate. These could be consistent with data collected in the vehicles such as the World Values surveys. A core set of questions should be asked at least annually, typically in surveys that deal with a range of other topics.

- More frequent collection of 24 hour *time use data* would appear to be a priority – but as part of an overall plan that involves the collection of other forms of time use data using consistent concepts. For example, this would include blocks of standard retrospective questions that could be included in various vehicles, as well as regular collections such as the labour force survey (which collects time spent in work on a monthly basis). It might be useful to have a standard diary type survey with common concepts and methods conducted every two or three years to ensure consistent time series and, in other years, to have another series of more experimental surveys that explore different approaches to measurement or that probe different aspects of time use – in order to build up our expertise in areas such as measuring social capital or the learning intensity of different time-based resource flows.

- It is likely timely to begin work early on the development of a set of *standard definitions of events, transitions and trajectories* as suggested in Section 4.

- As well, the specific priorities for *life-course data* (including frequency of collection) developed in Martin Cooke's parallel paper make much sense from a social accounting perspective as well.

- A series of *low-cost specific data creation activities* should be planned as an integral aspect of the new approach. These would be based on creating data from existing data sets, (i.e., not new collections) and would have three functions: a) creating useful new data; b) helping users (and players within the system) become more familiar with these powerful new data creation techniques and c) providing case studies for those involved in the more technical work of developing the needed data bases and analytic tools. Examples could include:

  o Incorporating existing work into the 'official system' such as the creation by DataAngel of highly disaggregated readings of the demand and supply of workers skills, by imputing reading comprehension scores into the occupational data on the 2006 census.

  o Extensions of work along similar lines, such as imputing adult literacy and a range of health and GSS data onto the monthly labour force survey.

  o Adapting the major progress being made in the US using geographic information systems.

Also related to the question of priorities for new data content is a question about the initial scope of the content of the social warehouse. The Olivia framework itself is open-ended, it can incorporate virtually any social data of interest, going well beyond the data included the official system of national social statistics. In order to help set reasonable expectations about the content of official social statistics, and to signal that the government does not intend to stray into areas best handled by others (e.g., academics, departments acting independently, or the private sector), it might be wise to be explicit about the initial scope of the new of official statistics. Figure 15 offers a suggestion.

Some concluding questions and answers

Finally, we will use a Q & A format in order to:

- Address some specific comments that were raised in discussions of earlier drafts of this paper.

- Set out some additional, and more personal, views on questions that were asked in the initial terms of reference for the paper but that could not be fully addressed by argument of the paper.

- Underline some of the main conclusions of the paper.

1. *The argument seems to be that technology allows us to collect everything. Is it still not better to think parsimoniously about what to measure? Is it better to measure everything poorly or measure fewer things precisely?*

The new technology makes it efficient to store, and make accessible, a very wide range of data that has been collected, including data varying degrees of quality and from different sources – provided that the source and quality are well documented and made clear to users. Informed user choice trumps statistical paternalism.

(To be fair, in nearly all cases the data should be in the 'fair to excellent' range, not the 'poor' range).

Having many sources can act as an important check on the quality of the whole system.

Of course, hard choices must continue to be made in planning the content of individual, expensive vehicles such as surveys or censuses. Here, the familiar trade-off among usefulness, cost, quality and response burden remains real – but becomes somewhat easier to manage knowing the trade-off need not be made entirely within the limits of a single survey vehicle. One can also take account of the strengths and weaknesses of alternative sources of this information.

Figure 15. A suggestion for the initial scope of the new base of 'official national social statistics'

There will be many pressures to enlarge the scope of	o Much of this information now exists (to a certain

59

the exercise. The more social data that is included in the data warehouse, the greater will be its power and the better it will serve the public – especially citizens for whom the territorial divisions among orders of government and departments will be irrelevant.

However, too broad an initial scope would invite organizational paralysis. The following is a suggestion for a compromise.

First, an initial focus on Government of Canada data. At least initially, the scope should be limited to data that is collected/sponsored by, or related to the mandates of, Statistics Canada, HRSDC and other social departments. For example, it would **exclude**:

o Data used internally by companies and other non-governmental groups.

o Most polling or market research.

o Academic research that is based on small data sets or experiments and, more generally, any data that is not directly related to policy applications or to nation-wide use by citizens.

Second, an initial focus on social data about individuals including data related to:

o Labour markets, learning, care-giving, income security and poverty broadly defined, industrial relations, disability, social well-being and related topics – i.e., within the mandate of HRSDC.

o Population health, but with linkages to external data on medical interventions. For example, disease per se would not but be included, but health status would be included, including barriers to daily living.

o Crime and victimization.

Third, a warehouse describing the social programs of all orders of government would be developed as the first warehouse for social data that uses the institution as the unit observation.

extent) in the tax/transfer models, in various administrative files, and in work undertaken by federal, provincial, territorial working groups.

o It would need to be extended to include descriptions of the actual processes that occur in these programs.

o Descriptions of provincial and territorial programs would be included on a pilot basis, with the active co-operation of the province or territory.

o Doing this would create the need to rethink the respective roles of Statistics Canada and departments in the creation, storage and access to social statistics.

Fourth, asymmetrical development – working only in partnerships with those ready and able to play. The framework can work with whatever data is available. In developing new types of data, particularly from administrative records, it is likely best to work on a selective, pilot basis with those departments or other orders of government who have an interest in working in partnership.

Fifth, data from many sources, but not qualitative data – at least initially. This approach would encompass data from traditional survey sources and all the administrative records in these areas.

o However, the system should foster linkages between the (included) quantitative and the (excluded) qualitative data.

o This will happen automatically if the Olivia framework is used.

Sixth, the focus would be on data that either already exists or that might potentially be collected within a period of say ten or fifteen years, given reasonable resources – and using only existing information processing technologies or those that seem highly likely be available in the next several years. (This is does not exclude much; we have barely begun to exploit the power of existing technologies.)

2. *The argument seems to put a huge premium on the importance of reconciling data from many sources. Yet this is an extremely difficult task in practice and one where the levers (as is the case in controlling the content and quality of some*

admin data) are not always in the hands of survey statisticians. Is this not cruel and unrealistic?

The reconciliation of different ways of measuring the world goes the heart of the theoretical work of a survey statistician – being able to draw quantifiable, consistent, useful meanings from the complexity and messiness of the real world. It also goes to the heart of the marketing work of the statistician – persuading others to use the concepts and definitions they devise, including adjusting administrative data to serve broader social ends.

So, yes, it will be a challenge.

The task is tough, but familiar, necessary and manageable – and with high pay-off. A huge advantage of the proposed approach is that it can evolve gradually. It can work with existing data. It will work better with more integrated, reconciled data. Each minor victory in reconciliation will be a non-reversible step in the right direction.

3. *Huge data bases with individual information will inevitable raise privacy issues. Is this not a deal-breaking problem?*

In the early days of data warehouses, for example in the early 90s when HRSD started piloting the use of individual data taken from tax files in order to calculate the expected success of individual training decisions, there were potential issues of privacy that would have to have been addressed before the project was fully implemented. And, indeed, perceived privacy issues were among the reasons that this project was eventually cancelled.

However, the technology has moved on since then and, today, this kind of application poses no threat to privacy. The calculations in question are all based on anonymous equations, not on the source records. LifePath microsimulation modelling in Statistics Canada has produced a fail-safe answer and one that is easy to explain, namely basing its calculations not on real people but on synthetic people who have the same overall characteristics as actual Canadians.

Nevertheless, perceptions of privacy problems may well still exist and need to be carefully addressed in the implementation strategy.

4. *We were expecting that an examination of the history of social accounting would provide some lessons about what specific data to collect and with what frequency. Why does the paper have so little to say on this point?*

Social accounting was always more about ways of organizing data than about priorities for its collection. It is a way of organizing social data, not determining what data should be collected. Indeed, the conclusions about data content that are listed at the start of this section are, if anything, stretching the limits of the positive lessons that can be learned. There are a some useful negative lessons however:

- Avoid approaches that require the collection of data whose main purpose is to fill the needs of an intellectual framework rather than the real needs of users. Demographic-based versions of social accounts, including some current versions of Social Accounting Matrices, are most likely to fall into this trap.

- Avoid reliance on technologies that are not yet fully operational. You may be proven to be right eventually, but people will lose interest if payoffs are only in the distant future. This is likely a main factor in the loss of momentum in developing time-based social accounts in the 1980s and 90s.

- Avoid basing data collection strategies on formal theories of social progress. Social indicators are important uses of statistical systems, not tools for designing those systems. There simply is no consensus on what would constitute a single best approach for measuring social progress – and the search for such consensus can be a major diversion of energy for the social statistician engaged in strategic data planning.

5. You have described one approach to a framework of the planning of official data, the Olivia framework. What are the alternatives? Why have you not described them?

Other frameworks do exist but have other purposes. For example, as argued at some length in the paper, social indicators are sometimes wrongly thought to be an alternative approach, but these are really particular uses of data (measuring social progress), not ways of structuring data. And, in any event, we are very far away from a consensus on a theory of social progress that would be strong enough on which to base a system of social statistics.

The real question is whether there are any reasonable alternatives to the Olivia framework for purposes of social accounting. The obvious alternative is continued reliance on the ad hoc approach currently in use, where each instrument is planned separately and the resulting data structured along the lines of the original collection instrument – with some degree of integration provided by common definitions and the use of standard classification systems. However, this approach does not work in the new world of integrated social data warehouses as described in the paper.

Is there, then, any alternative to the Olivia framework for planning and structuring data along the new warehouse lines? I am aware of none and I have looked hard. One reason is that the Olivia framework is extremely flexible. Each new application that has come along has been readily handled by the framework, or it has proved simple to adjust the framework to meet the new application (and without upsetting the consistency of its use in past applications). That is, the challenge is not so much to find alternatives to the Olivia framework, but continue its evolution by using it in new applications.

Why is it that we have such a framework in Canada, with no equivalent in other countries? See question 9 for an answer.

6. The Olivia framework seems individualistic. Does it not downplay the essential role of family, community and organizations in a good society?

In its early development to date, the Olivia framework has placed much emphasis on individuals. This was partly for purposes of exposition (we use the example of Olivia to explain how the various modules of the framework were inter-related) and partly because we do

consider the individual as the primary unit of observation for social policy. However, Olivia was never seen in isolation but as the centre of a complex web of resources flows with other individuals and institutions. Institutions are defined very broadly to include families, firms, organizations, communities, networks, government programs, etc. A fully developed system of social statistics would contain information about all these entities. Integration is provided by attributing the characteristics of these institutions to the individuals that comprise them when this appropriate and, in other cases, through the spatial module which allows us to associate exogenous information about communities of interest with the individuals who live and work in affected neighbourhoods or who are associated with the communities in question.

In other words, the framework is not individualistic in the sense that individuals are seen as the sole entity of interest or that are measured. It is individualistic in a shorter-term tactical sense, however, since we argue that priorities for early development should be focussed on information about individuals and closely-related data about families. On the institutional side, we argue that, aside from families, the priority for early development should be on consistent and richer statistical information about government social programs and their relationship to individuals. However, the framework can equally well serve those who have different priorities for developmental work.

7. *Does the technology really exist now to allow a shift to the data warehouse approach being advocated? Are we really ready to drop the deeply engrained tradition of organizing data according to its survey or census source?*

The technology is old hat. Database technology and data warehousing techniques have been used for decades in other applications. There is no technical problem at all in organizing data according to individual characteristics (or institutional characteristics) as advocated in this paper; it is the normal way that one would do given a fresh start and given today's technology. It is similar to applications already in use in the statistical world where we create a database in conjunction with a particular modelling application.

There will be transition problems, of course, as we move from one system to another. However, there is no need for a fast transition.

And, most important, there is no need to drop the old instrument-by-instrument data files. Indeed, the original files are needed as inputs to the new warehouse – and can continue to be used for analytic purposes along traditional lines. Similarly, the data in the warehouse will contain full information about the origins of each data item.

7. *Do users really want the new kind of data that is created from existing sources, as opposed to asking questions on surveys or censuses? There does not seem to be much demand for the existing tools that do this, such as microsimulation modelling.*

The most important of the new Enabling Society users – individuals and groups – will have no trouble at all. Quite the reverse, it is the normal way in which they receive data, similar to ordering a book or arranging travel plans on line. The technical stuff discussed in the text (i.e., using the experience of synthetic individuals to calculate real time estimates of the expected success of certain kinds of labour market or learning decisions) will all be in the background and invisible to the actual user.

Traditional users, who have invested heavily in the use of their traditional tools, have always been suspicious of new approaches that seem to magically create new data using techniques that cannot be personally verified by the researcher, namely techniques that produce accurate information about a whole population based on only partial observations.

It took a long time for sample surveys to catch on, or the use of powerful imputation tools to clean dirty admin files. Today progress has been slow in the use of geographic information systems that create new data for small areas with no new data collection, or dynamic micro-simulation models that use synthetic individuals. And, as well, many of these new techniques require expertise that is not yet common.

However, as the text suggests, once things start to move, they will move quickly.

And the traditional methods will still be available for those who prefer to use them.

8. Are there best practices in other countries in the area of social accounting frameworks? Would Canada be isolated if it took the lead in their development?

I was unable to find much experience in other countries that would provide best practice lessons in area of strategic statistical planning.

Of course, there are many specific lessons to be learned – certainly from the Americans in the area of Geographic Information Systems and from a number of countries in the area of longitudinal surveys. The European experience with population registers may well provide practical lessons on the technical side of data base construction and statistical linkages – although I have not investigated this myself.

It is important to remember that Canada is uniquely placed to undertake work of this sort. Statistics Canada has an unusually broad mandate in the area of social statistics development. HRSDC, when compared with counterparts in other countries, has both an exceptionally wide mandate in the area of social policy and an unusually strong commitment to data and analysis.

Similarly there are few countries in the world that have been as well placed to develop social accounting frameworks such as Olivia. It was first developed at the Policy Research Initiative – an internal policy think tank within government that has a mandate on both cross-cutting data and policy issues. It was further developed in the brief period of the creation of Social Development Canada, where there was a conscious effort to rethink priorities for social data to support social policy.

And there are few precedents for the democratization of official statistics that is on the table in Canada – making the individual citizen, firm or organization the prime user.

That is, there are few, if any, countries in the world that have the same potential for leadership in this area. This is even more so the case, when we look at the differences in statistical contexts in different countries. The strategic planning challenges facing Canada are shared only with relatively small number of countries that already have a rich statistical infrastructure that is primarily based on surveys and censuses and that are consequently poised to greatly extend the scope of that infrastructure by adding to it the richness of the administrative data sources that are found in some European countries.

Several years ago, I presented the Olivia framework at an OECD workshop in Paris that dealt with life course perspectives. In discussions there, I found international interest, but little practical experience with this kind of strategic statistical planning tool. At that workshop, the Australians and the Dutch were the most advanced, but neither had evolved as quickly as us.

That is, if we take the lead, other countries will be following us. In some cases, that can be a problem. It is sometimes costly being the leader and can result in the development of an infrastructure that can quickly become obsolete. That is not the case here. The technologies involved are all familiar. Implementation can be gradual and the system is highly flexible in terms its data content. The intellectual frameworks used are descriptive and do not rely on current thinking about social progress – thinking that could quickly become obsolete.

In short, there seems to be no downside to taken on early leadership – and everything to be gained.

9. So what do you really think about social indicators? If social progress is not the important driver, then what is?

Some readers may agree with the distinction in the paper that indicators and accounts serve different purposes, but still wish that we had not concentrated so heavily on the boring, descriptive social accounts and had given more weight to the different, but potentially even more important, topic of social indicators – what we need to know in order to better measure social progress. It seems to make sense that large gaps in the information needed by social indicators can be a good signal of where to invest in new data development.

Certainly there has been renewed interest in social indicators – formal approaches to measuring social progress – in recent years. And some have advocated that creating social indicators should be a main driver of new social data collection.

However the reality is that most systems of social indicators today are essentially descriptive and rely on a wide range of disparate theories of what is important (unemployment rates, inequality, victimization, GDP). There are also efforts to develop a more coherent, integrated approach to indicators, with work around human capital, human capabilities and other asset-based approaches being the most important. However, to repeat myself, a careful reading of this literature suggests that main proponents do not suggest that this work is sufficiently developed to be used as the sole basis for new data collection. This is explicitly the case with the Duclos paper. And, as noted in the text, even the Sarkozy commission report, with

Amartya Sen playing a key role, did not rely heavily on the human capabilities approach in its approach to measuring social progress.

All this being said, the author does have views on the development of social indicators. These are entirely consistent with the social accounting framework described in the paper, but it would be wrong to suggest that they flow logically out of the social accounting literature. Accordingly, they have not been included in the body of the report. In my view:

- It would be desirable to develop an outline for a number of different sets of indicators for different purposes and using different perspectives, i.e., to avoid the temptation to develop a single ideal, all-purpose system of indicators – and, especially, to avoid schemes that attempt to add up component indicators into one encompassing figure that attempts to measure social progress[25]. These have a useful function in the world of advocacy, not the world of official statistics.

- Data gaps would exist in many of the indicator sets that would be sketched out. These could be described and ways of filling them could be assessed. That is, creation of multiple sets of indicators could be a far better risk-management strategy for statistical planning purposes than would be reliance of a single approach.

- Among these sets of indicators would be one that provided a common time-based denominator to traditional point-in-time indicators – likely life-time hours or perhaps annual hours. For example, we would have a set of indicators that showed expected life-time hours or years spent in unemployment, in poverty, in sickness, in work, in isolation, in self-perceived life satisfaction, in school, etc. This relatively modest approach might provide some quite new insights and would require minimal collection of new data.

- Other indicator sets would take a perceived well-being perspective (including both the process well-being and outcome well-being measures discussed in the text). Still others would take an asset-based, dynamic approach with measures of stocks and flows in the areas of human and social capital. Others might take a life-course trajectories and transitions approach.

10. Is it reasonable to take leadership in developing new conceptual framework to support social data and analysis during a time of retrenchment and cutbacks in statistics – especially during a period when, at least according to its critics, the government does not place high priority on the use of statistics in social policy decision-making?

---

[25] For those who desire an integrated set of social indicators that is comprehensive but that avoids the adding-up trap referred to above, it is important to remember that Canada has been a world leader in this area for many years. We do not have to start fresh. I am referring to the excellent work on the 'Index of Economic Well-being', developed by Lars Osberg and maintained by the Centre for the Study of Living Standards (http://www.csls.ca/iwb.asp). As its name indicates, it has an economic orientation, but nevertheless has a broad scope and covers many areas of interest to social policy.

Any big change along the lines that we have proposed will be difficult. It will be hard to know where to start – and where to find the leadership that will be needed during the transition period before the payoffs from the changes begin to be felt (even if the benefits are large). However, even in a period of rebuilding after cutbacks, there might be merit in having a larger framework to guide development, especially one that promises lower-cost outcomes and with potential for much higher payoffs.

That would, of course, make sense only if there were high level interest in moving from a government-centric to a citizen-centric approach to social statistics. That implies at least a minimal interest in statistics as part of the national agenda.

I have read critiques in the media that suggest that the government is simply not interested in statistics as an important tool in social policy analysis, regardless of their orientation. In my own reading of events I simply do not see that. I do see stories that suggest that existing statistics may be largely irrelevant to some of the issues of the day – especially those which should be based on values not measures. To me at least, that makes much sense. Statistics on the incidence of this or that cannot, and should not, replace judgements based on what is right, or felt by the people to be right.

I also see a very reasonable frustration with today's government-centric statistics. Our traditional analysis has government and policy-making as it main audience and therefore concentrates on topics that are inherently those where potential collectivist solutions are at the forefront. The response to those topics may be positive (new and improved government programs) or, often, negative (government solutions do more harm than good, so reduce the sphere of government action).

If we shift the prime audience for data from government to citizen, I think we can eventually move to a world where empirical evidence is designed and developed to support the entire range of social decision-making – by individuals, associations, governments and firms – in a more neutral, constructive manner, supporting analysis at all points in the political spectrum. It would support a shift from backward-looking 'post-welfare-state' perspectives to more forward-looking 'enabling society' perspectives.

Also the data to support an enabling society, as outlined in this paper, holds promise of large gains in effectiveness, as well as efficiency. It is also less intrusive.

In essence, I do not see why a change in directions proposed in the paper should face any objection in principle. There may be scepticism about our ability to carry out such a major change. Strong leadership will be needed. However, the conditions to allow such leadership do exist, if my analysis is right.