



Using what keeps official statistics fit for any time and place for building a wider information (evidence) system for developing and applying policy

Leonard Cook and Vince Galvin

Paper prepared for the 16th Conference of IAOS
OECD Headquarters, Paris, France, 19-21 September 2018

Session 6.A., Day 3, 21/9, 14:00: Central role of national statistical offices

Leonard Cook
Len_cook@xtra.co.nz
University of Wellington

Vince Galvin
Vince.galvin@stats.govt.nz
Statistics New Zealand

**Using what keeps official statistics fit for any time and place for building a wider information
(evidence) system for developing and applying policy**

DRAFT VERSION 22/8/2018

Prepared for the 16th Conference of the
International Association of Official Statisticians (IAOS)
OECD Headquarters, Paris, France, 19-21 September 2018

Note:

The views expressed are those of the authors.

ABSTRACT

Quantitative information is multiplying within and outside government through expanding computational capacity. Used well, it is increasing the capacity to model processes, pathways and life-courses on a scale that could potentially transform both policy development and service delivery. Doing this in trustworthy ways is challenging the oversight of the methods and practices of measures used to influence society in the face of uncertainty, social change and economic disruption.

The growth in computational capacity has yet to be matched by an expansion of transparency and statistical thinking and capability, in part because of the often-limited consideration paid to the causes of error in data sources and their inherent limitations, as well as the black box nature of many developments. Coherence and confrontation of differences are integral to official statistics, as are error frameworks. This makes trial and error methods of policy execution especially important given the extent to which the limitations of modelling and data heighten uncertainty about the quality of policy choices at the same time as we see an increase the range of options.

Much can be judged about the integrity of many forms of evidence through their statistical properties and the practices adopted in production, management and accessibility. The ethical underpinnings integral to official statistics point to ways that an integrity centre can bring leadership of some form to evolving ethical practices and ensure that the means exists to hold to account policy makers and those who execute policy by the vindication of their methods.

The paper will discuss the relevant strengths of official statistics, emerging influences, and the changing frontier of quantitative evidence in public policy and its execution. The experiences of New Zealand in applying its integrated data infrastructure indicate the shape of the new highways and potential byways.

Keywords: Evidence system, Big data

1. INTRODUCTION

This paper is about charting a course in a time of rapid change. Quantitative information is multiplying within and outside governments through expanding computational capacity. Used well, it is increasing the capacity to model processes, pathways and life-courses on a scale that could potentially transform both policy development and service delivery. Doing this in trustworthy ways is challenging the oversight of the methods and practices of measures used to influence society in the face of uncertainty, social change and economic disruption.

The global velocity of circulation of anecdotes, beliefs, un-validated theory or just prejudice has also risen. These information forms have always competed with observational evidence not just in setting policy but also in determining whether to continue to gather evidence or invest in new forms. Where the quality of evidence in both policy formation and operational practice is poor, then citizens bear the risks of policy failure through both excessive fiscal costs and the personal costs and risks they bear as service recipients, for example because of screening errors from misclassification. In situations where the quality of the evidence base is recognised as not strong, the evidence obtained operationally by delivery staff and their autonomy on the ground must be expected to play a larger part in the effectiveness and efficacy of policy. Much established practice that draws on operational experience is not well codified.

How is this dynamic changed by the new data environment? A lot and a little. The availability of new sources suggests that new forms of analysis will generate new insight, but this will only come with a widespread adaption of data management disciplines, extensive collaboration and information sharing, and a willingness to look at some new ways of working. None of this, addresses any of the issues of institutional incentive and capability that can be problematic currently.

This paper is in four parts, the first is an examination of the inherent complexities in using data to improve policy analysis, then we discuss the way the evidence and data system functions. We then describe some of the way that this is playing out in New Zealand and offer some thoughts on the way ahead.

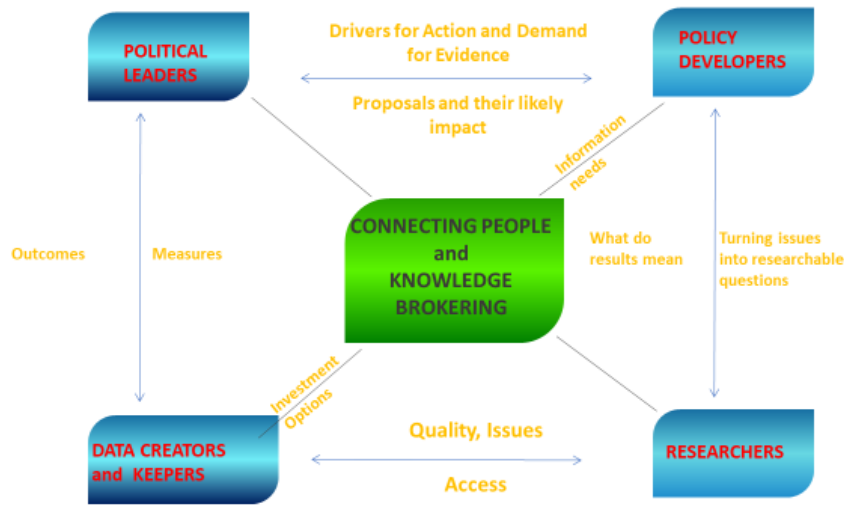
2. ASSISTING THE DEVELOPMENT OF PUBLIC POLICY HAS CONSIDERABLE INHERENT COMPLEXITIES

The evidence system must be influential (robust to political and institutional cultures), intellectually strong (methodologically and conceptually) robust and function (practical) as a coherent social system or network.

The evidence system is a social system that must convey some complex information among diverse communities

The development of policy has inevitably required rather specific pieces of information to be passed around a social system where all the participants have different knowledge and incentives. The following diagram (adapted from the Canadian review of the applied policy research system) illustrates this.

APPLIED POLICY RESEARCH SYSTEM



The policy research system must be able to manage the confrontation of diverse forms of evidence, to recognize the influence of diverse population structures on population pathways and work within different world views from the cultural make-up of the country. Operation complexity and operational failures limit what we know, and many variables are prone to respondent or agent variability. Diversity in the human condition can bring about selection bias, lead to incomplete/partial description of characteristics, and generate variation that remains untested or unaccounted for. Model misspecification that can lead to the spurious estimation of attributes will exacerbate any existing selection bias, and the impact will be dependent on the nature of any negative consequences on those who are falsely selected (false positives), and on those who should be selected but are not (false negatives).

Where the sources and dynamics of diversity and variability of the population are significant but not well understood, there is a risk that policy will be based on an unrealistic capacity to have an impact. Therefore, service delivery agents then have no option but to cope with an unanticipated complexity in the population targeted by the policy. Without the operational autonomy to adapt eligibility and entitlements where justified, the introduction of the services may undermine rather than increase public confidence in the service delivery agents, and perhaps the social services generally.

As an influencing system it must navigate the political world of conflicting interests, institutional incentives and diverse stakeholder drivers.

Emerging issues do not fit well with existing predominant agency structures, knowledge forms or the past focus of information investments. There are many intentional ways in which the dice are loaded by those conflicting interests and these can include;

- The certainty in results that politicians promise and defend are rarely likely to occur given the unforeseeable circumstances and conditions faced by people. This provides incentives at a political and institutional level to avoid programme evaluation and investment in analysis.
- The challenge of meeting tight fiscal objectives has increased incentives to obscure quality changes in services and measures of obsolescence of capital.
- The opportunity cost of current solutions cannot be estimated with confidence because of how the demand for fiscal accountability has displaced more relevant evidence gathering from contracted providers mainly from the non-government or community sector.
- An increased willingness to cause sanctions and penalties for non-compliance in service delivery perhaps enabled by information linking brings distrust in government among communities.
- Public trust issues arise from the apparent use of the same information for both facilitation and forms of population oversight.

Carole Weiss¹ observed the limits to which evaluation can be frustrated in building the potential connection between science, evaluation policy and its execution by the political management of evidence. She notes that evaluation must compete for attention but also can be suppressed or result in watering down of evaluation conclusions because it doesn't fit the political process.

“Evaluation research is a rational enterprise. But evaluation is a rational enterprise that takes place in a political context. Political considerations intrude in three major ways, and the evaluator who fails to recognize their presence is in for a series of shocks and frustrations:

1. *First, the policies and programs with which evaluation deals are the creatures of political decisions. They were proposed, defined, debated, enacted, and funded through political processes, and in implementation they remain subject to pressures-both supportive and hostile-that arise out of the play of politics.*
2. *Second, because evaluation is undertaken in order to feed into decision-making, its reports enter the political arena. There evaluative evidence of program outcomes should compete for attention with other factors that carry weight in the political process.*
3. *Third, and perhaps least recognized, evaluation itself has a political stance. By its very nature, it makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeable of others, the legitimacy of program goals and program strategies, the utility of strategies of incremental reform, and even the appropriate role of the social scientist in policy and program formation.”*

“Knowing that political constraints and resistances exist is not a reason for abandoning evaluation research; rather it is a precondition for usable evaluation research.”

As an intellectual system, applied policy research must engage in solutions for highly complex problems

If using data to help develop policy was purely a technical problem, it would still be challenging. The questions asked by public policy design and implementation, and the information base that can be applied to them are very diverse. Much is concerned with how people respond to changes in constraints and incentives. Predicting the outcomes from making changes to programmes is constrained by the complexity of the causal relationship at play and the ability of data to model the elements of the factors.

¹ CAROL H. WEISS: Where Politics and Evaluation Research Meet

This is illustrated in the following table.

Power of evidence base and evaluation options	Knowledge management and complexity
High certainty of continuity of execution (subject to periodic review) (PREDICTION)	<ul style="list-style-type: none"> • Strong knowledge base • Researchable issues
Structured ² evaluation programme (PROJECTION)	<ul style="list-style-type: none"> • Weak current knowledge • Researchable issues
Adaptation ³⁴ during execution (continuous improvement) (EXPLORATION)	<ul style="list-style-type: none"> • Strong knowledge base <ul style="list-style-type: none"> • Complex issues (evolving, enduring, causal)
Iterative ⁵ case review (Bayesian - multiple forms of evidence) (SPECULATION)	<ul style="list-style-type: none"> • Weak current knowledge <ul style="list-style-type: none"> • Complex issues (evolving, enduring, causal)

The top row is about where there is considerable capacity to be able to use data and models to manage programmes. Typically, these are about simpler decisions that people make about programme participation that are monitored over shorter periods of time. The bottom row represents the opposite extreme – trying to predict more significant life decisions over a long period of time with all the attendant challenges of understanding the range of factors at play and the problems of finding the types of data that help illustrate them.

Ideally, the knowledge base applied to the formulation of policy should be fit for purpose for the context of policy execution and the known uncertainty of the policy. This should determine the place, form and oversight of evaluation needed for the later, ongoing vindication of policy execution.

Conceptual complexity can also be found in the way that some key characteristics of populations limit the value of models in policy analysis and affect their relevance in service delivery. The information base may not enable the long-term dynamics inherent in events from the past to be recognised. For example, earlier cohorts of Maori had high rates of involvement with the youth justice system that would have influenced their later life propensity to be incarcerated as adults. Large birth cohorts produce a larger number of babies, but the certainty of predictions of later births may be reduced by the tendency of the age at first birth to be older for later cohorts.

The description of codified conditions used to screen or categorise people may be less stable than implicit in modelling, as the general fragility of the economic and social circumstances of individuals and their households can bring changes where the odds of seeing a reversal are not symmetric. For example, the economic position of households can fall quickly so that they takeout high interest “pay day loans” which

³ Rehabilitation of offenders

⁴ Homelessness

⁵ Parole board, sentencing options,

can take years to repay and return to a debt free situation because of the impact of interest rates of 100 percent or higher.

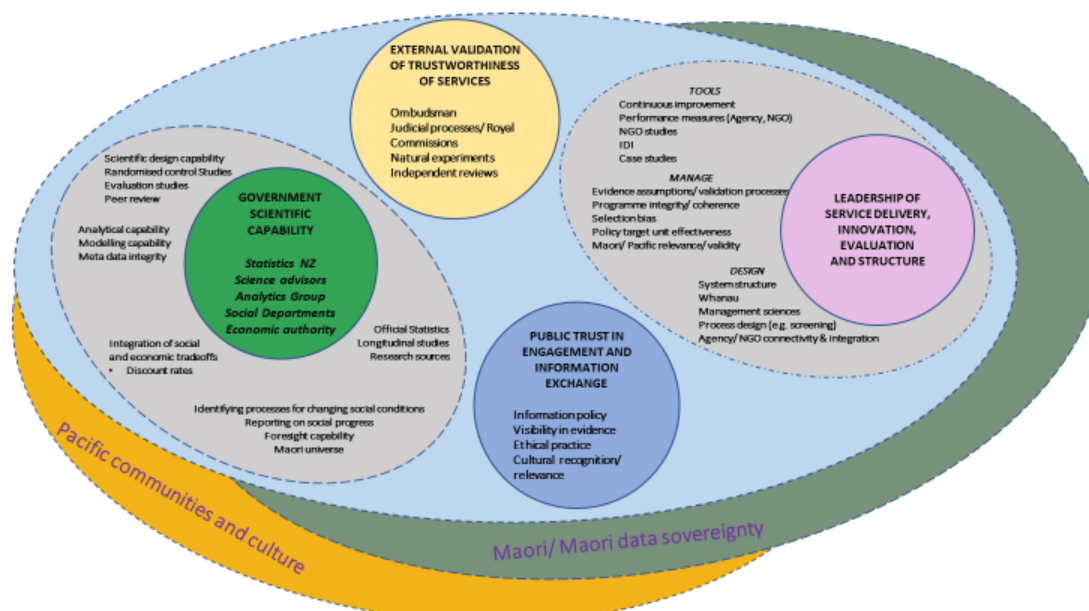
The diversity of people that results from ethnic origin, sexual orientation or generation may be obscured by the aggregation of defining variables into an analytically useful number of categories. The many distinct characteristics of the peoples of the Pacific (Tongan, Samoan, Niuean, Fijian) are often lost in the usually adopted broad classification of Pacifica. Part of the population may be systemically excluded from the modelled population. The higher than usual non-response rate to the 2018 census of population in New Zealand will not be evenly distributed across the population, nor will that of 10-15 percent of the country's household labour force survey or the 25 percent non-response to the household economic survey. Take up rates of social services are poorly estimated in New Zealand, but we have cases where they reach as low as 30 percent (disability support allowance). The privileging of one form of information (e.g. outcomes, performance measures, wellbeing) can lead to the limitations in the representativeness or randomness of models being ignored or weakly estimated, because of the inadequate confrontation with other sources.

3. AN OVERVIEW OF THE EVIDENCE SYSTEM

The wider institutional system for establishing trustworthiness

The judgements made about the reasonableness of public policy actions take place in a wider system of different sorts of authority, different bodies of expert knowledge and different sources of “public license” to operate as characterized in the following diagram;

Chart: The Spectrum of Evidence Sources for Ensuring Trustworthiness



The elements of the evidence system are quite dispersed and there is a need to understand on a government wide basis what resources and capabilities exist, and the potential role played by the mechanisms identified in the diagram as being how evidence gets to be trusted. So, there are parts of this system that are about the intellectual quality of what underpins the evidence. This includes the investment in inferential capability, connection to theory and understanding wider context of sources. The generation of evidence also depends

in substantial part on the capacity of scientific leadership in public agencies and the opportunity they must inform and influence management priorities. In any event, it is all too easy for the different parts of the system to work separately and to have different regard for the need to use mechanisms to ensure trustworthiness. There are some aspects of the evidence system that are driven by system wide drivers

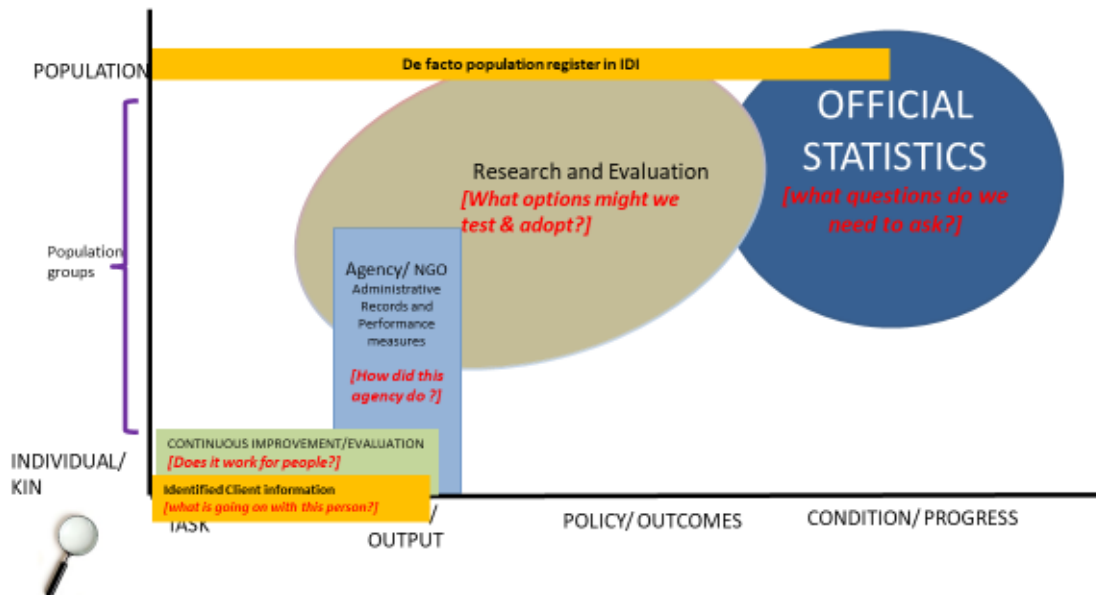
In the past the fragmentation of the evidence system has not been of significance, partly cause of the strength of arrangements for official statistics, and the limited opportunities that resulted from the way that administrative records had been managed. Trustworthiness was underpinned by a pragmatic and localized evolution of the means to assure trust prospectively, and a slow expansion of ways of holding government to account to vindicate the practices already in place. In the past the value of this was not well recognized. By the public nor sought by political and institutional leaders. The localized nature has meant that its evolution has been neither coherent nor focused on any sense of long term direction. Where change has been triggered by crisis, it has clearly lagged in meeting public expectations of trustworthiness. The diagram points to the core elements of a trustworthy evidence system and we suggest in the paper how a wider evidence system might be framed. While crisis has usually provided the imperative for the evolution of the evidence system in the past, foresight about the future context in which a government wide evidence system would operate should shape thinking about the direction and order of change. Such new thinking will inevitably challenge some past practice, as it seeks to bring a more coherent future approach. Perhaps the most difficult element will be gaining acceptance from politicians and institutional leaders of the importance of strengthening processes for ongoing independent oversight, before crises arise.

The Scope of Quantitative information

What is within the scope of quantitative evidence? The structure of diagram below illustrates one of the key challenges of the emerging data environment. Information that previously had been defined and managed in separate legal, methodological and operational ways are now being required to function as an integrated system.

The dimensions of this are articulated further below. The main challenge is that each of the elements of this pictures are all being challenged to meet a wider range of needs. Official Statistics must adequately reflect a world that is rapidly becoming more complex and dynamic. This must be responded to with various forms of system-wide investment in integration, meta data and coherence of sources. It will mean that oversight of operational processes potentially be extended by information management disciplines and practices. Although this will help understand the conceptual diversity of this information there will still be the need to deal with the considerable variation in how sources of error will range across different forms of evidence the broad scope of what might be available as evidence is articulated in this diagram.

Social services planned evidence purposes



There is one important consequence of the changes that are taking place that is illustrated by these two pictures. The opportunities that come with the expansion of data, methods and tools bring the likelihood of challenge about their integration and fit with the wider system described above. Because of a high degree of stability in its data sources and the way that its output was used, the official Statistical system and the other providers of evidence have tended to operate independently. None have had to rely on the wider system for authority or any source of license to operate.

Arguably this is changing now. There is more information being regarded as sensitive that was collected for operational purposes being brought together in databases. Questions are being raised about ethical limits and the potential need for wider societal governance of personal data. All these have led agencies like statistical offices to rethink their accountability to the public and consider looking to the wider system for forms of validation. The Privacy Commissioner has been able to be strong in publicly commenting on practices that are judged to not be in the public interest.

The challenges faced by long-standing methods and practices

The widening of data sources brings several challenges. Traditional sample survey methods and sources are facing challenges from problems in maintaining access to an increasing number of "Hard to reach" households, and concentrations of "alienated" communities who simply avoid wanting to be included in the statistical measures on which we base decisions that affect them. There is an increasing general survey

burden on households, while funder expectations of having access to traditionally confidential information about respondents

The dynamics of population frames and classifications has brought a need for a more rapid adaptability of the meta data infrastructure that underpins official statistical frameworks and continuity of measures. The dynamics, diversity and variability of business and people challenge the validity and robustness of statistical units and classifications that underpin official statistics. The world is now at our back door, so that national boundaries have less meaning in determining relevance to the population base and wealth generating capability. Several significant shifts have taken place that are both challenging the traditional elements of a national statistics system and providing greater opportunity to gain leverage from the fundamental strengths of the NSO.

Institutional Challenges

At an institutional level, official statistical offices have experienced a declining pre-eminence in the overall provision of information that informs personal, business and public choices. There are new roles for traditional information brokers, and new entry points to become an information broker. In the past analytical strengths underpinned most information brokers (scientists, universities, journalists, international (OECD)), but the core capability of newcomers is at least as likely to result from the unique sources that they oversee, integrate and distribute. "Evidence based policy" initiatives of government have foundations in and beyond the government research community, to varying and inconsistent effect.

We see in many places strong community research institutions continue to be founded on analysis of data, we also observe a great variety of weakly connected institutions engage in publicly credible research. Consequently, the sound ethical base for protecting confidentiality that exists through maintaining well founded practices that protect obligations to uses is challenged by the range of sources, their integration and multiplicity of presentation forms. Data Protection legislation plays an enlarged part in cross agency function. However, positive influences the public climate and debate about information access, by valuing and advancing access to information for research and public service, can be undermined by finding fraud and illegal activity, from specific identification. We need a strengthened legal base for this.

Valuing the traditional strengths of official statistics

There are enduring strengths of the way that Official Statistics has always been conducted that will provide at least a foundation for new structures that will be needed in the future;

- A legal infrastructure that defines the obligations and authorities about what data can be acquired and how it is used the articulation of the guarantees that are available to people providing data.
- Institutional assurances of the independence of the work of the Official Statistics system
- A long history of operating transparently, making the limitations of the information outputs clear
- Mechanisms to design coherence and integration complex derived outputs. Standard reference populations, standard methods and standard variable definitions enable coherence to be produced from diverse and idiosyncratic data sources.

The extension of administrative data access has greatly increased the available range of accessible microdata about populations. There is a high opportunity cost in failing to make full use of potentially available microdata. This is most significantly seen in inadequate identification and analysis of policy options. We cannot now doubt the weight of the argument that inadequate use of microdata has high costs. This may be because we now see as an everyday condition the expectation that micro data analyses will usually inform policy analysis and policy evaluation. Enabling research on microdata is now recognized as a strong part of any official statistics system, and the extent of this is a longstanding distinguishing characteristic of the UK statistical system, compared to most countries including New Zealand.

Similarly, the expectation that quality measures of some form will be provided alongside statistical output has promoted awareness of the need to use data with the possibility of measurement error in mind. Many of these measures from statistical offices have been the of sampling and non-sampling error derivable from the knowledge of the sample survey design of the survey. The quality of statistics produced by composite sources such as price indexes and national accounts or demographic projections has not usually been explained quantitatively but the consistency and transparency of production methods enables users to examine quality concerns effectively.

Opportunities and challenges of the new data environment

Opportunities

- The ability to link records creates the opportunity to model the operation of government in a wide range of fields, and improve services to citizens especially where multiple agencies are involved
- The opportunity to link individual records greatly widens the range of cohort studies of lifetime experiences with sectors of government such as health, education and justice are feasible.
- The use of administrative data to create records of individuals engagement with service providers have enabled linkages across policy domains to be established, creating possibilities for more informative segmentations of the population and enabling evaluation to consider a broad range of outcomes.
- Enriched sources and analytical capability widen potential scope of analysis and increase value of existing sources
- The use of predictive modelling techniques in a wide range of contexts have given agencies options in optimising their operations, although weak transparency has reduced trust them in situations involving the authority of the state.
- The capacity to use large data sets together has created possibilities of fitting forms of modelling that directly estimate people's responses to incentives.
- The scope and comprehensiveness of the overall data system increases the resilience of the knowledge base to shocks from one off sources

Challenges and risks

- The integration of information from multiple sources has brought several challenges and limitations to the representativeness, randomness and realism of the data in the resulting datasets and knowledge of their quality.

- The information obtained is often a crude proxy for characteristics of importance, and in different sources may vary in how they approximate the desired characteristics
- Screening errors cause misclassification and will have different effects in alternative sources
- Frame, unit and conceptual differences across administrative sources, differences in administrative programme coverage rates and the uncertainties around linking successfully making effective discussions about data quality difficult.
- The lack of consistent practices for the documentation of administrative or statistical source data over time both within and among agencies.
- The absence of meta data standards across government for many categorized attributes and characteristics.
- The near impossibility and high costs of retrospectively applying meta data standards to bring consistency to the information in past records.
- The importance of personalised knowledge in retrospectively documenting the context when information was obtained.
- Creating well informed expectation about the potential of new analysis from new sources using new techniques.
- The potential for privacy or confidentiality breeches as personal information is accumulated even in the absence of specific identifiers.
- Increasing the usage of microdata while ensuring that the essential access disciplines that underpin data security can be maintained.
- New approaches to data analysis being associated with the policy focus of the government in power at the time they were introduced, leading to a weak form of data analysis being politicised, especially if this was associated with methods to select individuals for targeting

4. HOW ARE WE DOING IN NEW ZEALAND

Broad Context

Most of the comments in this section of the paper relate to the use of the New Zealand Integrated Data Infrastructure (IDI) as it is the main new information resource that has been generated over the last ten years. The IDI combines over 40 different agency datasets, using a variety of linking techniques (New Zealand does not have a unique person identification number) to create a dataset that covers something like 90% of the interactions that the Government delivers to a named person in New Zealand (this excludes where services are contracted out to a non-government organization. The IDI has now become the core source of information in public policy development.

In the big picture we are now seeing strong growth in microdata studies and modelled outputs particularly in the social services and justice sectors. This has amplified the need to understand the detail of how our administrative data sources came about and have evolved.

Impact of new data and new analysis on Policy Development

- The advent of the IDI has enabled more informative population segmentation to take place. The capacity to construct a history of an individual's interaction with government services throughout their

life has been especially helpful in understanding vulnerability, both in a wider context of the range of services of government, and how these clues about vulnerability emerged over time. This sort of analysis contributed to the re-organisation of Child protection services in New Zealand.

- The discipline of charting how people flow in and out of services and in and between different types of services has proven very useful in examining how resources have been aligned to the size of different groups in different policy programmes and the achieved success in achieving favourable transitions. The Justice sector completed a major descriptive modelling exercise which lead into insights about which categories of engagement with the Justice sector were relatively well funded.
- Using the full range of metadata available within operational programmes to maximise short term indicators of progress has also progressed well. Several agencies have been able to demonstrate significant efficiencies from using predictive modelling approaches. This has also lead to concerns about algorithmic transparency and Stats New Zealand, in its new role as Government Data Steward (see below) undertaking a stocktake of algorithmic decision-making practices across New Zealand Government.
- Looking for direct evidence of the success of data analysis in estimating the impact of policy changes does not yield much. In *“Computerised models: tools for assessing the future of complex systems”*, Martin van Ittersum and Barbara Sterk make the useful observation that people do not tend to write up attempts to predict new policy outcomes that didn’t work, and similarly the data analysis that prevented a policy proposal with a false underlying premise from proceeding is again unlikely to be published. There are fewer available success stories about successfully predicting the outcomes of significant changes to programmes. It has been suggested that different forms of modelling, different types of data and some well-chosen areas of application will see successful prediction occurring more often, but this is still to eventuate.

Impact on Institutional Data Governance Arrangements and Researcher Collaboration

While the more institutional challenges identified above are significant there are reasons to be optimistic;

- The experience of Statistics New Zealand in managing its Integrated Data infrastructure (IDI) has been that it possible to create “virtuous cycles” of information sharing. As agencies make more use of their own data in the context of it being linked it to other agencies data, the drive to standardise practices arises out of the need to solve their own business problems
- In New Zealand it has become clear that individual researchers who are more active in maintaining personal networks are doing better in making good use of linked data. They can overcome dependency on agencies and individuals within agencies to get enough metadata and general intelligence about using data files to produce work that crosses agencies and sectors. Stats NZ has had to put active effort into trying to build
- Several approaches to directly involve critical communities of citizens have been initiated including building ongoing connections for challenge and possible collaboration with the Maori Data Sovereignty Network Te Mana Raraunga⁶.
- In New Zealand it is possible to get representatives of most if not all the key players in one room, and to date such interchanges have demonstrated the scale of collective learning needed, the diversity of issues involved in developing a cohesive approach, and the iterative nature of the path ahead.

⁶ <https://www.temanararaunga.maori.nz>

- The need for some more formal mechanisms across Government to achieve the potential of the new data sources has been recognised by establishing the Government Statistician as the New Zealand's Chief Data Steward. In this role Stats NZ is undertaking several more "wider data system" initiatives. These include;
 - Establishing a body across Government to identify critical standards and negotiate these with relevant agencies
 - Develop in conjunction with other agencies a range of wider Data Governance initiatives to develop the principles that will apply to a range of how data is acquired and used by Government agencies in operational and non-operational contexts
 - Developing a Data Roadmap for New Zealand to help identify needed data related infrastructure.
 - Undertaking specific initiatives (like the Algorithmic Transparency work already mentioned) to provide operational guidelines and facilitate useful knowledge sharing

Technical Lessons

- Stats NZ is making considerable investment in looking how to build error frameworks that provide insight into the limitations of using linked data. This framework uses linkage errors, coverage errors and classification errors as the dimensions that underpin quality measurement.
- There is increasing understanding of how to address the challenges identified above around the coverage of administrative sources. A series of papers have been published on the Stats NZ articulating what has been learnt towards solving these problems
- The emerging recognition of the role of traditional data sources has been striking, after what might turn out to be a transition period where they seem to have less importance than in the past. It is extracting insight from administrative data sources that it can only be done confidently with a lot of knowledge about what might explain observed variations. The well understood and validated context of data collected more traditionally helps provide an essential point of reference in deciding whether the observed admin data patterns are credible.
- The interesting possibilities of using survey and admin data together are starting to emerge. The general point is that we are just beginning to think about how to maximise the use of the administrative data in designing surveys, how to design the content so that it can complement the existing data.
- The need to incentive agencies to develop a greater focus on Meta data management has been a challenge. Practices are variable in quality, and many data sources have previously not been subject to any significant usage. The knowledge about sector specific databases by experienced staff in is often poorly documented and therefore limits the range of information known about the characteristics of integrated datasets. Experience has shown that there is no substitute for the originating agency starting to become a significant user of their own data, but the current limitations can constrain the understanding of the files quality risks.

Capability Challenges have loomed large

There has been significant investment in creating Analytics and Insights units in Government Agencies, however, in areas of long standing reliance on analytical competence, arguably where weaknesses have

existed they remain. The IDI and its analytical focus have not provided spinoff benefits from a more broadly-based uplift in analytical capability.

Some of this may be attributed to the loss of high level leadership of analytical work that has resulted from a twenty-year focus on generic management appointments where analytical expertise and sector experience are weighted low in defining merit for appointment. We have yet to see training programmes to lift the capacity for leadership of analytical resources at executive level. Until that happens, there is a risk that analytical capability will remain in many agencies as a commodity to be hired rather than a resource to be nurtured.

The Long Term

In terms of the long-term improvement in the development of public policy questions that have been raised include;

- The optimising work that has been flourishing functions by short term feedback loops that form a series of “mini-evaluations”. The interesting question is whether this can be extended to form the basis of a more adaptive approach to policy development by explicitly relying on the ability of predictive modelling of programme paradata to make an extended series of iterations towards a better understanding of what works. It is not clear how to determine the limits of this type of approach, but it does create an environment where it is easier to “learn by doing”.
- It is still not clear about the best way of obtaining confidence about the likely impact of major changes to programmes. For example, in a range of situations, there is well evidenced enthusiasm for investing early in the life cycle. The main difficulty with this being that it requires this expenditure to occur at the same time as the money currently being spent mitigating ill-effects of not investing. How can confidence be developed that investment really will result in the need for mitigation expenditure? It remains hard to estimate large scale change. Might it be possible to design a sequence of transitions that are each closer to the sort of change that is more likely to be predicted?
- Nothing about new data sources, tools and methods creates institutional change. However, a wider expectation that “modern” methods will be implemented in agencies can be an impetus for changing the role of evidence in policy formulation, and Statistics NZ has had to decide how far to go in helping build new capabilities in government. For example, there has been enthusiasm for Stats NZ to lead a Government Analytics Network.
- The leadership at the highest level in organisations and their curiosity and interest in defining researchable questions of real importance will determine the pace and intensity with which we can clarify the limits of data analysis in a policy development context.
- The availability of a central data source like the IDI creates the possibility that some programme evaluation can be done outside the agency that runs the programme. There have been instances of this happening in New Zealand and it has been interesting that this has created pressure for the agency running the programme to do enough evaluation so that it knows “at least as much” as external stakeholders.

5. WHAT ARE THE FUTURE DIRECTIONS

Drawing on our general reflections and looking at our experiences in New Zealand we suggest that the following matters will be critical to making progress.

There will always be significant constraints from evidence forms and wider institutional pressures. The trick is to make the system that manages them as transparent as possible. We will make this happen by:

A. Articulation of the broader evidence system and a willingness to renegotiate roles and responsibilities

The reality that agencies will want to be reconsidering the scope of their roles and that this will require them to negotiate with a wider system of obtaining license to operate needs to be recognized. Official Statistics will have plenty to bring to this table, but we are operating in a context where we are engaging in different ways with our wider eco-system and we need to expect and accept different types of challenges to the way we want to operate.

B. Explicit articulation of the new possibilities that are being realised, and what it is proving to be critical

In a world where there is excitement about new possibilities and a need to avoid creating unrealistic expectations it is critical that knowledge can be shared efficiently and quickly be translated into new norms. There is a need to invest in identifying what is currently being achieved, what are proving to be constraints and how useful expectations about the limits of what can be achieved is identified. For example, we need to develop the capacity to talk usefully about error structure of new sources (linked data etc.), how to assure coherence in a world of diverse data sources and how to have more extensive data catalogues.

C. Continuing learning how to make existing data useful.

More specifically, we may expect the value of administrative data to grow in the next decade as we will see huge shifts to the population base and business organization resulting from ageing, diversity, mobility and globalization. We see this being associated with further developments of integrated databases built up from administrative data. This includes:

- *Extension of known good practice*
 - Strengthening government wide meta data management
 - practices to integrate historical background information about constituent datasets
 - Identifying critical variables which are context determined for specific attention in information management over a life course
- *Making feasible practices more common*
 - Providing rules of thumb for users to assess the fitness for purpose of integrated datasets for the research questions that they have.
 - Identifying a portfolio of types of researchable questions that the integrated data can address well.
 - monitor the increasing share of public policy that will be targeted at recipients, or through new vehicles of delivery.
- *Broaden understanding of established methods that remain relevant*
 - Demographic methods
 - Continuous improvement
 - Operations research

- Sampling and experimental design
- Time series analysis
- *Identifying new methods*
 - Investing in developing a framework for the measurement errors associated with linked data files
 - Considering how the design of statistical surveys can be refined to maximise their synthesis with administrative data.

6. CONCLUSION

It is tempting to look at some of the difficulties with achieving influence in policy development by using data analysis in the past and hope that the advent of richer data sources and new analytical possibilities will provide the mechanism for bypassing previous difficulties. The message of this paper is that there are great reasons to be optimistic but that results will only come from good mechanisms for visibility of data sources, methods and practices. Reliance on analytical competence remains a long-standing concern, and where weaknesses have existed they remain. Initiatives to develop such competence may well be foundering through the limited capacity to lead and promote them by the limited research understanding and experience of generic managers in influential leadership roles.

The inherent difficulties of policy development cannot be underestimated. Whether viewed as a social system that must navigate some diverse communities, an intellectual system that must solve problems of “wicked” complexities or an influencing system that must navigate the vagaries of political stakeholder interests there has always been challenges well beyond getting the right data to the right person. These interests fail citizens particularly badly when they prevent the monitoring and evaluation of the execution of policy, given that policy selection is often be done with little recourse to the available evidence.

The possibilities of the new data world similarly come at some cost. New data sources need to be connected the previous body of knowledge if they are to provide constructive insight, new methods must be used with a clear view of their potential and their limitations, the limits of public acceptance must be established, and diverse and disparate knowledge must be made available to people making contributions.

The approaches of Official statistics, (but maybe not some of our practice) are well placed to become a hub in the new, expanded applied policy research system. Achieving value will require a willingness to create new functions and to work with a wide range of partners to make the system. It may also mean accepting that we must negotiate our authority to operate from our wider environment.

In an organization, there is no substitute for high level commitment and engagement to empower the curiosity of those involved in policy or operations to raise issues that can be turned into researchable questions requires. System wide protocols and conventions are needed to set the context for addressing those questions in an open government system so that the initiation and publication of an expanded range of research and analysis does not depend on the strengths and persuasiveness of individual managers. Understanding the uncertainty that surrounds any evidence base is vital for its effective application in policy analysis, and in setting the operational parameters that inform the execution of policy.

The trustworthiness of the past fragmented evidence system has been underpinned by a pragmatic and localized evolution of the means to assure trust prospectively, and ways of holding government to account

retrospectively which vindicate the practices put in place. The localized nature has meant that its evolution has been neither coherent nor focused on any sense of long term direction. Where change has been triggered by crisis, it has clearly lagged in meeting public expectations of trustworthiness. This paper has sought to identify how thinking about a wider evidence system might be framed, and the imperatives that will inevitably bring it about.

Where the quality of evidence in both policy formation and operational practice is poor, then citizens bear the risks of policy failure through both excessive fiscal costs and the personal costs and risks they bear as service recipients. As the velocity of circulation of anecdotes, beliefs, un-validated theory or just prejudice has also risen globally, the influencing capability of the evidence system to compete, which will become less likely if these forms of information are also used to limit the nature of the evidence base and the practices which underpin its integrity. The matters in this paper are not merely about official statistical practice but they involve ultimately the trustworthiness of the basis of government decision-making among diverse communities and groups with conflicting interests.