

**Confidential**

# **Role and limits of performance measures**

**Report of the Performance Measurement Research Project for the Technical Working Group**

**Report to Road Maintenance Task Force - Better Asset Management, Planning and Delivery**

**February 2012**



## About NZIER

NZIER is a specialist consulting firm that uses applied economic research and analysis to provide a wide range of strategic advice to clients in the public and private sectors, throughout New Zealand and Australia, and further afield.

NZIER is also known for its long-established Quarterly Survey of Business Opinion and Quarterly Predictions.

Our aim is to be the premier centre of applied economic research in New Zealand. We pride ourselves on our reputation for independence and delivering quality analysis in the right form, and at the right time, for our clients. We ensure quality through teamwork on individual projects, critical review at internal seminars, and by peer review at various stages through a project by a senior staff member otherwise not involved in the project.

NZIER was established in 1958.

## Authorship

Prepared by:	Derek Gill and James Zuccollo
Quality approved by:	Susan Hitchiner (independent QA) and John Ballingall
Date:	22/02/2012 7:28 p.m.
Version:	Final
Acknowledgements:	Interviewees and workshop participants. Susan Hitchiner for facilitation of the workshop and peer review of the draft report.

8 Halswell St, Thorndon  
P O Box 3479, Wellington  
Tel: +64 4 472 1880  
Fax: +64 4 472 1211  
econ@nzier.org.nz  
www.nzier.org.nz

NZIER's standard terms of engagement for contract research can be found at [www.nzier.org.nz](http://www.nzier.org.nz).

While NZIER will use all reasonable endeavours in undertaking contract research and producing reports to ensure the information is as accurate as practicable, the Institute, its contributors, employees, and Board shall not be liable (whether in contract, tort (including negligence), equity or on any other basis) for any loss or damage sustained by any person relying on such work whatever the cause of such loss or damage.



## Executive summary

The Road Maintenance Task Force is charged with improving the whole-of-life value of road maintenance operations and renewals. It sought some independent advice from NZIER based on a survey of the literature and interviews and workshops with practitioners on ways to improve the performance measurement system.

### What is the imperative?

New Zealand was a world leading innovator in the 1990s and early 2000s in both asset management and performance measurement in roading maintenance. Since then progress has levelled off at a time when information technology and data analysis techniques have continued to develop rapidly.

There is an imperative for change because of the flat lining of maintenance budgets at a time when the network is still expanding and the cost of maintenance expenditure is increasing. New Zealand has the opportunity to achieve a cultural shift that would allow performance information to shape decision-making on roading maintenance. The road transport sector has been a pioneer in the past and it has both the hard systems required (the data and capability) and the soft systems (a history of collaboration) to be so again. This report sets out the strategies required to regain the position of pioneering world leader.

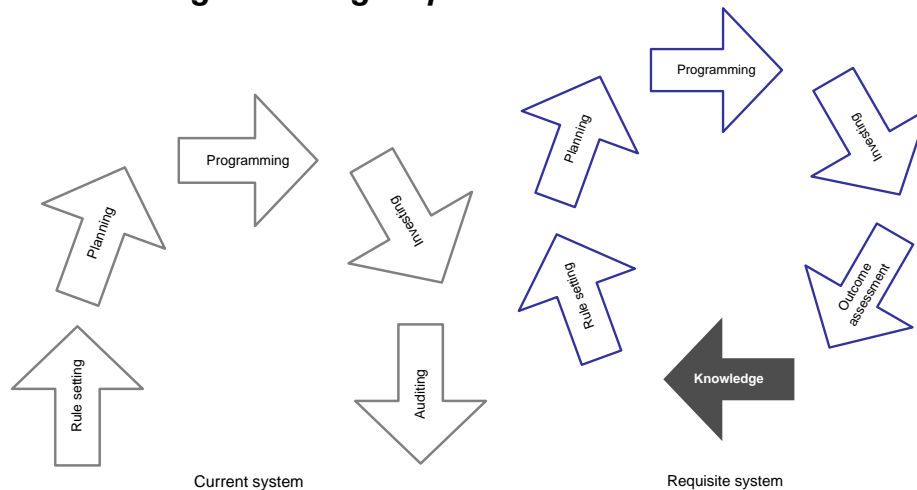
### What did we find?

The research identified a number of common themes including the need for:

- a common national performance framework including
  - common data protocols
  - enhanced measures
  - common standards on levels of service
  - improved data infrastructure to enhance data input and user access
- focus upon measuring the effectiveness of road maintenance interventions
- learning through communities of practice
- enhancing the capability of people to work in a different way
- support by leadership, engagement, focus and commitment.

### How do the formal regime and real in use system operate?

We have identified two major problems with the formal measurement regime applying to road maintenance. The first problem is that there is no cycle of performance measurement as shown on the left hand side of Figure 1. Rather there is an open loop with limited direct and indirect feedback from the monitoring and measurement subsystem to the design of programmes and policies. As shown on the right hand side of Figure 1, this lack of a closed loop is impeding performance improvements being identified and implemented. The lack of measurement of the effectiveness of interventions, such as pavement treatments, inhibits learning and the development of asset management expertise.

**Figure 1 Creating a learning loop**

Source: NZIER

The second problem is a disconnect between the data supplied and the presently unmet demand for different data and more measures. While there is a formal process, this is very shallow and focuses on a few core, objective measures such as roughness, that do not adequately capture overall performance. There is unmet demand both from Roading Control Authorities (RCAs) and NZTA for a wider range of better measures—partly to enable more effective investment allocation—but also to improve learning about effectiveness.

The practice in the sector is to work around the disconnect using systems based on informal collaboration: RCAs that use the RAMM database software allow NZTA to access their raw data and use it to construct measures. However, each RCA has discretion over their data collection, measures used and standards of service so RCAs' practices vary greatly. That makes the data collected through informal collaboration difficult to use for research and comparison purposes.

This leaves gaps not remedied by the informal practises such as data not stored in RAMM, information on the effectiveness of treatments, data on gravel roads, bridges etc. The lack of a national framework covering what should be measured (standards and associated measures), and how it should be measured (data protocols), inhibits learning.

### What are the features of well-designed measurement regimes?

The current system of performance measurement practiced in the road maintenance sector has a number of positive characteristics including:

- the high agreement that pavement condition is amenable to measurement
- there is a coherent and well-embedded culture of data collection and use of measurement accepted across the sector
- integration of measurement into planning, budgeting, programming and contracting systems
- a tradition of collaboration which allows sufficient scale of operations to spread the costs of designing and operating the measurement system.

Achieving a high performing road maintenance measurement system requires addressing a number of the weaknesses identified in our research. The recommendations identify the need for a common national performance framework for road maintenance and to enhance the capability of the people working in the system. In addition to **what** needs to be done, **how** that work is undertaken is critically important with:

- greater emphasis on engagement so key stake-holders buy-in to the development of measurement frameworks
- a culture of use of performance measures to learn about effectiveness
- leadership that encourages dialogue over what the evidence is saying.

### So what is to be done?

On improving performance measurement and management:

- get some quick wins using existing data and analysis platforms to review the effectiveness of road maintenance treatment projects.

Develop a common national performance framework for road maintenance including:

- protocols for how road maintenance data needs to be defined, collected and stored
- a suite of performance measures to support oversight of investment in roading maintenance
- a system of national standards for levels of service required for road maintenance
- improvement in the platforms for data input and data access.

Require all future roading maintenance applications to the NLTP to identify the strategy for learning about the effectiveness of the proposed maintenance investment.

On identifying the limits of performance measurement and reducing the risk of perverse behaviours:

- design systems collaboratively with managers and professionals in the RCAs and the industry to promote broader ownership, support, trust and utility
- use performance measurement to support learning so performance information can be used as a part of an intelligence system. There should be limited use of league tables while targets should be avoided completely.

On identifying and improving the take up of innovative practices:

- review the lessons learned from RIMs (along with comparable models in other sectors and jurisdictions) to develop new models of collaborative learning
- apply the collaborative ways of working to better facilitate the sharing of lessons learnt about building people capability.

# Contents

<b>1.</b>	<b>Introduction .....</b>	<b>1</b>
<b>2.</b>	<b>Themes from the performance measurement literature .....</b>	<b>2</b>
2.1	Overview of the literature .....	2
2.2	One size does not fit all .....	4
2.3	Citizens' perspective on performance .....	5
2.4	Use is the Achilles heel of performance information .....	6
2.5	Perversities of performance measurement.....	7
2.6	Perversities of performance targets.....	9
2.7	Qualities of good performance measures.....	10
2.8	The qualities of good performance measurement systems.....	12
2.9	The limits of performance measurement .....	14
<b>3.</b>	<b>Analysis of the performance measurement regime for road maintenance ....</b>	<b>16</b>
3.1	Our approach .....	16
3.2	The formal system of performance measurement.....	17
3.3	The real system in practice .....	26
3.4	Why do the real and formal systems differ? .....	28
3.5	Conclusion .....	33
<b>4.</b>	<b>Recommendations for improving performance measurement .....</b>	<b>35</b>
4.1	Strategies .....	35
4.1.1	A common national performance framework .....	35
4.1.2	Data .....	35
4.1.3	Measures .....	36
4.1.4	Standards of levels of service.....	36
4.1.5	Data infrastructure.....	37
4.1.6	Initial focus upon measuring effectiveness of treatments.....	38
4.1.7	Learning.....	38
4.1.8	People capability .....	39
4.2	Supporting strategies – engagement, focus and persistence .....	39
4.2.1	Engagement and partnership .....	39
4.2.2	Focus .....	40
4.2.3	Persistence .....	40
4.2.4	Directions for reform.....	40
4.3	Improving performance measurement.....	42
4.4	Identifying the limits of performance measurement and reducing the risk of perverse behaviours .....	42
4.5	Identifying and improving the take up of innovative practices .....	43



## Appendices

Appendix A References .....	44
Appendix B – People met.....	47

## Figures

Figure 1 Creating a learning loop .....	ii
Figure 2 The production of public value .....	2
Figure 3 Programming and budgeting for road maintenance .....	20
Figure 4 Continuum of contracting .....	24

## Tables

Table 1 Three implications of performance measurement.....	9
Table 2 Conformance gaming and cheating .....	10



# 1. Introduction

NZTA (on behalf of the Technical Working Group or TWG) asked for some independent advice that comments on the role and limits of performance measures. This advice was to include a review of the current performance measurement regime used in the roading maintenance sector and a set of recommendations to improve performance measurement in the sector. This advice is to be used to inform the TWG proposals to the Road Maintenance Task Force on improvements to road maintenance and renewals. The research scopes for the Task Force that relate to this report are:

- are the sectors' current indicators the right indicators, measured at the right frequency, for predicting future maintenance and renewal requirement and for benchmarking – and if not what indicators should be used?
- what perverse behaviours are driven by the Road Control Authorities' (RCAs) performance measurement regime?
- what innovative approaches (including international) to performance measurement and monitoring are there that could be used to help RCAs improve performance in maintenance and renewal?

We have worked closely with the TWG and NZTA to understand the current formal performance measurement regime and what the 'real' regime is that practitioners actually use. Where possible we have built upon existing research findings or research underway (in particular the draft NZTA research report on Performance Indicator Analysis by Hemming et al).

## 1.1 Our approach

Our approach to reviewing the current performance measurement regime was to seek to understand 'who wants to know about what for what purpose?' We started from the proposition well established in the performance measurement literature that there are many different perspectives on performance and different users of performance measures require different performance information for different purposes. This report is based upon:

- a desk-based review of the performance measurement literature (discussed in section 2 of this report)
- a summary of the formal performance measurement regime in place in the New Zealand roading sector applying to maintenance (section 3 of this report)
- an assessment of the real regime that practitioners actually use based on one-on-one semi-structured interviews with practitioners in RCAs and other private sector organisations and a workshop with a range of industry players listed in Appendix B (also discussed in section 3)
- a set of conclusions and a suite of recommendations for improving performance measurement in the sector based on a dialogue with the Task Team of the Technical Working Group on 27th January (section 4 of this report).

## 2. Themes from the performance measurement literature

The scope of this section of the project to review the international literature is largely based on the references already identified in earlier work by the author (Gill 2011 Ch2) with research assistance from Tyson Schmidt and Nestar Russell. In addition to this, the NZTA has supplied some work underway on the development of performance measurement frameworks which provided additional references particularly relating to roading.

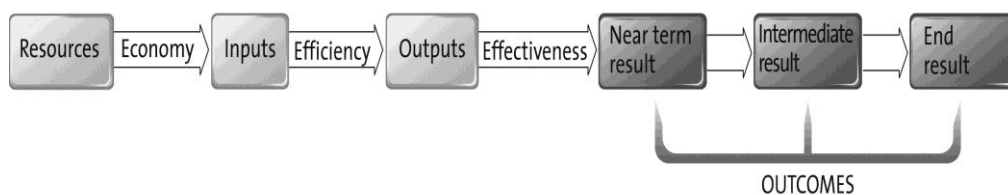
This section draws out some of the themes in the material surveyed relevant to roading maintenance. Most of the literature relates to use of performance measures within an individual public organisation whereas the roading maintenance sector includes national and local government as well as consultants and constructors. However the themes were selected based on their relevance to the research problem concerning good practice performance measures for roading maintenance. The intention of this section is draw out some lessons that provide practical insights for roading maintenance sector. Where appropriate at the end of each section, we have included a list of references highlighted as 'Recommended Further Reading'. A list of all references used is provided at the end of the report in Appendix A.

### 2.1 Overview of the literature

There is a large, rapidly growing, literature on performance measurement that draws on a range of academic disciplines. The literature is fragmented and the fractures extend beyond the academic literature into the practitioner discourse.

The term 'performance' has been the source of debate among both academics and practitioners about the 'cause and effect' relationships of the public production process. Performance, a practitioner's term drawn from the industrial model of the public production process, is shown diagrammatically below in Figure 2.

**Figure 2 The production of public value**



Source: Gill and Schmidt (2011 Figure 2.1)

Figure 2 covers the stages in the production chain – the inputs, outputs and outcomes – and the performance relationships between these stages – economy (converting financial resources into inputs), efficiency (turning inputs into outputs) and effectiveness (outputs converted into results).

There are a number of criticisms of this approach – see the discussion in Gill and Schmidt (p11). Attempts to measure performance are predicated on a number of strong but often unarticulated assumptions identified by Hodd (2007) as:

- that performance that can readily be measured adequately represents overall performance (synecdoche)
- that measurement error is not material
- that the distribution of performance is less important than the central tendency
- that there is no significant cheating and gaming.

There is generally a disconnection between the academic literature and practitioner discourse. In performance measurement, with some notable exceptions (Kaplan & Norton (1996) *The Balanced Score Card*), the grey or practitioner literature generally bypasses the academic literature about the limits of measurement and discusses 'best practice'. One reason for this disconnection is a paradigmatic difference in the view of the world. In the rational objective view "there is an objective reality 'out there' called performance just waiting to be discovered" (Thomas, 2006, p174). By contrast the alternative subjectivist perspective suggests that public performance is subjective and negotiated, being influenced by the distribution of power in society and therefore fundamentally political in nature. Some authors take the relativist view that performance should be viewed as institutionally defined since it is those in power who control the interests being pursued, "At its extreme, it can be argued that there is not and can never be any objective measure of performance as it is a purely ideological device" (Talbot, 2008, p 143).

Moynihan's interactive dialogue approach starts from the premise that "performance information itself is not comprehensive or objective but incomplete and ambiguous. It is subject to selective measurement, presentation and interpretation based on the interests of the actors involved" (2008 p24). de Bruijn (2006, p56) uses the metaphor of floodlighting a statue: "Illuminating the statue from several different perspectives creates a different image each time. Each image is correct but a single perspective always gives a distorted image." In sum, good performance is a social construct not 'out there' waiting to be discovered. Performance measurement has been asked to serve numerous purposes which are both 'political' and 'administrative' in character.

Performance measurement has been part of a wider 'movement' to improve public sector performance (that academics have labelled New Public Management (Hood 1991)) that is widely seen as having peaked (Lodge and Gill 2011).

### **2.1.1 Implications for the Road Maintenance Task Force**

The above discussion suggested there is no one single performance but many performances. While roading organisations need to take a positive but sceptical view of the role of performance measures, they should take a negative view of simplistic claims for one set of "best practice" measures that can be directly imported from overseas and applied in New Zealand.

## 2.2 One size does not fit all

There is one proposition on which the academic literature is nearly unanimous and that is the view that there is no 'one size fits all' best practice performance measurement system. For example, de Lancer Julnes observes "Different audience, different purposes and different users require different kinds of performance information" (2008, p171). Bouckaert and Halligan observe "there is general acceptance that there are different performances for different purposes that result in different responsibilities covered by different mechanisms of accountability for performance" (2008, p174).

This is in marked contrast with the grey practitioner literature which tends to emphasise 'best practice' and the associated notions of the golden thread, or one integrated set of performance information (SSC & Treasury, 2008).<sup>1</sup> Commenting on best practice Pollitt and Bouckaert (2003) suggest "what we are dealing with here is best described as a kind of a religion, a system of belief founded on faith" (sourced from OECD 2009 p36).

The performance measures used at the macro, sectoral or organisational level may well be different from those of programme, which in turn may be different from the qualities desired by a customer or client. And what is 'good' will in turn depend on whose perspective is being privileged. For example, the measures used at the national system-wide roading level, for organisations (RCAs, NZTA), programmes (road maintenance networks), interventions (roading treatments) and individual performance agreements may be significantly different. While some measures can be aggregated without change, such as the number of fatal accidents, others will need to be fundamentally changed. There need be no 'golden thread' that links one set of measures to others although the relationship between the measures needs to be understood.

### 2.2.1 Implications for the Road Maintenance Task Force

This line of argument has three significant implications for roading organisations. Firstly, in seeking to answer the question about what is good or best practice in measuring and managing performance, it is necessary to first ask 'who wants to know and for what purpose'? As no one set of performance information is likely to match the requirements of all, there is unlikely to be one 'good' or 'best practice' performance framework. Performance measures focussed on citizens needs may be different from more 'technical' performance measures used by roading experts. A common data set can provide different performance measures for different users and purposes.

Secondly, there may be important, legitimate differences between different stakeholders about what constitutes performance. These considerations suggest the importance of consulting the relevant stake-holders before adopting measures or indicators for your organisation and its programmes. There need not be a 'golden thread' or direct line of sight linking high level outcomes to individual performance

---

<sup>1</sup> For the references on the golden thread, see Gill (2011, p. 16).

measures – different sets of measures may be required at the strategic, tactical and operational levels based on clear understanding on the relationship between the measures.

Thirdly, no one set of performance information is likely to match the requirements of different users and purposes. However, different measures should be grounded in common data sets based on consistent data definitions and the relationships between the measures should be understood. The potential tensions between these purposes, also need to be recognised as for instance, use of detailed programme performance information for external accountability may impede use for the purpose of learning.

### 2.2.2 Recommended further reading

Gill, D., and Schmidt, T. (2011). *Organisational Performance Management: Concepts and Themes* (pp. 9-36) in *The Iron Cage Recreated* (Gill Ed.). Wellington: Institute of Policy Studies.

de Bruijn, H. (2006). *Managing Performance in the Public Sector*, 2nd ed. New York, NY: Routledge.

## 2.3 Citizens' perspective on performance

The previous discussion explored how there are many perspectives on performance and many performances. Some of the literature lends support to the view that the design and development of performance measurement systems needs to include the perspectives of everyone involved, including citizens and customers (Ho, 2008; OECD, 2009). A general message from this is that what managers think is important in terms of performance can often differ from what citizen and service users think is important. Unbeknown to public managers, this can generate what Ho (2008, p. 206) terms an undesirable “performance perception gap”. Engaging citizens in this way is consistent with Moore’s (1995) view of public managers as explorers of public value.

So why are citizens so often excluded? Ho (2008 p197-208) identifies a variety of reasons why government agencies do not seek citizen input into the design of organisational performance measures. These include public officials who often view citizens as incompetent, have negative stereotypes of citizens, perceive seeking the input of citizens as expensive (less money for inputs), share concerns about the difficulties in selecting samples of citizens to consult (which citizens’ views are wanted/unwanted), are aware that citizens may not be motivated or have time to help, and that organisational leadership change from pro-citizen input to anti-citizen input can render the entire exercise a waste of time and money.

One line of explanation for the lack of engagement is that the use by citizens of performance information is one of Pollitt’s (2006) ‘Missing Links’. Gill and Schmidt’s (2011, p 27) survey of the literature on the use of performance information by citizens concludes “the studies we have located generally show a negative picture.” Holzer and Kolby survey United States local governments and conclude “overall few jurisdictions report citizen involvement as a method for adding value or social relevance to performance indicators” (2008 p259).

### 2.3.1 Implications for the Road Maintenance Task Force

Simply reporting performance information on achievement of technical performance measures or consulting citizens and rate payers on Asset Management Plans is unlikely to be effective. The gap between perception and performance suggests that what managers think is important differs from what road users really want. This requires a degree of humility to go out and engage with citizens and service users with no strong prior beliefs, listen actively for what is important and be prepared to act on it. NZTA's experience with focus group research for national roading is instructive – what citizens valued (security and no surprises) was different from what the professionals thought was important (road roughness).

### 2.3.2 Recommended further reading

*Ho, A-Tat-Kei. (2008). "Reporting Public Performance Information: The Promise and Challenges of Citizen Involvement" in Van Dooren, W. and Van de Walle, S. Performance information in the public sector: How it is used. Basingstoke [England]; New York: Palgrave Macmillan.*

## 2.4 Use is the Achilles heel of performance information

The development of formal performance measurement frameworks have often assumed a simple and direct link between the increased availability of supply of performance information and its corresponding use for purposes such as decision-making. However, research does not support the view that supply of performance information results in corresponding demand. The empirical research provides a generally negative view of how parliamentarians, cabinets, councils and citizens use performance information and a mixed view on usage by managers and individual politicians. To understand these research findings it is important to distinguish between the different types of users and the potential purposes.

Gill and Schmidt (Gill 2011, p15) identify two external purposes - accountability and legitimisation - and four main internal uses - learning, monitoring, decision-making and attention focusing. There are three broad groups of external end users: legislatures, executive politicians and the general public.

In summary, there is a marked contrast between non-use by external users, such as citizens and parliamentarians, and more mixed use by managers and politicians with direct executive responsibility such as Ministers. There is no support from studies that legislatures actively use performance information and in some jurisdictions, such as the United States Congress, they have been actively hostile. Similarly, the few studies that are available on direct or indirect use by citizens suggest a lack of interest by citizens.

"While fewer studies of use of performance information by ministers were identified, studies exist on the use of performance information by executive politicians in local government. These studies generally show that local government politicians access various sources of performance information, including quantitative and qualitative information and formal and informal information. The extent to which formal



performance information was used appeared to vary by portfolio, jurisdiction and the characteristics of the person concerned.” (Gill and Schmidt 2011 p35)

Empirical studies on the use of formal performance information by public managers find complex patterns of use and non-use of performance information with considerable variation between ‘power users’ at one extreme and ‘low users’ at the other. Public managers do not systematically ignore performance information nor slavishly follow it. Patterns of use are influenced by a range of factors including person specific factors such as the skills and capabilities of managers, as well as organisational factors such as cultural barriers, functions performed and the different frameworks used by different professional groups.

#### 2.4.1 Implications for the Road Maintenance Task Force

Performance data collection and reporting systems need to be designed for real people not for stylised positions or hypothetical users of reports. Good system design needs to recognise the mixture of skills and professional backgrounds. More sophisticated use of technology (voice recognition, more ease of inputting data and ease of data access) will improve data quality and user engagement. Some local government councillors and cabinet ministers are more likely to be direct users, but the type of information and how it is presented will need to be carefully tailored for the people involved. Parliaments and Councils have a legitimate and important constitutional role of control over the executive but they will not generally use performance information for management improvement purposes.

#### 2.4.2 Recommended further reading

Gill, D., and Schmidt, T. (2011). *Organisational Performance Management: Concepts and Themes* (pp. 9-36). In *The Iron Cage Recreated* (Ed.). Wellington: Institute of Policy Studies.

Van Dooren, W. G., Bouckaert, G., and Halligan, J. (2010). *Performance Management in the Public Sector*. New York: Routledge.

### 2.5 Perversities of performance measurement

One common view is that what gets measured gets managed. Pidd (2008 p72-75) discusses:

- virtualism – such as the pursuit of high performance scores actually displaces efforts to improve outcomes
- performativity – how performance is measured comes to redefine the reality of how performance is perceived.

Different authors have different takes on the perversities and potential negative consequences of the use of performance measurement in performance management.

Hatry (2008 p 237-8) argues that the perverse effects and unintended consequences of performance measurement include:

- performance data does not tell why performance has been good or what should be done to improve services
- performance systems need to be selective but at various times other service attributes are likely to assume importance
- performance data is about the past but decisions are about the future. Projecting into the future is fraught with obstacles and difficulties and does not provide a basis for assessing the expected performance of new options.

de Bruijn (2006)<sup>2</sup> provides the most comprehensive argument in the context of organisations such as hospitals where a professional group or groups dominate service delivery. He suggests that done badly, performance measurement systems can have a number of perverse effects including:

- providing incentives for perverse strategic behaviour (i.e. gaming (p 17)), where efficiency increases on paper but not in reality. The more specific strategies of gaming can include 'cherry picking' (p 19), focusing on 'cash cows' (p 19) and the massaging of reality (p24). See also Radnor (2008 p104) for the distinction between cheating (falsification of data or activities) and gaming (creative classification of activities)
- punishing rather than rewarding performance. The greater the consequences of success or failure (carrot or stick via the provision of incentives, threats of naming and shaming or of managerial intervention), the greater the probability of perverse strategic behaviour (p34-36). And in such a situation, subordinates often feel justified in sabotaging a performance measurement system (pp. 37-38)
- decreasing professionalism by rewarding those who are focused on reaching the targets, even when doing so is organisationally counterproductive (p22)
- the promotion of copying not learning (p24-26). The 'copy' may fail when transplanted elsewhere and discourages sharing techniques of so-called 'best practice' amongst potential competitors, thus diminishing collective knowledge.

de Bruijn (2006) suggests distortions arise because professionals feel justified in perverting the performance measurement system. To offset this, it is important to build trust based on interaction between management and professional staff. The interaction needs to occur around how data is defined, the performance measures that are developed and how standards of performance are assessed. de Bruijn (2006 p54) recommends managers:

- give professionals a say in influencing the definitions of indicators and measures
- include a variety of perspectives when designing and developing measures
- augment 'quantitative' data with qualitative and contextual information
- be honest about how data collected will be used.

Thomas (2006, p63) argues that performance measures need to be "less about sophisticated conceptualisation and precise analysis and more about interaction and seeking a consensus on what should be measured, how and with what consequences. The process should be less top-down and bureaucratic. It should involve consultation with the key stake-holders and the public at large, so that the

---

<sup>2</sup> All page references on this page (except Radnor 2008) are to de Bruijn, H. (2006). *Managing performance in the public sector*, 2<sup>nd</sup> ed. New York, NY: Routledge.

results of the performance measurement system have more legitimacy and support, especially among the people most directly affected by programs. This would also improve the relevance and importance of performance reporting in the eyes of ministers and parliamentarians.”

### 2.5.1 Implications for the Road Maintenance Task Force

Road maintenance has a number of the conditions required for the effective use of performance information. How the performance information framework is developed becomes very important. Most of the problems revolve around the issue of perverse strategic behaviour or gaming. A system imposed on an organisation from above is likely to encourage perverse strategic behaviour because doing so provides professionals with justifications not to support, trust or use (take seriously) such a system. By contrast performance measurement systems that are designed collaboratively by both managers and professionals promote broader ownership, more supporters, greater trust and then utility. Measures need to be developed in partnership with stake-holders rather than imposed from above if they are going to be used and effective. Care must be taken to ensure that measurement is cost effective: some areas (such as road signage) may not be.

### 2.5.2 Recommended further reading

de Bruijn, H. (2006). *Managing Performance in the Public Sector*, 2<sup>nd</sup> ed. New York, NY: Routledge.

## 2.6 Perversities of performance targets

In the United Kingdom in particular there has been an active debate about the use of targets linked to high powered incentives. Within the academic literature there are differences between those who tend to take a positive view of the use of measurement to support goal-directed management and a body of literature critical of the use of targets in particular.

**Table 1 Three implications of performance measurement**

<i>Application of measures</i>	<i>Basic principle</i>	<i>Simple example</i>	<i>Some variants</i>	<i>Comment</i>
<i>Targets</i>	Stipulated floor standard of performance or change in performance to be achieved within some time period	Percentage efficiency savings or staff reductions required over a budgetary period	Specific targets (applying to individuals or particular organizations) versus global or sector-wide targets	Produce threshold and ratchet effects in behaviour of individuals and organizations subject to targets
<i>Rankings</i>	Data allowing comparison of performance on stipulated indicators among a set of rival units	Sporting leagues	Simple comparisons versus composite leagues (with numbers distilled from other numbers)	Produce output distortions and pressures to change the composition of the league and the nature of the game
<i>Intelligence</i>	Background information	Activity logs, for example of health care ‘episodes’	Anonymized data (for example for near-miss reporting) versus attributed performance data	Use is unpredictable by those whose performance is recorded; often combined with targets and rankings

Source: Hood 2007

Hood (2007) distinguishes between three different styles of performance measurement – ‘intelligence systems’ where performance measures are used as background information, ‘ranking systems’ such as league tables where comparative performance is tracked over time and ‘target based systems’ where performance is specified against an aspirational standard. Table 1 (above) compares these systems.

Each of these different styles has different strengths and weaknesses. Targets, for example suffer from ratchet and threshold effects (incentives to reduce performance in the current year to enable future targets to be met), output distortion (‘hitting the target and missing the mark’) along with cheating and gaming on measurement shown in Table 2 below.

**Table 2 Conformance gaming and cheating**

	Internal	External
<b>Conformance</b>	Recording what is required not what is needed	Hitting the target missing the mark - regravelling roads close to gravel pit
<b>Gaming- measurement only</b>	Flexible Coding	Fiddling response time measures
<b>Cheating- measures &amp; activities</b>	Reclassifying support staff as front line staff	Changing activities to meet the target e.g. charging low value gravel as premium quality gravel

Source: NZIER partly based on Radnor (2009)

### 2.6.1 Implications for the Road Maintenance Task Force

There is no silver bullet as all measurement is subject to error of various kinds and measures are a proxy for a more complex reality. Different regimes have different strengths and weaknesses. This is not an argument against the use of measures but a caution about acknowledging the limits as well as the role of measurement. In the case of roading maintenance there is a strong case to build an improved ‘intelligence system’ but moving beyond that to ranking or target systems would require careful consideration of the costs and risks relative to the potential benefits. Care is required before introducing high powered incentives – personal liability can result in less measurement, performance pay directly linked to performance measures can result in cheating and gaming.

## 2.7 Qualities of good performance measures

This section will discuss the different qualities or features of good performance measures. Roberts (2006, p15) (quoted in Thomas 2006) presents “The ‘Ideal’ of Performance Measurement—A Destination Never Reached:

- it uses measures which are valid, reliable, consistent, comparable and controllable
- it produces information which is relevant, meaningful, balanced and valued by the leaders/funders of the organisation.”

Looking at standards that would apply at the programme or organisational level, the International Public Sector Accounting Standards Board is consulting on qualitative characteristics for reporting by public entities. These include:

- relevance - service performance information should have a close logical relationship between the information provided and the purpose for which it is intended to be used
- timeliness - service performance information should be reported to users before it loses its capacity to be useful for accountability and decision-making purposes
- verifiability - service performance information should provide users with a basis for assessing whether the information in a service performance report could be replicated
- faithful representation - service performance information should provide a realistic representation of the service performance of a public sector entity's services
- understandability - service performance information should be communicated to users simply and clearly
- comparability - service performance information should provide users with a basis and context for assessing a public sector entity's service performance.

This list is very similar to the United Kingdom's criteria for good indicators – relevance, attribution, timeliness, reliability, and verifiability and the FABRIC characteristics of a good performance measurement system: Focused, Appropriate, Balanced, Robust, Integrated and Cost effective.

In the New Zealand context the NZICA (then ICANZ) Technical Practice Aid No. 9 (September 2002) and the recently issued Treasury and SSC guidance on reporting would provide a useful point of comparison. For example, the principles underpinning TPA9 include external focus, controllable, comprehensive, measurable and informative to the user.

### 2.7.1 Implications for the Road Maintenance Task Force

The striking thing about the discussion is the number of competing lists, their lack of convergence (despite varying areas of overlap) and the lack of useful guidance about how to resolve inevitable conflicts between the criteria. There is more agreement about the major problems inherent in performance measurement systems in the public sector due partly to differences in the level and/or the intended user of the performance information. Good practice design of performance measures will be shaped by consideration of 'who wants to know' and 'for what purpose' the information is to be used.

Great care is required when measures developed for one purpose (such as providing trend data to assess road condition life cycle) are being used for another purpose (such as service standard levels). The more incentives to meet the measure, the more likely to generate conformance, cheating and gaming.

### 2.7.2 Recommended further reading

Thomas, P. (2006). *Performance Measurement, Reporting, Obstacles and Accountability*. ANU E Press: Canberra.

## 2.8 The qualities of good performance measurement systems

While the previous section discussed the desirability qualities of individual performance measures, this section will explore the qualities required in a good overall system. Thomas (2006, p65) presents: “Conditions Favouring Performance Measurement:

- agreement on what constitutes performance
- activities involved are amenable to measurement on a quantitative or qualitative basis
- cause-effect relations are reasonably well understood and attribution is possible
- scale of operation is large enough to spread the costs of designing and operating the measurement system
- leadership support for the activity and the culture of the organisation supports dialogue over what the evidence is saying.”

Where the approach to performance measurement fits with the mission and nature of the activities he suggests the following steps:

- consult the relevant stake-holders before adopting measures or indicators for your organisation and its programmes
- link measurement activities to strategic/business plans
- set forth as clearly as possible performance expectations and compare to actual results
- strive for balance in your performance measurement system between: comprehensiveness vs. relevance/simplicity; financial vs. non-financial, short-term vs. longer-term; control vs. learning, outputs and outcomes; quality from an internal, professional perspective with quality from an external, user perspective
- promote a culture of performance management within your organisations. Create incentives or remove disincentives for the use of performance measure
- encourage the development of causal models of programmes which link outputs to desired outcomes
- ensure fairness in the use of performance data to appraise the performance of organisations and individuals. Allow for the recognition of factors beyond their control
- approach the task of communicating about performance in a strategic fashion by paying attention to the needs of different audiences
- take a pragmatic approach: use pilot projects in areas more amenable to measurement, make use of existing data sources, acknowledge the limits of existing data, but do not wait for the ‘best’ data to become available and review the cost-effectiveness of your system periodically
- consider benchmarking your performance to that of superior comparable organisations and share knowledge with other organisations
- recognise the limits of measures. Don’t be mesmerised by the numbers. Ensure the continued relevance of your measures. Avoid doing the wrong things well, based on your performance measurement system” (Thomas, 2006, p. 66).

Roberts (2006, p15) presents “The ‘Ideal’ of Performance Measurement—A Destination Never Reached:

- it has clearly defined purposes and uses
- it focuses on outcomes, not just on inputs and outputs
- it employs a limited, cost effective set of measures
- it uses measures which are valid, reliable, consistent, comparable and controllable
- it produces information which is relevant, meaningful, balanced and valued by the leaders/funders of the organisation
- it is integrated with the planning and budgetary processes
- it is embedded in the organisation, is stable and is widely understood and supported.”

Quoted in Thomas (2006, p. 65)

Thomas (2006) identifies four barriers to the success of performance measurement – institutional, financial, technical and political:

- institutional: Organisational resistance in times of economic hardship can result when public servants associate performance measures with budgetary cuts and with intentions to slash ineffective programmes (p 3, 6, 36, 48). This situation can stimulate gamesmanship (p52). Other institutional issues are that ‘softer’ organisations are less amenable to measurement and linking programmes to outcomes (p21) and mission statements for organisations tend to be vague making it difficult to agree on operational measures (p48)
- financial: Performance measurement can be very expensive to design, implement and run (p 3, 36), and divert funds from actual programmes (p 47)
- technical: Establishing causal attributions in relation to outcomes is very difficult (p47) although not impossible, quantity is easier to measure than quality (p48) which can lead to information overload (p52), the measurable can be prioritised over what is truly important (p56), which can result in the collection of useless information (p48)
- political: Adversarial, negative, and theatrical political processes and the media only focus on failures which promotes risk averse behaviour in public managers/ministers (p 9, 12, 56, 57) and gamesmanship (p 52). “In short, the current culture of parliamentary government clashes with the ideal of a performance measurement system, in which a balanced, constructive and learning approach is assumed” (p. 57).

### 2.8.1 Implications for the Road Maintenance Task force

Like the qualities of good individual performance measures, the features of good systems provide the counsel of perfection. The competing lists provide little guidance about how to resolve inevitable conflicts between ideal features. Nonetheless they can be used to review the overall system. In the remaining parts of this paper we utilise a cut down version of Thomas’ criteria to review the current performance measurement regime applying to road maintenance.

Developing a good set of measures requires leadership as robust performance measurement systems evolve over time. Generally leadership and persistence is required because progress is gradual and incremental with no single spectacular

breakthrough. The leaders of organisations with responsibility for roading must model the ability to champion and lead change. This will require the ability to handle ambiguity, support dialogue on what the performance information does and doesn't support, and encourage the organisation to tell a broader performance story based on quantitative and qualitative evidence.

## 2.9 The limits of performance measurement

Some sort of performance measurement is inevitable in assessing overall performance. "All humans make judgements and these judgements rest on measurement of some kind" (Pidd, 2008, p69) "performance measurement in some form or other is probably present in all organisations and it is pointless to argue about whether this is desirable" (op. cit. p71).

Not only is organisational performance measurement inevitable, there is also evidence that it can be beneficial (see Bevan and Hood 2006). For performance measurement to be beneficial it also needs to be recognised as being dysfunctional if not done properly.

At the system design level, when modifying the architecture of the performance measurement system it is important to take into account de Bruijn's law of system dynamics. The dynamics established by performance measurement means that four perverse laws operate:

- the law of decreasing effectiveness (the system is perverted)
- the law of mushrooming (the system is bloated)
- the law of collective blindness (myopia based on short-sightedly putting too much weight on production figures)
- the law of preserving perverted systems (insufficient incentives for abandoning the system) (2006, p. 33).

To illustrate one of these in more detail "the Law of Decreasing Effectiveness leads to a paradox: the greater a manager's efforts to manage on performance measurement, the stronger the incentive for professionals to exhibit perverse behaviour. More control leads to more negative effects" (p37). And those who engage in gaming can be rewarded only encouraging perverse strategic behaviour.

The greater the consequences of success or failure, the more performance measurement provide incentives for perverse strategic behaviour or gaming and cheating). The greater the threats of naming and shaming or of managerial intervention, the greater the probability of perverse strategic behaviour. In such a situation, subordinates often feel justified in sabotaging a performance measurement system.

### 2.9.1 Implications for the Road Maintenance Task Force

The law of diminishing effectiveness is not an argument against measurement. Indeed low impact measurement acts to spur improvement. It is an argument of how diminishing and eventually negative returns set in when measurement is pushed too far and too hard. The trick is to find the sweet spot. One option would be to introduce



a 'tit for tat' rule so that no additional costs of data collection could be imposed on local RCAs unless offsetting reductions in the cost of data collection are identified.

### **2.9.2 Recommended further reading**

de Bruijn, H. (2006). *Managing Performance in the Public Sector*, 2<sup>nd</sup> ed. New York, NY: Routledge.

## 3. Analysis of the performance measurement regime for road maintenance

The previous section summarised the key themes from the literature on performance measurement in the public sectors of OECD countries and identified a number of common threads:

- usage is the Achilles heel – supply of performance information does not create demand and systems need to be designed with care to meet the needs of real people not hypothetical users
- different perspectives – there are many different perspectives on performance and different users of performance measures require different performance information for different purposes. There need be no ‘golden thread’ that links one set of measures to others
- citizens’ perspective – what managers and experts think is important differs from what citizens value
- there is no silver bullet – measures are a proxy, are subject to measurement error, support multiple interpretations, and encourage perverse behaviours
- performance information is subject to limits due to problems of conformance (what gets measured gets managed), gaming of measures and cheating of both measures and activities.

While performance measurement can play a positive role, it is also subject to distinct limits and distorting actions. This section looks at the performance measurement regime in road maintenance. It begins by appraising the formal system in terms of the data, measures and purposes of performance information. The data and measures required by the formal system are then compared with those actually used by practitioners at different levels of the system (NZTA, RCA and contractors). We conclude with a discussion drawing out the underlying issues and problems.

### 3.1 Our approach

We have worked closely with a handful of respondents in the NZTA and RCAs to understand the current formal performance measurement regime and the real regime that practitioners actually use. To do this we conducted:

- a desk-based review of the performance measurement regime in place in the New Zealand road maintenance sector
- one-on-one, semi-structured interviews with practitioners in NZTA, RCAs and other private sector organisations (listed in Appendix B)
- a facilitated half day workshop with twelve participants (Appendix B has the list of participants drawn from NZTA, RCAs and private industry)
- a dialogue with the Technical Working Group based on the workshop findings.

We expected that the desk-based review would provide a clear exposition of the formal system. In fact, a document that puts it all together in one place proved elusive. In part this is because of the number of dimensions involved: the different

organisational levels (project, programme, network, system-wide and organisational), the different phases (design, collection and storage, analysis, reporting), and different users and uses.

In the workshop we asked participants to map out the formal system as they experienced it. This confirmed what the literature and desk-based reviews had identified: that there are many different performances and diverging perspectives and that no single system describes the detail of the road maintenance regime enacted across all of the RCAs.

Indeed, the participants in our workshop emphasised that the real system they employ in many cases fills in 'gaps' that are left by the formal system, rather than replacing it. To reflect that, discussion starts with the flow of performance information throughout the formal planning, budgeting and programming system and then highlights places in which the real system might diverge slightly, or where the real system augments the processes required by the formal system.

Throughout our work we used the distinction between data (know what), information (know how) and knowledge (know why). Data is the raw input that is gathered and stored and is subject to only limited analysis (quality testing and cleaning). Data can be qualitative (text) or quantitative. Quantitative data can be collected in many ways, including electronically (high speed data scanners) and by human visual observations of road roughness. Measures are a form of information that organisations use for a variety of purposes. Distinguishing between the data collected and information reported, and the knowledge generated at each stage of the process is crucial to understanding the problems with the present system.

## **3.2 The formal system of performance measurement**

Defining the system of performance measurement for road maintenance requires asking who wants to know what, and for what purpose? Once that is understood then the flows of data between parties, and the measures used for different purposes can in principle be mapped out and the bottlenecks, gaps and deficiencies identified.

### **3.2.1 Who are the main organisations?**

The 'who' of road maintenance performance measurement are the organisations at each of the three levels of the system: the NZTA's Planning and Investment division, the RCAs, and private road constructors and consultants.

The road maintenance system occurs within a legal framework:

- NZTA is a Crown Agent bound by the accountability and governance requirements of the Crown Entities Act (2004) augmented by additional provisions in the Land Transport Management Act (2002)
- RCAs are part of local territorial authorities operating under the accountability and governance provisions of the Local Government Act (2001)
- private industry constructors and consultants are private legal entities governed by the relevant law for the legal form, generally the Companies Act (1993).

In addition to primary legislation, there is secondary regulation and tertiary rules and guidance that place a series of rules and requirements on the entities in the roading maintenance sector. For example, NZTA is included within central government procurement rules, the Capital Asset Management regime and the associated National Infrastructure Plan. Similarly, RCAs are bound by local government procurement rules, planning and reporting requirements (including the LTCCP) and may face additional reporting requirements once the new service reporting requirements being developed by DIA are promulgated.

The NZTA is a regulator, a provider (through the Highways and Network Operations Division) and an investor in road maintenance. NZTA as regulator has created rules specifying the protocols for digital collection of data of road roughness and the approved algorithms for analysis of the preferred treatment. NZTA also provides guidance on standards.

In this report, we will refer to the part of the organisation that controls investment and programming as 'NZTA'. The NZTA division that controls the state highway network is also an RCA. We will use the term 'local RCAs' for the local authority controlled roading organisations to distinguish them from NZTA Highways and Network Operations Division. Like NZTA, the local authorities also have a dual role as an RCA and an investor in roads through their co-funding of maintenance.

Private industry falls into two broad categories of contracts with RCAs - contracts for professional services (consultants used to oversee work) and contracts for works (construction and maintenance contractors). Consultants and contractors can carry out tasks ranging from individual maintenance tasks all the way through to overseeing the collection and analysis of data and construction of the maintenance programme. Many of the tasks that we describe as being the domain of the RCA are actually conducted by the contractor. However, the split of tasks between the RCA's staff and the consultant/contractor is a contractual matter, rather than a performance measurement matter. There are a range of contract types and these are well described and analysed in detail in the report commissioned by the Task Force from OPUS (2012) so will only be touched on here.

### **3.2.2 What data and information is generated and used?**

The 'what' of performance information can be organised into four categories that align with the phases of the road maintenance programming and investing:

- the data is gathered and the information used for asset management planning
- the data is gathered and the information used to programme road maintenance
- the information used to determine the investments made in road maintenance
- finally, information gathered to audit and monitor the maintenance programmes.

#### ***a) Asset management planning***

Asset Management Plans (AMP) are pivotal to understanding the formal system as they are designed to guide future decisions about programming, investing, and auditing for NZTA Highways and local RCAs. The AMPs and related NZTA documents provide the framework and context within which data is gathered,

measures generated and new information created. AMPs describe the state of the assets, how the RCA intends to manage the assets, and what levels of service the RCA intends to provide. The AMPs define:

- what measures of asset condition will be monitored
- what levels of service must be provided
- what standards have to be met
- how asset maintenance, renewal and replacement will be undertaken (including procurement)
- the estimated future costs associated with providing any identified extra capacity needs and replacement and maintenance of existing assets.

While NZTA provides guidelines for road maintenance requirements, they are not binding on RCAs, so long as procurement and asset management processes are adequate. As a result the details of AMPs are specific to each RCA. RCAs are required by the NZTA to assess road condition, but the method of condition assessment and the standard that must be met can be defined by the RCA without reference to the NZTA's guidelines.

### ***b) Road maintenance programmes***

AMPs are the first phase of road maintenance programming and investing. Having conducted their analysis and determined their maintenance requirements, the local RCAs provide the NZTA with the proposed road maintenance programme, via the regional land transport committee.<sup>3</sup> Having collected the road maintenance programmes developed by the RCAs, the NZTA must then decide how to disburse funds to meet the needs identified by those programmes. The funding for maintenance investment is primarily drawn from the National Land Transport Fund (NLTF), which is administered by the NZTA.<sup>4</sup>

### ***c) Crown investment in road maintenance***

The next phase commences with the negotiation between the RCA and NZTA over the proposed road maintenance programme. If an activity is approved for funding then it is partially funded by the NZTA, while the local authority funds the remainder, which amount to approximately half the total spending.<sup>5</sup>

While the NZTA has discretion over which activities to fund, the management of the RCA's roads is determined by the criteria set by the local RCA itself: the NZTA provides guidelines for the management of roads but the RCAs are not required to adhere to them. As a consequence, the NZTA is required to gauge the cost effectiveness of the local RCA's road maintenance programme against the RCA's own criteria, developed in their AMP.

---

<sup>3</sup> The regional committees have a lesser role in prioritising maintenance spending, so their role is not emphasised here.

<sup>4</sup> Local RCAs road asset managers are also responsible to their councils as co-investors, who may provide up to half the funding for road maintenance.

<sup>5</sup> The funding rate depends upon the type of maintenance so the average depends upon the mix of maintenance types conducted. The present base funding rate by the Crown is 48%.

### d) Auditing, monitoring and reporting

Once work has been completed it is reported upon, and potentially subject to audit.

RCAs require regular reporting from contractors on work done. Contractors are usually also required to enter the work completed into a database, such as Road Assessment and Maintenance Management (RAMM)—the most commonly used system. Finally, RCAs tend to also have either a network engineer or clerk of works inspect work once it is completed in order to confirm that it meets the required standards.

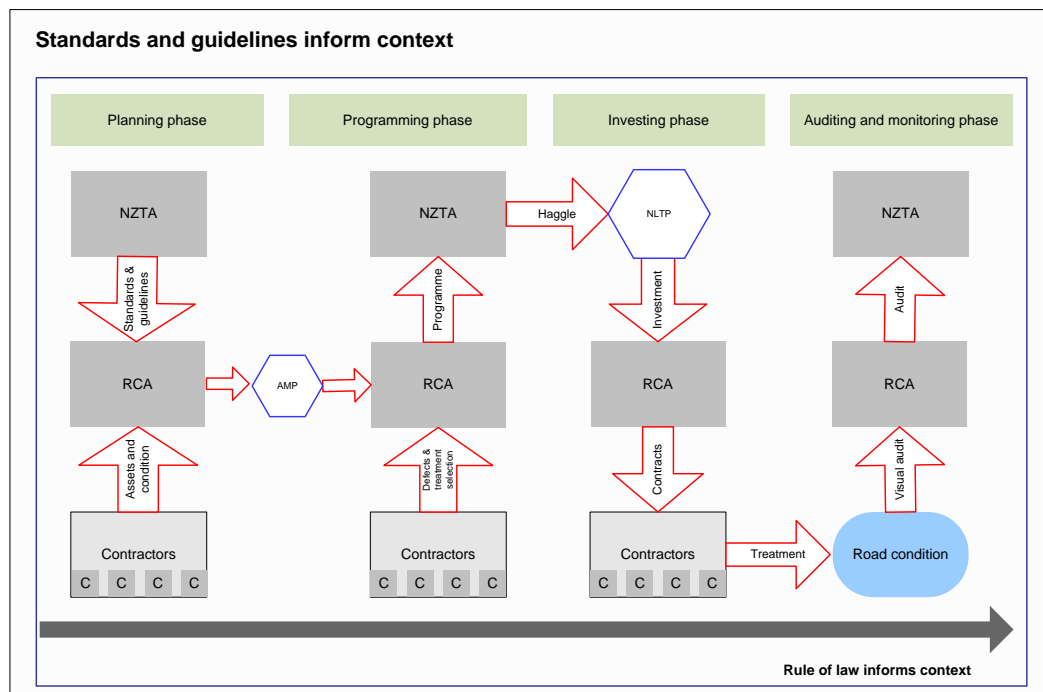
At a level above the RCAs the NZTA monitors and audits the work that it has funded to confirm that it is done, and that has been undertaken to an appropriate standard.

In addition to having standards for the AMP, the NZTA also requires that local RCAs demonstrate that they are procuring services in an efficient fashion. This is done through a review of the RCA's procurement processes, which are required to fulfil certain criteria in order for the RCA to gain funding. For each local RCA, the NZTA decides whether they have fulfilled the required criteria for procurement processes before they allow funding to be drawn down.

### e) Summary

The process of planning, programming, investing and auditing is schematically outlined in Figure 3.

**Figure 3 Programming and budgeting for road maintenance**



Source: NZIER, NZTA

The lack of a loop back in Figure 3 from the Auditing and Monitoring phase into the Programming phase is deliberate as there is no cycle of performance measurement. Rather there is an open loop with limited direct and indirect feedback from the

monitoring and measurement subsystem to the design of programmes and policies. This lack of a closed loop is impeding performance improvements being identified and implemented.

In the next section we describe for RCAs and NZTA what data is required to be collected and what performance measures are constructed before reviewing what data and measures are actually used, and the knowledge generated.

### 3.2.3 What data is collected

#### *a) Data collected by RCAs*

The RCAs collect both data for measures that are required to be provided to NZTA as well as the performance data that is required for their own AMP planning measures. While not all RCAs collect all of this data there are a number of common sources.

#### **Assets are quantified**

The first class of data that is collected is a register of all assets on the network. That quantification of the RCAs' current asset position is entered in to an asset management database, such as the popular RAMM software.

#### **Condition data is collected**

The next stage of data gathering is to assess the condition of the network. Ordinarily, an RCA will employ a consultant to assess the road network.

The precise method of data collection is, for the most part, not defined by the NZTA. Some RCAs may digitally assess road condition using high-speed data collection for example, while others may simply visually inspect the road to determine whether rutting is acceptable.<sup>6</sup>

These data will be entered into a database system, such as RAMM. Defects are entered annually after a survey of the network and, over time, the database will accumulate the condition history of the present network.<sup>7</sup>

#### **Visual inspection to validate analysis**

RCAs have engineers visually survey the roads on their network to validate or modify the treatments selected during analysis of measures.

#### **Traffic counts**

To determine a maintenance programme many RCAs conduct demand assessments, which involve traffic counting. Often, this work will also be conducted by a consultant.

---

<sup>6</sup> HSD collection requires calibration of testing equipment to NZTA standards. Visual road condition assessment requires evidence that the assessor has attended a course on condition rating.

<sup>7</sup> RAMM does not record the history of roads that no longer exist, although that may not be the case for all such databases.

## Ratepayer surveys and call monitoring

Local RCAs have a dual responsibility to NZTA and their ratepayers and tend to survey them regularly. Part of that survey usually involves questions related to the transport network, and that commentary will be passed onto the road manager. The data from surveys and focus groups can be built in to measures that help to understand ratepayers' concerns. There is no standard way of doing this.

RCAs also monitor the calls that are made to the council's call centre. Any calls regarding road maintenance and road performance will be logged as such.

### *b) Data collected by NZTA*

#### Technical data

Having collected the information required by their AMP in order to construct a proposed road maintenance programme, the RCAs are then required to pass a defined subset of that information on to the NZTA.

The RCA provides summary information on the road's condition to the NZTA in its annual achievement report. The required information is against indicators that summarise the road condition of the network.

RCAs also provide the financial information from the past year's investment programme via the TIO database on an annual basis.

Finally, NZTA audits the maintenance done by RCAs. Data is collected on an ad hoc basis, as deemed appropriate by the NZTA auditor.

#### Procedural data

NZTA collects information from RCAs on their procurement and asset management processes. For procurement, in particular, NZTA requires the following data to be provided for each contract let by the RCA:

- the class of activity
- the delivery model selected
- the procurement method chosen
- the preferred supplier.

### 3.2.4 What measures are used?

Measures are constructed to interpret and give meaning to the raw data. While much of the data collection is intended to inform certain measures, not all of the above data is used to construct measures. Some may, for example, be used solely to inform modelling analysis or as a record for future research.

#### *a) Measures constructed by RCAs*

##### Road condition measures

The data gathered by the consultant supports three categories of road condition measures:



- NZTA-required measures such as roughness, surface condition index, and pavement condition index. They will also collect the skid resistance, rutting, and texture for national highways. There are standard definitions across RCAs for these measures
- RCAs also tend to request standard measures such as customer response and travel times, for which there are definitions
- RCAs may additionally request any number of non-standard measures that they have included in their AMP for assessing road condition.

The formation of measures from the data has no universal standard. For instance, visual and digital data may variously be used to report the road condition indicator. Consequently, the measures used are not easily comparable across RCAs.

### Demand forecasts

Demand forecasts for each part of the road network are done as required by the AMP and the level of detail is not prescribed by the NZTA. The measures reported from this data are not standardised but will draw upon traffic counting data.

### Road performance measures

Drawing upon the asset database and the history of faults, RCAs report measures of trends in faults and identify repeat fault locations.

### Customer satisfaction measures

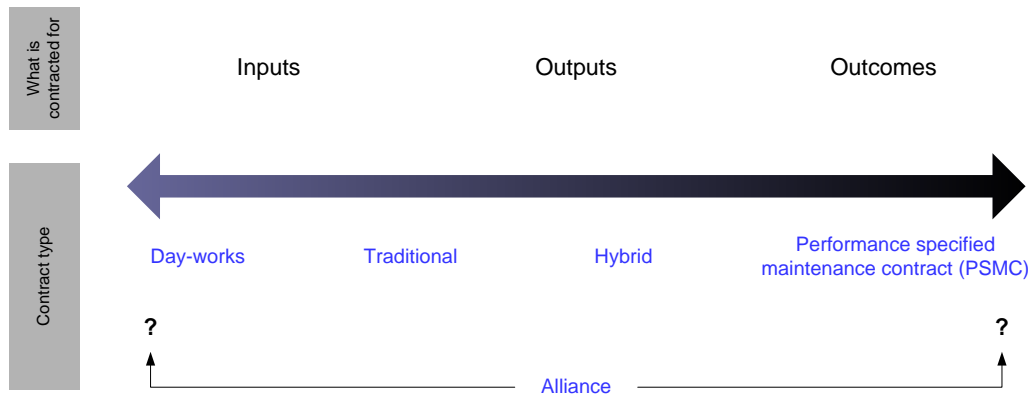
Whether this information is an important part of the programming process seems to vary markedly between authorities. For some, their AMP's levels of service make specific reference to stake-holder satisfaction measures. For others, the data is used casually and not formalised into performance reporting.

### Measures built into contracts for works

In addition to road performance, the RCA must also measure the performance of the contractor who carries out the maintenance. The particular measurement regimen depends upon the contract chosen. OPUS describes the range as a continuum from inputs, outputs, outcomes and alliance contracts.

There are two important points from the OPUS work shown in figure 4. The first is that different performance measures and data requirements are associated with each type of contract. NZTA (highways) has contracts across this entire continuum (apart from input contracts). Opus describe the contracting continuum as consisting of Input Contracts for day work, contracts for defined outputs as 'Traditional Contracts', 'Hybrid Contracts' (as a mixture between output and outcome forms), Outcome Contracts (such as performance specified maintenance contracts or performance based contracts) and alliance contracts (a complex risks sharing arrangement involving an amalgam of input contracts informed by shared outcomes). The contract type depends upon factors such as procurement market conditions, network type and complexity, traffic density etc. Similarly there is also a degree of variation in the manner in which contractors are used by local RCAs in response to the variety of conditions faced and the freedom in determining their maintenance arrangements.

**Figure 4 Continuum of contracting**



Source: OPUS 2012 p20

The second important point is that different contracts have different measures and as ‘what gets measured get managed’ this has consequences for performance. The earlier discussion raised the problem of conformance (see Table 2) involves hitting the target and missing the mark. The quality of specification, the relationship management and the quality of monitoring all affect the extent of conformance. Under output-based contracts, constructors have incentives to deliver on the specified measure even if this results in over servicing by supplying work that isn’t actually required. Under outcomes contracts, the incentive is to do least work required to achieve the outcome, with the accompanying risk of underservicing. All types of contracts face the problem of conformance described in the literature survey in section 2 whereby ‘hitting the measure’ means ‘missing the mark’ as contractors maximise the return within the rules while minimising the amount of effort required.

### ***b) Measures used by NZTA***

#### **Road condition measures**

The NZTA is provided with summaries of the road condition on each network by the RCAs.

#### **Cost benchmarks**

The NZTA compiles information on financial performance and outcomes gathered from the Transport Investment Online (TIO) database. From that it can report against measures such as the cost per kilometre and cost per vehicle kilometre travelled across RCAs.

#### **Compliance measures**

Technical audits of RCAs are conducted that review

- the selection of faults to remedy
- the selection of treatments for those faults
- implementation of the programme of treatments.

The NZTA employs its own road engineers as auditors, who physically inspect work conducted by the RCAs and their contractors to check that it is of the required standard. The NZTA's auditors also assess whether their objectives of strategic fit and cost effectiveness are being fulfilled by the work programme enacted using NLTF investment.

### 3.2.5 What information is used?

#### *a) RCAs' use of measures*

The programming of work relies on analysis of the measures collected. Given the differences between RCAs' AMPs there is no one set of analyses that are employed, but the common uses are surveyed below. In almost every case, these analyses will be conducted by a consultant engaged by the RCAs, rather than in-house staff at the RCA.

#### **Treatment selection**

This is the only mandatory piece of analysis required by NZTA of every RCA. The road condition measures are required to be analysed using an approved treatment selection algorithm (TSA) to determine the initial treatment selection for the network. Running the TSA is a requirement for gaining investment funds from NZTA.

#### **Long term condition forecasting**

Over a longer time horizon than the three year funding focus of the TSA, Deighton Total Infrastructure Management System (dTIMS) software provides an estimate of road deterioration over the following 20 years. It uses data from RAMM, traffic counting data, and climate information to project which roads are likely to need treatment and when.

That information is less commonly used by RCAs but can be useful in helping them to forecast their future financial liabilities. Its use has been championed by the Road Information Management Steering (RIMS) group over recent years.

#### **Lifecycle analysis**

RCAs use the age of their roads (from the asset database), along with performance measures to conduct lifecycle analysis and determine areas where road performance indicates that the pavement will need rehabilitation.

#### *b) NZTA's use of measures*

#### **Investment**

NZTA uses the measures it collects to determine the appropriate activities to include in the NLTP and to disburse funds.

#### **Comparing RCAs' performance**

They benchmark RCAs' cost-effectiveness against similar authorities throughout the country and confirm that they are not under or over-spending for the level of service provided by their road network.

Part of the comparison is an audit of the RCA's processes and network by an NZTA auditor. The auditor compiles a report on an RCA's compliance which can then be used to compare the RCAs' compliance performance.

### 3.3 The real system in practice

Here we ask how the real system of performance measurement differs from the formal system. Rather than exhaustively detail the system in place we focus on the key points of difference. Most of this information is drawn from interviews with the participants listed in Appendix B without attribution.

#### 3.3.1 What data is collected in practice?

##### *a) There is a broad range of data collected.....*

As described above, there is a wide array of data types that can be collected by RCAs and used to inform their programme. For the most part, RCAs are practised at using data to inform their programming and investing decisions and that is borne out in the range of data that is collected.

##### *b) ...but collection is patchy*

Our interviews suggest that the data that is collected varies greatly by RCA, and many collect only a small subset of the possible data that could be used to inform their programming decisions. For example, the RAMM software allows for a wide array of data to be entered to inform further analysis, but many RCAs fill in only a few of those fields. That is, of course, largely because their measures do not require the remainder of the possible data.

In addition, the method of data collection varies: some RCAs will use quantitative methods to survey their network while others will largely rely on visual judgements.

There are also classes of assets for which there are no formal performance data requirements, and for which RCAs must develop their own data collection and analysis standards. For example, bridges and road signs are not covered by the formal asset management guidelines of the NZTA. Nonetheless, RCAs must measure these and so collect data on them in some fashion.

##### *c) Data available to NZTA is limited*

NZTA has limited technical information available. Road condition indicator measures are passed up from the RCAs, but they describe only the average condition across a network, rather than the distribution of condition. In order to support more detailed performance reporting, the NZTA draws on a significant amount of data that it is not required to collect, but has incidental access to.

The main example of this is RAMM, which NZTA can access and which contains the raw data from many of the RCAs. The NZTA uses the RAMM data that it has access to, but has no control over many aspects of the data collection. It also can do little to deal with missing data points, since RCAs are not required to collect much of the data that RAMM could hold. In the case of RCAs who use other database software, the NZTA can access only the condition data provided by the RCAs.

### 3.3.2 What measures are used in practice?

#### *a) RCA measurement is highly variable*

The analysis of data to report against measures is highly variable across the many RCAs. For instance, we were told that dTIMS is a valuable tool for forecasting road deterioration on the network, but it is largely distrusted by road asset managers. The reasons given were that the data quality fed in to dTIMS is often poor—for reasons described above—and that asset management processes may be too poorly developed to take account of the measures produced.

#### *b) NZTA is limited by the availability of data*

The NZTA has the economies of scale to employ specialist data analysts who have the expertise to develop useful new measures. However, the analysts are limited by the availability of data, which they do not directly control the collection of. Using the data it has access to, the NZTA has attempted to apply numerous measures of road performance and assess the outcomes of road maintenance.

For example, it has developed indicators of network efficiency, asset condition, travel performance, customer satisfaction, and environmental performance. All of these are limited to some extent by the availability of consistent, reliable data. It is also attempting to report against measures for the outcomes and efficiency of road maintenance programmes but, again, these measures are hampered by patchy and inconsistent data.

NZTA has also attempted to report trends over time, rather than looking at the static measure solely. These efforts have run into similar difficulties.

### 3.3.3 What are the real uses of the information?

#### *a) RCAs manage maintenance budgets rather than assets*

A notable aspect of the job of a road asset manager is that they are expected to spend their budget: not less than their budget, but the budget precisely. The cost of not spending the budget is that it is likely to be cut the next year. As a consequence, road asset managers carefully monitor their annual spending in order to ensure that their entire budget is spent. Indeed, there was evidence shown to us that some RCAs do a large amount of resurfacing in winter as the end of the budgetary year approaches, despite that being a poor time to resurface.

#### *b) NZTA uses the measures to set implicit standards*

The current practice is that NZTA provides guidelines on roading maintenance standards for levels of service. These standards are provided as guidance and RCAs have discretion about whether and how much is adopted. As there is no requirement to adopt and no direct consequences from failing to do so, take up of these standards has been patchy. In practice the use of these standards varies and some RCAs adopt the standards more completely than others. NZTA as investor has an implicit role as setter of maximum network standards by declining to invest in 'levels of service' at the project level above the national guidelines

The NZTA uses the measures for which it has good information to help prioritise and choose activities for programming and investment. The NZTA is required to decide which elements of which programmes will gain funding, based on cost effectiveness criteria. Those criteria require a standard to be applied to the measures of the levels of service provided, which is formally intended to be the levels defined in the RCA's AMP. However, the NZTA may consider that either the RCA's measures are inadequate, or that their levels of service are too high. In either case, the NZTA may decline to fund some activities.

Here, the NZTA is implicitly applying the standards that it has developed for the measures that it calculates, and overriding the RCA's own measures and standards.

### **3.4 Why do the real and formal systems differ?**

The literature review in the previous section of this report had a number of important messages for interpreting the use and non-use of performance measurement. One theme that has proven particularly salient for road maintenance is that the uses differ across units within an organisation. Since different uses call for different measures, there is no reason to believe that the measures used by RCAs will be relevant to NZTA, and vice versa.

Another issue is that the uses of information must be supported by appropriate measures, and the measures must be supported by appropriate data collection. The lesson we draw is that an analysis of the system must seek to understand the uses of information, before asking whether the measures, and then the data collection, are adequate.

#### **3.4.1 Strengths of the current system**

The current system of performance measurement practiced in the road maintenance sector has a number of positive characteristics. Firstly, there is a high degree of agreement on what constitutes good performance in road maintenance and that activities such as pavement condition are amenable to measurement. There is a coherent and well-embedded culture of data collection and use of measurement accepted across the sector. At each of the organisations involved, there is strong acceptance of the need for measurement and use of the measures generated.

Secondly, the work with dTIMMs has shown that cause-effect relationships are reasonably well understood and attribution is possible. Thirdly there is a degree of integration of performance measures into the planning, budgeting, programming and contracting systems shown in Figure 3.

Lastly, there is a history of collaboration between organisations on the subject of performance measurement. Organisations share data fairly freely and collaborate to develop new methods of measurement (e.g. RIMS and NAMS).

These strengths should not be underestimated and provide a strong base to build from.

### 3.4.2 Problems with the system

The common thread that emerged from our analysis of the performance measurement system used in road maintenance is that a vast quantity of data is collected but it is not used to the extent that it could be to inform decision-making and asset management – and that the vastness of what is collected overshadows what is not collected.

#### *a) At NZTA*

NZTA wish to use measures primarily to allocate investment, but also to aid in learning. They have ideas on the measures that they wish to use for each, but inadequate data to implement them.

In particular, NZTA would like more consistent data, with standardised collection techniques, in order to facilitate both research and investment allocation across the range of RCAs.

#### *b) At RCAs*

RCAs need to:

- manage their road maintenance programme effectively, in accordance with good asset management practice. A part of that is the need to learn from the outcomes of previous decisions
- justify their chosen maintenance programme to the NZTA and their council in order to gain investment funding
- monitor and manage their consultants and contractors.

The measures that local RCAs collect are not always adequate to support these objectives. For example, NZTA Highways have a range of measures that local RCAs don't have.

#### **NZTA needs better data**

Our interviews suggest that NZTA have some idea of what information they need, but are unsure about the precise measures that they would use to gain it. What they are sure of is that the present data passed to them by RCAs is insufficient to inform those measures.

Measures that NZTA has considered—representing things such as road condition, outcomes and cost-effectiveness—can be thought of as a part of the demand for data. The other demand for data arises at the RCAs when they construct measures. At NZTA the problem is less the level of demand for data and more the limited supply of data. That leads naturally to the question of why the data supplied to NZTA is too limited to inform the measures that they would need to fulfil their desired information goals.

#### *c) Incentives are not aligned with the desired outcomes*

Data collection in the present system happens at the RCA and contractor level. The RCA commissions the contractor to collect data that the RCA requires, and some that the NZTA requires from the RCA. At present, the data that the RCA is required to

collect on behalf of the NZTA is limited to the annual achievement report, which is insufficient to adequately support the measures that the NZTA wishes to implement.

The NZTA releases guidelines to inform RCAs' choice of measurement and levels of service. However, councils are not bound by those guidelines and may have competing objectives in their LTCCP that the road asset manager is required to follow. As a consequence, there can be a significant divergence between the NZTA's guidelines and the measures collected by the RCAs.

Henning, et al (2011) point to three specific problems with data collection from the NZTA's perspective:

- data is not measured using a common process across RCAs
- data is not equally accurate across RCAs due to differences in the collection method (e.g., digital and visual assessments aren't directly comparable)
- measures of the current state is provided, but not on the rate of change of the measure.

These are all problems that could be alleviated with a common set of data collection protocols, more standardised measures and a wider range of types of measures that use rates of change rather than absolute values and distributions not just central tendencies.

#### *d) RCAs are not a partner in data collection*

A concern RCA road asset managers have was that they were required to pass on data to the NZTA that was not relevant to their job of managing the roads. Some managers appeared not to understand why NZTA would need that information and resented having to compile and send it. In addition, the TIO systems used for passing on some of that data was described as being unnecessarily laborious, which added to their resentment.

There is no doubt that RCA managers have a role to play as providers and protectors of data for others, but the resentment suggests that the NZTA has not effectively communicated the benefits of data provision to the RCAs.

#### *e) The supply of information requires people and tools*

The other factor that limits the collection of data is the supply of tools and skills for interacting with it. In order to effectively collect and analyse data, RCAs require software and tools to easily measure and record the information, along with the expertise to analyse it and use the results.

At present, the development of information systems appears to be separated by organisation: the NZTA is developing its own in-house databases that RCAs have limited access to; and RCAs often use in-house databases to manage their networks, the information from which is not fed up to the NZTA. In some RCAs, the programming is largely contracted out to consultants, who may use their own database tools that feed only summaries of the raw data back to both the RCA and the NZTA.



The lack of integration of data systems makes it very difficult for any one organisation to gain a clear picture of the effectiveness of their maintenance programme. It makes it harder still for that picture of the maintenance programme to be shared across organisations or for the network as a whole.

In addition to the problem of data access, there is also a lack of expertise in using data to inform future maintenance programmes and improve outcomes. Developing those skills across the organisations will allow easier access to data for decision-makers and monitors alike.

### 3.4.3 RCAs need better measures

There is a large amount of data that can be collected during the process of constructing and implementing a road maintenance programme. It is such that, for existing roads, at least the history of their condition, maintenance work done and expenditure is available (through RAMM, for instance). That data is potentially extremely valuable for:

- learning about the response of the road to maintenance work
- learning about the effectiveness of interventions
- modifying the future maintenance programme to account for the lessons learned.

As yet, there are only pockets within the sector—such as NZTA’s highways division—that are making use of the available data in that fashion. A recent performance audit by the Controller and Auditor General observed “In general, NZTA does not systematically or consistently assess information at national level about the wider quality and performance of maintenance and renewal work in areas in the network, to understand differences, trends, and the reasons for them. NZTA has recently begun some initial work to compare maintenance and renewal costs between areas with similar traffic volume characteristics.” (2011 p36)

For the most part RCAs appear not to be either constructing or using the measures that would enable such learning. For example, NZTA surveys of auditors’ reports have found that many RCAs have inadequate maintenance of their RAMM database, under-utilise the data in it and fail to record significant portions of the work done, as well as failing to update default values based upon experience. These failings around data management arise because road asset managers are not using the measures that rely upon the data to manage their assets.

The lack of use of measures at the RCA level appears to be due to a number of factors.

#### *a) Unclear investment requirements*

The primary incentive driving the collection of data is the need to invest in road maintenance. However, there appears to be a lack of clear guidance about the measures, and standards, that are required to gain funding. The NZTA endeavours to work with the standards and measures chosen by the RCA in their AMP, while implicitly applying their own standards based upon their own measures. For example, NZTA staff routinely used the cost per vehicle kilometre as a measure of the cost-effectiveness of an RCA’s network maintenance. That may not be a part of the AMP

and the RCA may not necessarily realise the importance of that measure for funding, yet they are still judged against it.

For RCAs to willingly use more comprehensive measures of road maintenance performance, the NZTA must provide clear information about the measures against which investment proposals will be assessed. Similarly, if particular levels of service are to be targeted, then the NZTA must be clear about the levels required for investment purposes.

### ***b) But investment incentives alone are not enough***

However, sharp incentives via investment guidelines and performance information requirements are not sufficient to motivate RCAs to learn from data analysis. When strong extrinsic motivators such as investment are used as a tool, there is a risk of conformance without performance. That is to say that there is a risk the data will be compiled and analysed—itself a cost to the RCA—without being used for informing future learning and planning measures.

In order for the information to be effectively utilised to support performance, the extrinsic incentives need to be aligned with the RCA's intrinsic motivations. For road asset managers, that means the incentives must be aligned with the desire to build and maintain excellent roads.

The need for alignment raises difficult questions about investment allocation. If funding is allocated to those networks that need it most then there is some risk that it may flow to those who are managing their network most poorly and inefficiently. Thus, a road asset manager who does an excellent job of lifting their asset management performance and demonstrates a reduced the cost of maintaining the network will be 'rewarded' with a cut in their future maintenance budget. That creates dissonance between the intrinsic and extrinsic incentives and risks the manager lapsing from consummate performance into conformance.

In addition, there may well be numerous measures that a good asset manager would use, that are not required in order to justify investment by NZTA. The other functions of performance measurement at an RCA level may require additional measures that would not be demanded by NZTA. Thus, the demand for measurement cannot come wholly from the investor. It must also be remembered that councils are co-investors with NZTA and may have their own reporting requirements of the road asset manager.

### ***c) Good asset management is not well understood***

Asset management planning is a specialised task that requires significant expertise. For large councils with sufficient economies of scale, employing an expert in asset management is feasible. For smaller councils with correspondingly smaller budgets for personnel, it may not be. As a consequence, there are varying levels of expertise across RCAs in planning.

Some support is provided through professional development courses and rapport with other professionals. The National Asset Management Steering Group (NAMS), in particular, is responsible for disseminating good practice guidelines.

Appropriate guidelines can compensate for some lack of formal training; however, formulating guidelines that deal with the multitude of possible council objectives is not possible. As a consequence, road asset managers in councils that do not closely follow the NZTA's maintenance guidelines are left with little support.

The RIMS group is intended to deal with the problem by providing a forum for best practice guidelines to be developed and disseminated. Unfortunately, our interviews suggest that the RIMS groups have had limited success in reaching out to stakeholder RCAs who are not seeking to actively engage with them. Those stake-holders are also likely to be the ones with the least experience and expertise in the necessary planning tasks.

#### *d) There are no obvious models*

For many RCAs the need for better information to support their asset management has not been clearly established. If an asset manager appears to have a healthy network without using advanced measurement and asset management practices then it may appear to them that there is no reason to change. The disparate regimes at RCAs then make it hard to point to case studies where improved information has led to better outcomes for the network: it is too easy to claim that a champion is the beneficiary of special circumstances. Overcoming that problem itself requires some engagement and belief in the relevance of communities of practice.

#### *e) Cost pressures are more salient than efficiency gains*

Gathering data, analysing it and using measures is costly. There are efficiency gains to be had in measurement that can bring down costs, and reduce costs of asset maintenance. Two conclusions follow:

- not all RCAs will find it cost-effective to use all of the measures that are identified as useful for asset management and learning. For some, the costs will outweigh the benefits
- there need to be champions who lead the way and demonstrate to other RCAs that investing in good measurement and asset management is cost-effective.

### **3.5 Conclusion**

The roading sector has much of both the data and capability and is starting to more actively use the data to learn about what works and when. NZTA Highways division has initiatives underway to increase its ability to learn about the effectiveness of highway programme spending. Similarly, as asset management practices improve and an increasing number of RCAs move to more performance-based contracting there will be a requirement for RCAs to use performance information in different ways than has occurred before. New Zealand is potentially on the cusp of a cultural shift to enable changes in behaviour that would allow performance information to shape decision making on roading maintenance.

We have identified two major problems with the formal system. The first problem is that there is no cycle of performance measurement. Rather there is an open loop with limited direct and indirect feedback from the monitoring and measurement subsystem to the design of programmes and policies. This lack of a learning loop is

impeding performance improvements being identified and implemented. The lack of measurement of the effectiveness of interventions, such as pavement treatments, inhibits the learning and development of asset management expertise.

The second problem is a disconnect between the data supplied and the presently unmet demand for different data and more measures. While there is a formal process, this is very shallow and focuses on a few core objectives measures such as roughness that do not adequately capture overall performance. There is unmet demand both from RCAs and NZTA for a wider range of better measures—partly to enable more effective investment allocation—but also to improve learning about effectiveness.

The practice in the sector is to work around the disconnect using systems based on informal collaboration: RCAs that use the RAMM database software allow NZTA to access their raw data and use it to construct measures. However, RCA has discretion over their data collection, measures used and standards of service so RCAs' practices vary greatly. That makes the data collected through informal collaboration difficult to use for research and comparison purposes.

This leaves gaps not remedied by the informal practises such as data not stored in RAMM, information on the effectiveness of treatments, data on gravel roads, bridges etc. The lack of a national framework covering what should be measured (standards and associated measures) and how it should be measured (data protocols) inhibits learning.

## 4. Recommendations for improving performance measurement

The previous section reviewed the formal system of performance measurement for roading maintenance and compared that with the real system that practitioners actually use. It highlighted the contrast between the effort placed on the supply side in terms of collection of data and the lack of demand side pressure from users.

Achieving a high performing road maintenance measurement system focused on learning would require addressing a number of the shortfalls identified in the research on the current real and formal system. The research identified a number of common themes including the need for:

- a common national performance framework including
  - common data protocols
  - enhanced measures
  - common standards on levels of service
  - improved data infrastructure to enhance data input and user access
- focus upon measuring the effectiveness of road maintenance interventions
- learning through communities of practice
- enhancing the capability of people to work in a different way
- support by leadership, engagement, focus and commitment.

Achieving the step change sought by the Task Force will require a number of mutually reinforcing strategies to address these themes.

### 4.1 Strategies

#### 4.1.1 A common national performance framework

A common national framework covering data collection, performance measures and standards of service would provide the opportunity to improve overall system effectiveness and to enable valid comparisons and benchmarking. All management frameworks are a mix of being ‘tight’ and ‘loose’.

#### 4.1.2 Data

Better data collection provides the foundation for better measurement. To enable learning across the system, the framework would need to be ‘tight’ on how data is defined, measured, collected and stored. This could require a national system of roading classifications, common data definitions, and common measurement techniques to record performance achieved against different levels of service. There are existing models, such as the data standard and data definitions developed by the UK Highways Agency (2011), that may be able to be adapted for New Zealand conditions. Some of these data requirements already exist in some areas in New Zealand like measures of road roughness but the coverage of uniform requirements would need to be extended more widely.

Data management requires defining data ownership, stewardship and custodianship roles and should be consistent with the new principles for government data approved on 8 August 2011. These provide that in general government data should be well managed, open, and readily available. <http://ict.govt.nz/programme/opening-government-data-and-information/new-zealand-data-and-information-management-principles>.

### 4.1.3 Measures

Section 3 discussed that currently while there is extensive data collected, NZTA requires relatively few measures to be reported by RCAs. Some of those measures such as roughness, while useful for predicting pavement life, are inappropriate as a measure of service quality. A draft NZTA research report (Henning et al 2011) discusses the development of a new measure – the vehicle operating cost index (Part 7), gaps in what is currently measured (such as bridges Part 8), the need for more sophisticated measures (combining current rate and rate of change for a given parameter p37) and the need to separate the measure from the method of data collection.

The development of any further measures needs to proceed on the basis that there are many different perspectives on performance and different users of performance measures require different performance information for different purposes. For example, the performance measures required for supporting contracting will vary depending upon the nature of the contract (output, hybrid, outcome or alliance). The initial priority is measures of treatment effectiveness. Different measures again may be required at the network level and the overall; the national system level. There need be no 'golden thread' that links one set of measures to others although the relationship between the measures needs to be understood.

Development of measures should not occur in isolation from the lessons and practices of comparable jurisdictions overseas. The Association of Australian and New Zealand road transport and traffic authorities (Austroads) <http://www.austroads.com.au/performance-indicators> provides a forum to compare and learn from others' experiences with performance measurement and share the costs of developing policies, procedures, guidelines and tools.

It is recommended that a group be established to develop a suite of performance measures required to support oversight of investment in roading maintenance. While the draft NZTA research report provides a useful base, it is important to accommodate the different perspectives and recognise that one size doesn't fit all.

### 4.1.4 Standards of levels of service

The current practice is that NZTA provides guidelines on roading maintenance standards for levels of service. These standards are provided as guidance and RCAs have discretion about whether and how much is adopted.

Any framework for standards of service would have to be 'loose' on the level of service selected allowing regional variation (both within NZTA highways and local RCAs) on what levels are set within an agreed framework of standards (and

associated measures). The aim is to achieve national consistency in how standards are expressed while allowing regional variation in what aspects of performance are most important and what level of performance is required within the standards framework for each in a particular location.

Ultimately this framework could be augmented by evidence-based policy priorities to enable amenity values to be compared and traded off against other factors such as journey times, journey reliability, and safety. This last step could only be contemplated when the other elements discussed below are in place and operating effectively. Experience in other sectors such as the Health Core Services Task Force in the 1990s suggest that moving to prioritisation frameworks quickly can be a leap too far too fast. A staged measured approach to developing the framework is required.

In order to make progress on developing the national framework, NZTA (in partnership with RCAs and industry consultants and constructors) needs to commission an independent broker (such as Standards New Zealand) to develop a common system of national standards for levels of service. The standards would need to consider a range of factors such as levels of service, ability to pay, risk, whole of life issues, local conditions (natural resources, climate) etc.

#### 4.1.5 Data infrastructure

While New Zealand was a pace setter with developments such as RAMM in the late 1980s, in recent years new system developments have not kept pace with developments in ICT that would provide an easier, more user-friendly interface. This resulted in New Zealand falling behind the leaders in the field. The lack of user friendliness in ICT results in errors in data entry and lack of use, which combine to create problems with data quality.

NZTA has some development work underway on a proof of concept for a data warehouse but this is not currently available to users. New technologies also have the potential to ease the costs of non-traditional data collection.

Data infrastructure is a 'chicken and egg' problem. Without greater demand for data and analysis from users, greater investment in data infrastructure would be wasted on its own. But without investment to make it easier for users to supply data and tools that make it easier for users to undertake analysis, the other strategies will fail.

The recent experience with the dTIMS project suggests that once there was a use for much of the data, there was increased pressure to improve the quality of data sets.

To progress this issue NZTA needs to convene a user group to develop user requirements for improved data input and data access. A range of perspectives covered will need to be covered (citizen-ratepayers as well as RCAs and industry) to ensure that the range of external purposes (accountability and legitimation) as well as internal purposes (learning, budgeting) are covered.

Emphasis should be placed on small 'throw-away' pilots that allow 'fast failure' and rapid learning. Rapid learning does not imply, of course, that any move to full production in a data-mart or data-warehouse type environment should not follow the

usual disciplines for business cases and procurement procedures. There would need to be an evaluation of competing technologies against a defined set of performance expectations in the normal way consistent with good asset acquisition practice.

#### 4.1.6 Initial focus upon measuring effectiveness of treatments

Establishing a small community of practice group on the effectiveness of road maintenance treatment projects offers great potential for quick wins because this would utilise existing data sets and platforms to provide new information that will help asset manager fit increasing work programmes within flat lined budgets. Measures focus on programme efficiency and treatment effectiveness will be important.

The current position is that much data is collected, and there is significant analytical capability throughout the sector, but knowledge is still lacking on the effectiveness of maintenance treatment interventions. NZTA as operator of around half of the network and as a major investor in local RCAs has a hugely important role to play.

There is the potential for a 'quick wins, by taking a strengths-based approach to developing a deeper understanding of the effectiveness of the programme of maintenance work. One way to make early progress in this area is to develop pilot projects using existing data sources based on successful models from overseas to develop knowledge on the returns on the roading maintenance investment. By making use of existing platforms and data sources, it often is possible to generate new insights into questions such as the quality of the overall management of the portfolio of assets, how individual assets are performing and what the trends are in asset usage. While acknowledging the limits of existing data, the robustness of the findings can be challenged and tested in further research. Leadership is critical in taking the first steps. It will be important in this work to not 'let the search for perfection' get in the way of 'making good progress'.

#### 4.1.7 Learning

The current position is that while New Zealand's roading sector has historically been a pioneering world leader in developing system-wide data management tools such as RAMM, and collaboration through forums like the Road Industry Management Steering Group (RIMS), that lead has been eroded in recent years. Information technology and data analysis techniques have continued to develop rapidly but the take up of new tools such as data-marts and data-warehousing has not occurred. RIMS recent work focused on predictive modelling using dTIMS and is now moving onto the development of good practice guidance for RCAs' road assets management systems.

If New Zealand is to keep pace with innovations in other leading jurisdictions then it will need to continue to experiment and innovate. Austroads provides a forum to compare and learn from others' experiences with performance measurement. RIMS provides an example of collaborative working and engagement between NZTA and RCAs. Looking ahead the sector needs to build on lessons from RIMS to develop a new model of sector-wide learning based on communities of practice, and good practice exemplars.



To progress system-wide learning NZTA, in partnership with RCAs and the industry, could work with an organisational learning management specialist to review the lessons learned from RIMS (along with comparable models in other sectors and jurisdictions). Based on that review recommendations need to be developed for a new model of sector-wide learning based on communities of practice, and good practice exemplars.

#### 4.1.8 People capability

There are a number of strands to the development of the capability of people in the roading sector. Professional bodies such as Ingenium (public engineers), IPENZ, and the roading ITO all provide continuing education to professional engineers. However, the roading authorities vary significantly in their size (or capacity) and organisational capability and the smaller RCAs are reported to experience problems with critical mass. NZTA has the five star project underway to assess RCAs' technical capability.

Achieving a step change in system performance will require a step change in capability and behaviours. The Task Force has identified the importance of capability with a work stream focused on clustering and a separate research scope in another stream of work covering skill development. Accordingly more detailed examination of people capability issues is out of scope for a project focused on performance measurement. It is important to note, however, that the collaborative ways of working required for the development of national performance framework, data infrastructure etc. could equally be applied to better facilitating the sharing of lessons learned about building people capability.

### 4.2 Supporting strategies - engagement, focus and persistence

The previous discussion has focused on *what* needs to be done. *How* performance measurement systems are developed is equally important. A key finding from the literature review was that without buy-in from professionals and partner organisations problems of conformance, gaming and cheating become more prevalent.

#### 4.2.1 Engagement and partnership

The performance measurement regime needs to be developed in partnership with RCAs and industry and not imposed by NZTA from above. How performance is expressed in contracts and other documents and how measures are developed becomes very important. Performance measurement creates incentives to maximise the measured performance, with the least amount of effort within the rules.

A system imposed on an organisation from above is likely to encourage perverse strategic behaviour or gaming because imposition provides professionals with justification to not support, trust, use or take the system seriously. By contrast performance measurement systems that are designed collaboratively by both managers and professionals promote broader ownership, more supporters, greater trust and then utility. Measures need to be developed in partnership with stakeholders (including citizens-rate payers) rather than imposed from above if they are going to be used and effective.

### 4.2.2 Focus

The discussion of the current formal performance measurement system showed that there were multiple purposes. The current system includes a predominant focus on management information to support investment decisions and programme planning and a mix of learning and formal reporting for accountability and legitimisation purposes, with some contract management.

In the future, the formal system needs to be focused on system learning. Putting performance measurement in a learning frame means that performance information on inputs, outputs and impacts is used (in Hood's terms discussed above) as a part of an intelligence system and not targets as the basis for name and shame consequences. Dialogue and learning need to proceed on the basis of shared understanding about the limits of performance information as well as the potential role as diagnostic learning devices.

### 4.2.3 Persistence

Achieving changes in professional practices and behaviour is a long term challenge. Making progress takes sustained effort and commitment to change. Without a champion and fellow travellers committed to making a difference, the programme is doomed to fail. The history of performance measurement is a history of setbacks, dashed expectations and outright failures. Achieving sustained changes in behaviour is hard. Great care is required in the sequencing of the programme of change. The changes to the performance measurement regime will need to be well integrated with the Task Force's recommendations on contracting, skills and capability development in particular.

### 4.2.4 Directions for reform

While there is no silver bullet, the Task Force has been established because there is a sense of urgency that creates 'a burning platform' and the need for change. That urgency is created by the flat lining of maintenance budgets at a time when the network is still expanding and the cost of maintenance expenditure is increasing.

The strategies discussed in this paper will inevitably require the commitment of time, effort and scarce financial resources by NZTA in particular. This creates the classic hump funding problem – the costs of implementing the recommendations in this paper can increase the immediate pressure on maintenance budgets. Two considerations are relevant here:

- one has to invest cents up front to save dollars down the track
- the scale of spending required is dwarfed by the overall investment in roading maintenance (NZTA invested \$500 million in 2008/09 on maintaining and operating the roading network).

The size of the investment required will also be reduced to the extent that already existing models (such as UK data definitions and standards) can be adopted and modified, and existing relationships with Australian roading authorities can be leveraged through Ausroads.

In this report we have addressed how to improve the road maintenance performance measurement system. We have **not** focused on the development of multiple additional new performance measures, in part because a detailed draft report on this is already available, as well as because of the limited resources and time available. The main reason, however, is that providing new measures does not address the systemic issue which is the lack of use of the existing suite of performance data.

The research scopes for the Task Force that resulted in this report being commissioned include:

- are the sectors' current indicators the right indicators, measured at the right frequency, for predicting future maintenance and renewal requirement and for benchmarking – and if not what indicators should be used?
- what perverse behaviours are driven by the Road Control Authorities (RCAs) performance measurement regime?
- what innovative approaches (including international) to performance measurement and monitoring are there that could be used to help RCAs improve performance in maintenance and renewal?

Accordingly we have organised the recommendations into the three broad areas applying to road maintenance:

- improving performance measurement
- identifying the limits of performance measurement and reducing the risk of perverse behaviours
- identifying and improving the take up of innovative practices.

Work on developing a common national performance framework in particular will require sustained commitment of time and resource before tangible yield is evident. That is an argument for starting the long march now so long as it is accompanied by actions that yield more immediate results. By contrast, work on measuring effectiveness offers the greatest potential for quick wins. If these recommendations are adopted, then the sequencing suggested in the recommendations would need careful design to mesh in with the findings of the other work strands and research scopes.

One of the main lessons from this work is that how the overall programme is designed, implemented and governed is as important as what work is conducted. A key challenge is to develop a governance mechanism to guide the development of the design of a common national performance framework, monitor progress with implementation and oversee change management. This will require active leadership from NZTA and active followership and participation from RCAs, constructors and consultants. The fact that we have not provided detailed recommendations on this as part of this report, as the governance mechanisms need to be designed as part of the overall Task Force's recommendations, does not suggest that this is not critically important.

## 4.3 Improving performance measurement

### 4.3.1 First phase

The first phase is focused on quick wins (treatment effectiveness) and developing common foundations (data protocols, performance measures, data infrastructure).

**Recommendation 1 - Treatment effectiveness** - NZTA to establish a team drawn from across the sector to assess the effectiveness of road maintenance treatment projects using existing data and analysis platforms.

**Recommendation 2 - Data protocols** - NZTA, in partnership with local RCAs and industry, to establish a team to develop how road maintenance data needs to be defined, collected and stored to support learning.

**Recommendation 3- Performance measures** - NZTA, in partnership with local RCAs and industry, to establish a group to develop a suite of performance measures to support oversight of investment in roading maintenance that builds on the draft NZTA paper but accommodates different perspectives and includes treatment effectiveness, programme efficiency, economy measures as well as output delivery.

**Recommendation 4 - Data infrastructure** - to support Recommendations 1-3, a user group from across the industry should be convened by NZTA to develop user requirements for data input and data access and explore the implications for IT platforms (including possible small 'throw-away' pilots).

### 4.3.2 Second phase

**Recommendation 5 - Level of service standards** - NZTA, in partnership with local RCAs and industry, to commission an independent broker (such as Standards New Zealand) to develop a common system of national standards for levels of service required for road maintenance.

**Recommendation 6 - Oversight of implementation & change management**– Develop an industry wide governance mechanism as part of the overall Task Force's recommendations to guide the development of the design of a common national performance framework, monitor progress with implementation and oversee change management.

## 4.4 Identifying the limits of performance measurement and reducing the risk of perverse behaviours

The implementation of the recommendations in section 4.2 above, need to be supported by incentives, sector engagement and collaboration, and a focus on learning:

**Recommendation 7 - Incentives** - NZTA provide incentives for the use of performance information by requiring each RCA as part of the application to the NLTP to identify its strategy for learning about the effectiveness of the proposed investment.

**Recommendation 8 - Sector collaboration** - Reduce the problems of perverse strategic behaviour and gaming by working collaboratively with managers and

professionals in RCAs and industry to motivate the desired behaviour and promote broader ownership, support, trust and utility.

**Recommendation 9 - Focus on learning** - Performance information should be used as part of an intelligence system. There should be limited use of league tables and targets should be avoided.

## **4.5 Identifying and improving the take up of innovative practices.**

The recommendations above will require the development of new ways of working based on new models of learning:

**Recommendation 10 - Learn from current models** – In order to develop the new models of learning required, NZTA in partnership with RCAs and private industry, should work with an organisational learning management specialist to review the lessons learned from RIMs (along with comparable models in other sectors and jurisdictions).

**Recommendation 11 - New learning models** - Based on that review, to develop recommendations for a new model or models of sector-wide learning based on communities of practice, and good practice exemplars.

## Appendix A References

Most of the references included below are cited in the review. A few references are included which have not been directly cited, but are significant contributions to the literature judged to be worth including.

### A.1 Academic references

Behn, R. D. (2003). 'Why measure performance? Different purposes require different measures.' *Public Administration Review* 63(5), 586–606

Bevan, G., Hood, C. (2006). *Health policy: Have targets improved performance in the English NHS?* *BMJ* p332:419

Bouckaert G. & Halligan J. (2008). "Managing Performance – International Comparison". London: Routledge

de Bruijn, H. (2006). "Managing Performance in the Public Sector", 2<sup>nd</sup> ed. New York, NY: Routledge

de Lancer Julnes, P. (2008). "Performance Measurement Utilization: Beyond Instrumental Use". In Van Dooren, W., and Van de Walle, S. (eds.) *Performance Information in the Public Sector: How it is Used* (58-71). Palgrave Mcmillan: London

Gill, D. (ed) (2011). "The Iron Cage Revisited: The Performance Management of State Organisations in New Zealand ". Wellington: Institute of Policy Studies

Hatry, H. (2008). "Emerging developments in performance measurement: An international perspective", in de Lancer Julnes, P., Berry, F., Aristigueta, M. and Yang, K, (Eds.). *International Handbook of Practice-Based* , Sage Publications: CA.

Ho, A-Tat-Kei. (2008). "Reporting Public Performance Information: The Promise and Challenges of Citizen Involvement" in Van Dooren, W. and Van de Walle, S. *Performance information in the public sector: how it is used*. Basingstoke [England]; New York: Palgrave Macmillan

Holzer, M. & Kolby, K. (2008). "Helping government measure up", in de Lancer Julnes, P., Berry, F., Aristigueta, M. and Yang, K, (eds.). *International Handbook of Practice-Based Performance Management*. Sage Publications: CA

Hood, C. (1991). 'A Public Management for all Seasons'. *Public Administration* Vol. 69 No 1.

Hood, C. (2006). 'Gaming in Target world: The Targets Approach to Managing British Public Services'. *Public Administration*.

Hood, C. (2007). *Public Service Management by Numbers: Why Does it Vary? Where Has It Come From? What Are the Gaps and the Puzzles? Public Money and Management*.

Kaplan, R. S. and D.P. Norton (1996). "Using the balanced scorecard as a strategic Management system." Harvard Business Review (January-February): 150–161.

Lodge, M. and D. Gill (2011). *Toward a New Era of Administrative Reform? The Myth of Post-NPM in New Zealand*. Governance 24(1): 141-66.

Meekings, S., and A. Neely (2009). "How to Avoid the Problems of Target-Setting". landmarkconsulting.co.uk/.../10-09-28-PMA-Symposium-Target-Setting- Paper-v1.0.docx.

Moynihan, D. (2008). "Advocacy and Learning: An Interactive Dialogue Approach to Performance Information Use" in Van Dooren, W. and Van de Walle, S., *Performance information in the public sector: how it is used*. Basingstoke [England]; New York: Palgrave Macmillan.

Moynihan, D.P., and S. K. Pandey (2010). "The big question for performance management: Why do managers use performance information?" Journal of Public Administration Research and Theory 20 (4): 849-866.

Moore, M. H. (1995). "Creating Public Value: Strategic management in government" Cambridge, MA: Harvard University Press.

Pidd, M. (2008). "A critical assessment of performance measurement for policy making", in Hartley, J. *Managing to improve public services*. Cambridge, UK ; New York: Cambridge University Press.

Pollitt, C. (2006). "Performance information for democracy: The missing link?" Evaluation, 12(1): 38–55.

Radnor, Z. (2008). "Hitting the Target and Missing the Point? Developing and Understanding of Organizational Gaming" in Van Dooren, W. and Van de Walle, S. *performance information in the public sector : how it is used*. Basingstoke [England]; New York: Palgrave Macmillan.

Talbot, C. (2008) 'Competing public values and performance.' In KPMG International (ed) *Holy Grail or Achievable Quest? International perspectives on public sector performance management* (p 141–152). KPMG International.

Thomas, P. (2006). "Performance Measurement, Reporting, Obstacles and Accountability." ANU E Press: Canberra.

Van Dooren, W. G., G. Bouckaert and J. Halligan (2010). "Performance Management in the Public Sector". New York: Routledge.

## A.2 Practitioner documents

Austroads (2011). *National Performance Indicators*

<http://www.austroads.com.au/performance-indicators>.

Controller and Auditor General (2011). *New Zealand Transport Agency: Delivering maintenance and renewal work on the state highway network*. Performance audit report.

Covec (2011). *Blueprint for a Best Practice Measurement Indicator Set and Benchmarking: Literature Review*.

Henning et al (unpublished 2011). *NZTA Performance Indicator Analysis and applying Levels of Service* draft NZTA research report number X.

Hutt City District Council (undated). *Draft Asset Management Plan*.

Government Chief Information Officer *New Zealand Data and Information Management Principles* <http://ict.govt.nz/programme/opening-government-data-and-information/new-zealand-data-and-information-management-principles>.

International Public Sector Accounting Standards Board (2012) *Conceptual Framework for General Purpose Financial Reporting by Public Sector Entities* IFAC

Kapiti Coast District Council (September 2011). *Draft Asset Management Plan*.

NAMS (2011). *International Infrastructure Management Manual* (section 2.2).

NZICA (2002). *ICANZ Technical Practice Aid No. 9*.

NZ Transport Agency (2012). *Audit Findings Review*.

NZ Transport Agency (2007). *Maintenance Guidelines 2007 - Land Transport programme*.

NZ Transport Agency (2007). *Procurement manual for activities funded through the National Land Transport Programme First edition*, effective from July 2009 Amended March 2011.

NZ Transport Agency (2008) *Planning, programming and funding manual*.

NZ Transport Agency (undated). *Rules/Regulations/requirements relating to NZTA approved maintenance optimisation models and RCA pavement data collection*.

OECD (2009). *Focus on Citizens: Public engagement for better policy and services*. Paris: Organisation for Economic Co-operation and Development.

OPUS (2012 draft). *Review of Delivery Models for Works and Services*.

SSC and Treasury (2008). "Performance Measurement: Advice and examples on how to develop effective frameworks". Wellington: State Services Commission and Treasury. [www.ssc.govt.nz/display/document.asp?DocID=7121](http://www.ssc.govt.nz/display/document.asp?DocID=7121).

Transfund (2002). *Competition Monitoring Summary 2002-03-GB571506*.

UK Highway Agency (2011). *Performance Management Manual* <http://www.highways.gov.uk/business/15098.aspx>



## **Appendix B People met**

### **B.1 Task team members**

NZTA – Janice Brass, David Darwin, Mark Edwards

Industry – Adam Bevin and Ian Greenwood

RCAs – John Sutton

### **B.2 Workshop participants**

NZTA – Janice Brass, David Darwin, Rob Merrifield, Matt Hendry, Balt Gregorious, and Brandon Mainwearing

Industry – Adam Bevin, Richard Parsons and Ian Greenwood

RCAs – Julie Muir, John Sutton, Vaughan McEwen and Craig Thew

### **B.3 Individual interviews**

Brandon Mainwearing – NZTA

David Darwin – NZTA

Gordon Hart – NZTA

Matt Hendry – NZTA

Michelle Lewis – KDC

Ron Muir – HCC