Schiff Consulting.

Getting more value from data sharing: Some potential solutions

28 February 2017

Hayden Glass (hglass@srgexpert.com, 021 689 176)

Aaron Schiff (aaron@schiff.co.nz, 021 288 0665)

Summary

1. This is the second report on our project for the Data Futures Partnership Working Group (DFP) to explore barriers to the generation of value from data sharing in New Zealand and develop potential solutions.

2. We are using the word "sharing" in the non-technical sense of the word, i.e., data collected by one organisation being used by another. Such sharing does not necessarily require that the data is made publicly available.

3. The first paper looked at what "value" means, explored what the literature tells us about barriers to data sharing and, after discussion with the DFP, settled on three barriers to focus on in this stage of the work. We repeat some material from that first paper at the beginning of this one so that this paper can stand alone.

4. Because each of the three barriers is large on its own, we use just one narrow point to demonstrate each barrier and how a solution might be approached:

    *   Data sharing processes – We explore the work involved in sharing data publicly by a small number of large central government agencies. We conclude that process guidance alone will not help much (there is already a lot of this). The barriers facing public data sharing from government agencies are more about mandate, motivation and focus than a lack of processes. That said, support for agency efforts to move towards robust data governance and practical advice on how to publish data could both be helpful.

    *   Agency business models – We look at residential real estate data. A combination of the incentives faced by territorial authorities (TAs) to sell property market data and commercial relationships between TAs and Quotable Value (an SOE) have limited the availability and use of residential property market data. Substantial benefits for home buyers, the real estate industry, and property developers could be generated if this data was collected by a central agency with a mandate to publish unit record data on residential property transactions openly.

    *   Data sharing standards – We look at standards for recording data with a geographic dimension that is not a pin-point location, such as a city or a region. Many different geographic schema are in use, even just in the public sector, and this makes it difficult to combine data from different sources for analysis, or to authoritatively answer simple questions like how

many people live in Grey Lynn or in Karori. The costs created by the need to reconcile different geographic standards ripple out across the whole ecosystem, meaning that every organisation that collects location-based data faces higher costs to generate value from its data.

5. Our focus in this work is mostly on public agencies (rather than private sector or not-for-profit organisations) and on data that could be made open (rather than other forms of data sharing). There are clear implications from the barriers for private sector operators also, limiting access to data and increasing costs. We draw some of these out in this paper. There are doubtless Diagnose and Fix issues elsewhere in the ecosystem, and making data open is not the only way to generate value from it. We have not explored these issues in this project for want of time. We review in the first Report in more detail how we and the DFP chose these three barriers to focus on.

6. We also present a solution that addresses the three barriers (and possibly others) at once in quite a different way, built from the Basic Data initiative in Denmark. In this approach, one agency would curate a standard set of geographic, property, business and personal data and make it available openly. There would be controls on some data to protect confidentiality and personal privacy. Non-confidential, non-personal data would be publicly available for free. The data would be authoritative and use common data definitions to make analysis and integration easier.

7. We recommend that the DFP support a project to define in more detail what this Basic Data approach could look like if applied in New Zealand, engaging with potential users to ensure that the product design fits with their needs. An initial project, including standing up a prototype that provides some actual data could be completed by October this year in time to include in the DFP's report on its progress so far to Ministers.

8. The DFP could also:

    •   Encourage government agencies to put together and publish a list of data assets they hold (including those not publicly available), what they contain, and what condition they are in, and support this encouragement with practical assistance and public praise for agencies that are doing this well.

    •   Engage with agencies that publish data using non-standard geographic definitions to understand their reasons for doing so, and support and encourage them to adopt consistent geographic standards (and standards for other characteristics too).

    •   Dig in more detail into the costs and benefits of opening up data on residential property transactions, and developing a preferred solution for that work for Ministers to engage with.

9. The table on the next page summarises the barriers, the potential options and our suggested approach.

Summary of barriers, potential solutions and recommendations (with recommended approaches shaded)

	Topic	Barrier	Fix A	Fix B	Fix C	Fix D
1	Data sharing processes	Lack of guidance or standardisation of processes inhibits publication of open data. Focus on government agencies.	(1A) Work with a small number of large agencies on an in-depth assessment of the state of their processes and progress in publishing open data. Develop a detailed view of the most helpful assistance to accelerate progress. Focus: More data publishing by the most important agencies	(1B) Encourage forming of clusters of agencies that are more advanced in their domain with those that are less advanced. Focus: Lift the number of agencies publishing open data, and coordinate data publishing within sectors	(1C) Encourage and support agencies to develop something like a fixed-asset register but for core data holdings, showing their extent and their quality. Focus: Lifting visibility of what data is collected (whether it is available openly or not)	(1D) A single specialist agency made responsible for data publication (discussed further under barrier 4 below) Focus: Faster progress in availability and usability of open data
2	Data sharing standards	Different datasets use different classification systems creating extra costs for everyone. Focus on geographic data.	(2A) Ensure that all primary data with geographic characteristics records locations at particular points from which aggregated statistics for any area can be calculated. Focus: An enduring but costly and potentially infeasible fix	(2B) Encourage government agencies and private sector operators to adopt Statistics NZs geographic standards. Focus: A practical way forward		
3	Government agency business models	Some government agencies have control of important datasets and business models based on restricting access. Focus on residential property data.	(3A) Require TAs to provide property market data they collect to another government agency and allow this data to be made public. Focus: Fixing the issue at point of aggregation	(3B) Require conveyancing lawyers to provide data about property market transactions to a central government agency. Focus: Fixing the issue at the point of collection	(3C) Investigate this issue more fully, generating data on options, and the costs and benefits of change, and raise awareness of the barriers created by business models in general Focus: Building a case for change in how value is derived from residential property data	
4	Another way forward	An alternative approach to resolve all three barriers (plus others) at the same time, for some particularly valuable datasets	Assign a single organisation to assemble and publish specific high-value datasets for government agencies and the public to use (in addition to the existing agency-by-agency approach). Focus: An alternative to the agency by agency approach to data release			

Schiff Consulting.

Purpose

10. This paper explores the three barriers identified in the first stage of this work: data sharing processes, standards, and agency business models, and looks at potential solutions.

Background

11. The Data Futures Partnership Working Group (DFP) asked us to help to identify some "barriers to generation of value for New Zealand from the trusted use of data" and possible ways to reduce or remove them.

What this paper is about

12. This paper is about data sharing, i.e., the use of data by someone other the organisation that collected it. We do not talk about issues of data collection, except to the extent that what the collector does with the data it collects makes a difference to the usefulness of the data once shared.

13. We have generally focused on data collected by government agencies and use of that data outside the state sector. We leave a lot of terrain uncovered for want of time and to ensure focus. There would doubtless be value in looking at data sharing in other parts of the ecosystem, especially data collected by the private sector.

14. Data sharing can be one-off or repeated. It can be unilateral (which we also call "data release"), bilateral, or multilateral. We found it useful to distinguish data sharing for a particular purpose, e.g., IRD sharing student loan data with border agencies to enable better enforcement of student loan rules, with multi-party data stores that can be used for many purposes, e.g., the IDI.

15. This is the second report of our work. A first report defines what we mean by "value" when we talk about value being generated from data sharing, explains how value is created and who is involved in this process, and explores some barriers to generating more value from data sharing for New Zealand.

Value comes in many forms

16. It is helpful at this point to recall some material from our first report on what "value" of data is.

17. Value is a broad concept that could include a wide range of benefits for individuals, businesses, government organisations, and society as a whole.

18. The DFP's terms of reference are to "increase the value being generated by New Zealand's data", to "broker and stimulate more data driven innovation", with a principle that "New Zealand should use data to drive economic and social value and create a competitive advantage".

19. Focusing just on specific types of value makes it easier to measure the scale of benefits and how they are generated, and it simplifies judgements about what counts as "valuable".

20. One set of value is that associated with economic activity, which we call "economic value" and measure through market transactions. Economic value can be generated by business activity, and by the actions of government agencies in markets. Economic value accrues to businesses and their owners in the form of profits and to individuals and consumers in the form of greater wellbeing (or "welfare") from market transactions.

21. But economic activity is not the only factor that affects people's wellbeing. Other aspects of value derived from data might be improvements in social values (e.g., feelings of community safety from better information on crime), environmental values (e.g., increases in the efficiency of irrigation where water is provided without charge).

22. There might also be value derived through data from improvements in government services. This could be because of better understanding of public policy problems, better targeting of interventions, easier evaluation of existing initiatives, or more transparency in government that leads to greater citizen engagement in the processes of governance.

23. We expect individuals, businesses and government agencies to try to generate and secure value where they can. We therefore focus on barriers to the incentives or ability to generate and capture value from data.

24. Such barriers may arise from unwillingness of data subjects to have their data shared. We treat these issues as part of the conversation around "social licence", which has been excluded from our remit.

25. Note that the value of data sharing and attitudes or barriers to sharing may differ in different situations. For example, databases of street addresses have many valuable uses and are not considered to be very sensitive for most people (e.g. the electoral roll is published). But if addresses are connected to other data, such as health or criminal records, the value of shared data and the acceptability of sharing this data may be quite different.

26. Also, our concept of value is instrumental, i.e., value comes from use. But one can also conceive of more intrinsic value in data, e.g., value that comes about because the data is somehow associated with a person or an area and is valuable to that person in itself regardless of whether anything is done with it.

27. In this paper we are working under the general assumption that sharing more data between organisations than is done today is valuable (within some limits) and that more sharing could create more value.

The structure of this report

28. In our first report and via discussion with DFP, we agreed on three barriers to focus on in this second report:

- Data sharing processes: Each individual agency must create its own bespoke process for sharing data with others. There is no detailed guidance on how, when, and with whom to share data, how to address ethics, confidentiality, and privacy issues, and how to revise or correct errors in data that has been shared. This creates costs and risks that weaken the incentives of agencies to share data.

- Data sharing standards: Data that is shared is coded in many ways, stored in many formats and systems, definitions are not standardised, and the quality and coverage of metadata is variable. This limits the ability to combine data from different sources for analysis, and imposes costs that weaken the incentives to use data that has been shared.

- Agency business models: Some government agencies (e.g. NIWA, QV, local councils) have business models based around selling data that severely restrict sharing and use of their data, limit experimentation to discover new uses, and make it difficult to combine their datasets with others.

29. This paper is in four parts. The first three explore each of these barriers in turn, using a case study approach to define the barrier and look at what potential solutions might look like. The fourth part develops a solution based on the Danish Basic Data initiative that could address all three barriers at once for a subset of data.

Data sharing processes

30. The first barrier of the three we focus on is about simplification of processes for data publication.

31. As a way to explore process barriers to publishing data, we spoke to a small number of people from LINZ, MBIE, MFE and MPI. We wanted to understand the processes they used for releasing data externally and look at whether there was scope to standardise those processes to simplify their work and reduce costs. The goal of process improvement is to increase the volume of data made available for people inside and outside the agency to use.

Our understanding of the problem

32. Every agency with responsibilities under the Declaration of Open and Transparent Government has to do many of the same things. These include defining "high-value" data, figuring out how best to publish it, and dealing appropriately with requests for data.

33. There is assistance available in the wording of the Declaration itself, in the approved NZGOAL legal framework, and through the small Open Government Data programme, which can provide advice. Agencies have staff members nominated internally to help encourage and coordinate agency open data efforts. An annual report to Cabinet on progress in implementation of the Declaration that is routinely made public provides some motivation. There are other related efforts, including the Register of published data at http://data.govt.nz.

34. At the highest level, we can see that this system is not very effective, since so little data is published openly outside of agencies with direct publishing

responsibilities, and so little of what is published is published in compliance with the Declaration.

35. The hypothesis behind this barrier was that guidance or standardisation of processes might help ensure more data was published. It could simplify decision-making and execution for agencies and might help lower the costs of publishing usable data openly. We know, as covered in our first report, that publishing data in such a way that others can easily make use of it requires a lot more effort than just publishing data alone. Data needs to be high quality, and have decent metadata, good information on provenance, and secure open licensing if others are to use it.

36. From our very small sample of discussions with public service departments we conclude that the problem is not what we thought it was.

37. There is no shortage of guidance within the agencies we spoke to on how to process and publish data openly. The Open Government Data programme makes available useful information and documents that can and have been used for developing agency policies. Agencies have taken these and written their own internal guidance on how to proceed.[1] The decentralised decision-making model and the different circumstances of particular datasets and agencies mean that there are limits to how useful generalised guidance can be in any case.

38. Very practical guidance might be useful (since the existing material focuses essentially on legal and process questions). Guidelines explaining how organisations should organise and present their data to facilitate effective sharing with others, or exactly what wording to use to ensure that data collected under contracts can be published openly might be useful, for example. Teaching agencies how to create "data products" that can be used and re-used by others could facilitate uses of data that haven't been anticipated in advance. This work needs practical examples, not just conceptual models.

39. That said, the big barriers we perceive are data quality issues, issues of governance and mandate, and questions about the usefulness of the datasets themselves for answering the questions that users actually have. Providing practical guidance or improving processes will not address these big barriers directly.

40. We observe that the ease of publication depends on the agency's mission and on whether it sees data publishing as being an important part of its role. Where publishing data is part of the agency's job and where datasets are collected with the intention of being made public, it is easier to get quality public data out the door. Retro-fitting data publication on to datasets that were not collected with publication in mind can be much harder.

41. People we spoke to discussed substantial barriers to boosting the volume of data publishing in three areas (none of which is directly about process):

Data quality

- There can be substantial effort required to clean up data to make it usable.

---

[1] It is outside of our focus here on public data, but there is also guidance from the Privacy Commissioner relevant to sharing personal information between agencies.

- What works for one organisation collecting data does not necessarily work when that organisation's data needs to be combined with another's. Data is often collected at multiple points for administrative purposes with no concerns about cooperation or use by others or standardisation. The value of data often emerges from integration, but integration can be hard.

- And what works for administrative purposes does not necessarily work for analysis. As one example, we were told that an MBIE immigration dataset derived from administrative data has many different variants of "Wellington" as an address. For humans dealing with a single administrative record, this is no problem, since it is easy to tell what was intended. But to try to determine how many applicants have an address in Wellington, effort is required so the computer knows that the different versions of "Wellington" are comparable.

- Cleaning up data to put it into forms that can be published is hard, and the costs of doing this bring few benefits to the agency itself. It can take a long time to build systems that make it possible to publish usable data in appropriate formats, especially if that effort needs to happen at the point of collection, e.g., by validating addresses on entry.

Ownership and mandate

- Despite the Declaration and the substantial effort that has been put in over many years, for agencies other than those whose purpose is to release data, open data is still generally a sideline compliance issue that competes for time and energy with all the other compliance obligations under which Departments labour. With some exceptions, open data publication is not seen as helping agencies to do their job better but instead as a separate other thing that they do instead of their main job.

- It is evident that the best way to release data openly is to plan for that to happen as part of the process of collecting it in the first place. Where that has not happened (which is most of the time), a second-best world involves a long programme of changing existing processes and agreements to enable open publication of data that is collected.

- We observe that the decentralised open data model also makes progress difficult, because it means there are a lot of people to motivate and data publishing is just one task among many for people who may be experts on a dataset but are not experts on data publication, on what users of the data might want, or on how to make the data usable for others. Those agencies making the most progress have empowered believers in open data.

Usefulness of datasets

- Administrative datasets do not always include the data that users might prefer. Sticking with immigration examples, one policy question is where have immigrants settled after they have been granted residence. This data is not held by MBIE (the best source is tax data held by IR). It turns out that answering that question is much harder than one might think it should be for MBIE.

- These issues underline the importance of making dataset integration easy, so that answers can be found to questions that can't be predicted or that are not relevant to the agency that collects the data. They also make more

obvious the important of being open about exactly what data is in and not in particular datasets so that users know what to expect.

Potential solutions

42. As noted above, we cannot conclude that lack of guidance on processes for agencies is a major issue. But in pursuit of encouraging agencies to publish more data generally, we sketch out four potential directions for solutions here. We broadly assess their costs and benefits in the next section.

- 1) To boost the volume of open data published by particular agencies, a project could work with a small number of large agencies on an in-depth assessment of the state of their processes and progress in publishing open data. We would suggest keeping the number of agencies small to keep the project tractable. From this would come a much more detailed assessment of the barriers to progress and a nuanced view of what might be the most helpful assistance in particular cases.

- 2) To boost the amount of data that is published in general, clusters of agencies could be formed comprising some that are more advanced in their domain with others that are less advanced. Giving each cluster a shared set of goals to advance the data publicly available in a domain (rather than just in the holdings of a particular agency) might prove more fruitful than the agency by agency approach. Providing resources, education, and people experienced with managing data sharing projects could help lift achievement across the board. Creating guidance on the practicalities of data publication would also be useful. A clustering approach would support efforts that are already underway in agencies to form clusters with others in their sector.

- 3) To lift visibility of what data is held and what state it is in, agencies could be encouraged and supported to develop documentation like a fixed-asset register but for core data holdings, showing their extent and their quality. A shared view of the metadata would be necessary so that each agency approaches this task in a similar way. A more focused effort could focus on a subset of datasets that are identified as the most useful to others. This would support efforts that are already underway in agencies to draw up data inventories.

- 4) To make a sharper change in the volume and quality of data published, a single specialist organisation could be made responsible for data publication. It would take the data in whatever state it is in from agencies internally, and work through the process of cleaning and publication of usable data. A simpler version of this approach would be to kick-start some catalyst projects that involve small groups of people working with particular datasets to quickly describe their characteristics and diagnose their suitability for publication. This might be most useful for agencies where resourcing for open data is a most significant issue.

Costs and benefits

43. Exploring processes and barriers to publication with a small number of agencies is relatively cheap. It could readily build on existing guidance and be aimed at supporting efforts within agencies where those already exist. It could act as an in-

depth diagnostic, starting from the recognition that the outcomes New Zealand is getting from open government data effort are only scratching the surface of what is possible, but sticking with the agency-by-agency approach that characterises the existing system. It might lead to some changes in the existing programme, e.g., removing the "high-value" test and just requiring agencies make open all data that they hold, subject to confidentiality and privacy requirements.

44. Option two focuses on boosting agency capability and collaboration, and would bring some more structure to open data by defining particular data domains and encouraging clusters of agencies to progress the data available publicly across the domain. It could be complex in reality to form agency clusters and it is hard to know how this initiative could be sustainable without Ministerial direction or agency internal motivation. There does seem to be value in focusing on particular datasets and domains rather than a generalised obligation. We can see from the Open Data Barometer that New Zealand is behind the leaders in the open data world partly because it does not publish specific datasets in the appropriate way (rather than because it is behind on publication in general).

45. Option three is about improving agency data governance by lifting visibility of what data is held by agencies. It would build on efforts that already happen in some agencies, and so it need not be high cost. It could focus on specific agencies and data domains to simplify the effort and not waste effort with agencies that do not have much data of interest or lack motivation to do anything. There are more formal options (requiring annual audit and publication) or more informal approaches (with simple web publication). Aside from greater clarity for users on what data is collected (and therefore potentially releasable in time) it might help catalyse effort to make some particular datasets open. Simple, low-tech methods could increase awareness of what datasets are available, e.g., a daily email list or an online forum that highlights a different public dataset each day, or simple guided tours of what data is available from government agencies in particular domains.

46. Option four involves giving one organisation the responsibility for processing and publishing open data (as well as each individual agency continuing to publish data as some do now). This type of approach could sidestep problems of agency motivation and resourcing directly, assuming it had an appropriate mandate and was appropriately funded. There are similarities to the Danish Basic Data initiative, which we discuss in more detail below.

A way forward

47. The best option seems like some variant of Option three, i.e., encouraging agencies to manage their data like the asset that it is, and collating and presenting information on its characteristics and quality. This would be a long-term intervention, i.e., it would not lead directly to an increase in data sharing. There are existing efforts in agencies to encourage and build from.

Data sharing standards

48. The second barrier we focus on is about standards for data definitions.

49. Data is often published in the form of aggregated statistics about groups of people, households, or businesses with similar characteristics. Classifications used to define such groups include:

- Sector, industry, or business activity

- Geographic location

- Personal attributes such as ethnicity, gender, age, marital status, etc.

50. Such classifications are the key to combining, comparing, and aggregating data from different datasets for analysis. Problems arise when different datasets use different classification systems. At best, additional assumptions are needed to combine datasets that use different classifications, which makes analysis and interpretation more difficult. At worst, some types of analysis may not be feasible if classifications are significantly different.

51. Different classification systems have arisen because organisations produce datasets that suit their own systems and purposes, often without considering how their data could be combined with other sources and the value of such combinations. Unless organisations directly benefit from such combined data, there is little incentive to change existing systems.

52. One way to resolve this issue is the development and adoption of standards for classifying shared data. We discuss this in what follows using the example of geographic classification.

53. An alternative solution is to agree and publish rules for translating data from one set of definitions into another (i.e. some kind of official concordance). This may be easier to implement but does not address the fundamental issue and does not eliminate the assumptions that are required to analyse combined datasets.

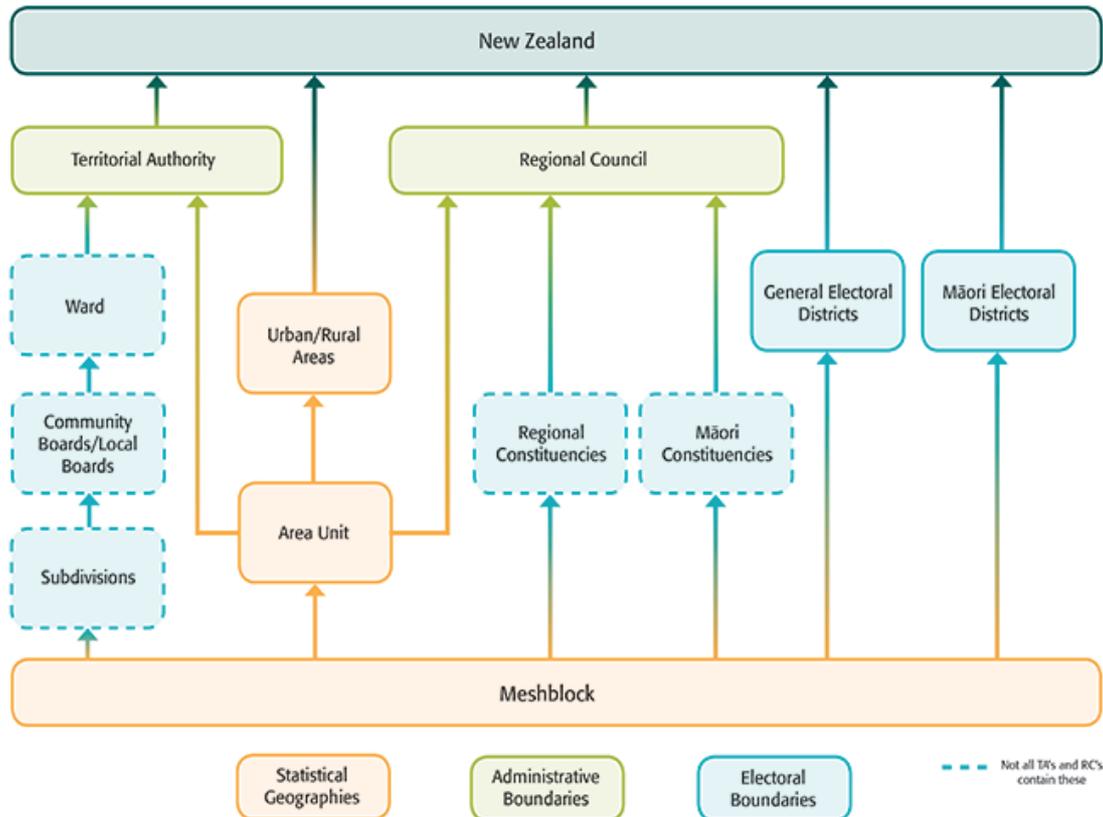**Our understanding of the issue**

54. Aggregated statistics about geographic areas are published by several New Zealand government agencies. Analysis of any area or across multiple areas often requires combining data from different agencies.

55. As demonstrated below, different geographic standards are in use, which creates costs for anyone wishing to combine geographic data from different sources. These costs are borne by both the public and private sector and arise from additional work required to combine data, as well as the opportunity costs of analysis that is not possible or hindered due to lack of standardisation.

56. Some datasets conform to Statistics NZ's geographic hierarchy (Figure 1). Meshblocks are the foundation of this hierarchy and all larger areas are made up of meshblocks. In principle, this means that meshblocks can be used to translate data between different types of geographic area. However:

- Data that does use Statistics NZ's hierarchy is often not published at the meshblock level, and

- Many public datasets do not use areas defined in this hierarchy at all.

57. As an example of the first issue, a general electoral district can include part of several Territorial Authority (TA) areas (and vice versa). If analysis is being carried out for TA areas, but some relevant data has been published for electoral districts, it is not possible to use that data in the analysis without making some additional assumptions to allocate it to the TAs. Such assumptions often cannot be tested for validity.

58. The second problem arises where other organisations have defined their own geographic areas. In addition to Statistics NZ's hierarchy, in the private sector there are:[2]

- A set of "localities" that identify communities of interest, such as suburbs, towns, and other areas in common use. This dataset is administered by the New Zealand Fire Service in consultation with NZ Post and Quotable Value. The localities dataset is available for free but only under a licence that severely limits its use.[3]

- Postcode boundaries, defined by NZ Post. This data is available only under commercial terms from NZ Post.

- Fisheries quota management boundaries defined by the Ministry of Primary Industries.

- Police, fire, ambulance regions and districts, fire and ambulance response boundaries, and fire jurisdictions.

- Environmental management and monitoring boundaries.

- District Health Board areas.

- Work & Income reporting regions, from the Ministry of Social Development.

- Regional Tourism Organisation areas.

---

[2] This list was partly derived from http://www.linz.govt.nz/about-linz/our-location-strategy/geospatial-data-themes-and-datasets/administrative-boundaries-theme

[3] More information about the localities dataset is available at http://www.linz.govt.nz/about-linz/our-location-strategy/geospatial-data-themes-and-datasets/administrative-boundaries-theme/nz-localities-suburbs-dataset. For a discussion of issues associated with licensing the localities dataset, see http://polemic.nz/2015/09/10/suburbs-and-open-data-barriers/.

Figure 1 Statistics NZ's geographic hierarchy

Source: http://www.stats.govt.nz/methods/classifications-and-standards/geographic-hierarchy.aspx

59.   A recent report by Deloitte and NZIER illustrated (Figure 2) how data about a single area of interest such as Opunake or Bluff can appear in different geographic areas in different datasets. The report describes how these inconsistencies create problems for using data to make decisions:

Take the challenge of understanding the rate of assaults and associated hospitalisation rates in Opunake. Data about hospitalisation is collected by the Taranaki District Health Board, while data about assaults is collected by the Central Police District, which encompasses a very different area – most of the lower North Island. Decision-makers must use statistical estimations to underpin a case for investment, which relies on many assumptions that do not always accurately reflect the real world.

The decision on whether to invest in interventions to address assaults and, if so, how much, relies heavily on these workarounds. Furthermore, the benchmarks for success or failure of programmes seeking to make a difference to assault rates, would need to rely on the same data and assumptions, which creates further problems for measuring performance.

60.   This example involves just one geographic location. Analysis that requires assembling a dataset on multiple areas from different sources can be even more problematic. For example, a large private operator of medical clinics may wish to analyse population demographics and health-related statistics, which requires combining Census data with data from DHBs that use different area definitions.

13

This lack of alignment makes analysis more difficult and costly, and potentially reduces its accuracy.

Figure 2 Example of conflicting geographic definitions
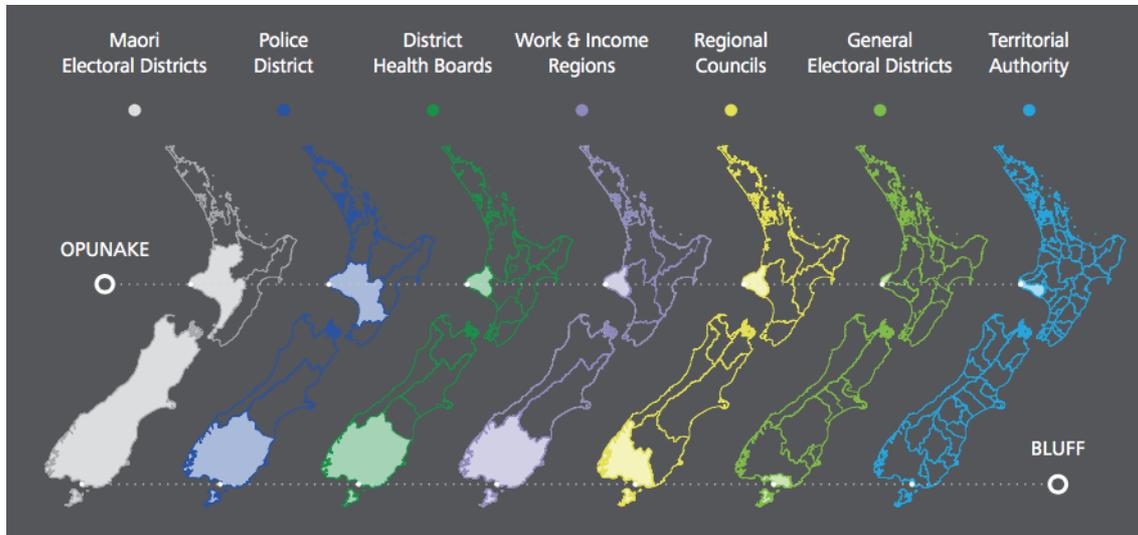
Source: *State of the State, New Zealand 2016: Social investment for our future*, Deloitte and NZIER.

Potential solutions

61.   We have considered two potential solutions to this problem:

- 1) Create and use a "register based" system for all primary data with geographic characteristics, from which aggregated statistics for any area can be calculated. A "register based" system involves recording data on people or businesses together with a pin-point location for each observation (e.g. a person's home address). Statistics can then be calculated for any area by aggregating over the individuals deemed to be in that area.

- 2) Encourage or require data published by government agencies that has a geographic dimension to comply with Statistics NZ's geographic hierarchy. Data would ideally be published at the meshblock or area unit level, from which it can be aggregated to larger areas if required.[4]

62.   The report by Deloitte and NZIER discusses the merits of a "register based" system and how it could solve the types of problems illustrated above. However, the authors note that implementing such a system across government departments requires significant effort and could be very costly. Statistics NZ

---

[4] Statistics NZ is undertaking a review of its statistical standard for geographic areas (see http://www.stats.govt.nz/methods/classifications-and-standards/current-classifications-and-standards-review/stat-std-geo-areas-2015-17-review.aspx). Area units are proposed to be replaced with a statistical area 2 (SA2) geography. An additional small area geography (SA1) may be introduced that sits between meshblocks and SA2 in the new hierarchy.

estimates that such a system would take 10-20 years to implement.[5] Thus, while we think this idea is worthy of further investigation, it does not seem like a near-term project.

63.   The second solution seems more feasible in the short term. Alternative ways to set data standards are discussed by the Australian Productivity Commission in a recent draft report:[6]

- A "leader-follower" approach, where a standard is set by an agency that publishes a lot of data using it, or where the standard is used for key datasets. Other publishers may then naturally find it beneficial to adopt the same standard. Standards have network effects – a standard is more valuable the more widely it is used. The "leader-follower" approach exploits such effects.

- A "community" based approach, where different agencies and the private sector work together to define and adopt standards. Under this model a central agency may coordinate but does not set the standard directly.

64.   The Australian Productivity Commission report notes that centralised standard setting can be slow and inflexible, which may delay benefits from the publication and use of data. The APC appears to be particularly concerned about inefficiencies from centralised setting of data standards that are primarily used by the private sector. In such cases, it recommends relying on voluntary efforts of the private sector first, with a centralised process as a backup.

Benefits and costs

65.   The benefits of solution two include:

- Reduced costs for users who need to combine geographic data from multiple sources. With standard definitions, little or no work is required to combine data. The value of this benefit is difficult to estimate but use of spatial data was estimated to generate $1.2 billion in productivity-related benefits for New Zealand in 2008, while an additional $481 billion in benefits was unrealised due to various barriers to the use of such data.[7]

- Better decision-making and better allocation of public resources, if the types of problems described in the Deloitte and NZIER report above can be reduced. Deloitte and NZIER describe many potential benefits of a "social investment" approach to government intervention based on data analysis.

- Reduced costs for agencies to maintain and publish their own geographic area definitions.

66.   The costs include:

---

[5] *An overview of progress on the potential use of administrative data for census information in New Zealand: Census Transformation programme.* Statistics NZ, 2014.

[6] *Data Availability and Use*, Australian Productivity Commission Draft Report, October 2016. Section 6.3 discusses data standards.

[7] *Spatial information in the New Zealand economy: Realising productivity gains*. ACIL Tasman, August 2009.

- One-off costs for agencies to adopt a new system of geographic classification. This may require changes to the way data is collected and recorded.

- Additional costs if data needs to be reported for old geographic areas as well as standardised areas.

- Costs of retrospective re-classification of data that has already been published.[8]

67. There is also the issue of ensuring adoption of standards by data publishers. This could be achieved by:

- Expressly requiring government agencies to publish their data using area units or meshblocks as a default. In addition to this, data could be published for other areas if there are clear benefits from doing so.

- Making it easier for agencies to publish data using area units or meshblocks, by creating guidelines and software tools for coding data to these standards and publishing it in appropriate formats. Privacy and confidentiality issues will need to be considered.

- Encouraging adoption of common standards by fostering communities of data publishers and data users. If agencies are more aware of the potential uses of their data and the barriers created by inconsistent definitions, they may voluntarily decide to change the way they publish data.

- Demonstrate the benefits of adopting common standards and encourage such adoption among state-owned enterprises (e.g. NZ Post) and the private sector.

A way forward

68. An approach to improving geographic classification could involve:

- Working with government agencies and private sector operators to encourage adoption of existing geographic standards defined in Statistics NZ's hierarchy.

- Investigating in more detail the barriers that organisations face to using these geographic standards and the costs of adopting them.

- Working with Statistics NZ to develop simple guidelines and tools to make it easier for organisations to adopt existing geographic standards.

69. More generally on standards there could be value in:

- Public articulation of the benefits of data standards and pushing for their broader adoption by government agencies and the private sector.

- Investigation of other areas where consistent classifications are not applied, such as for industries and personal characteristics (e.g., ethnicity, gender, nationality).

---

[8] Re-classification may not be possible if original data has not been retained in a format that allows it to be re-expressed for different areas. However, there are still advantages of publishing standardised data from now on.

- Helping to catalyse existing efforts to standardise data, e.g. we are aware of efforts at Auckland Council to standardise data about physical infrastructure owned by local government bodies.

- Developing good-practice guidelines for the creation of shared public/private sector data standards, and robust but flexible processes for maintaining them.

Government agency business models

70. The third barrier that we look at is government agency business models that are built on restricting access to data.

71. Some government agencies that have control of important datasets have incentives to earn revenue from the data and to restrict access to it. This is particularly problematic where the agency that creates or maintains an important dataset also sells services that make use of that data. In such a situation, it will have an incentive to restrict access to data to its competitors.

72. There are several such examples, each of which deserves further investigation of the opportunity costs created by restrictive access to data. The main economic question is whether other arrangements could cover the costs of data collection while permitting more use of the data and generating greater value overall. Here we discuss one such example: data on the residential property market. As we explain below, the restrictions on use of property market data create costs for home buyers and private sector realtors and property developers.

73. As an aside, there are parallels between restrictive government agency business models and public sharing of data by private sector entities. In our experience, barriers to private to public data sharing include the fear of a loss of commercial advantage, as well as customer backlash and legal risks. We also spoke to people who have organised data sharing schemes in the private sector. Such efforts face the challenge of balancing the benefits and costs of businesses to contribute data to a common pool, i.e. giving up confidential information to realise benefits from shared data. We were told that the most successful such efforts are organised around a specific purpose or to solve a specific problem faced by an industry. The motivation of businesses to join general data sharing arrangements without a specific purpose appears to be significantly lower.

Our understanding of the issue

74. The state-owned enterprise Quotable Value Limited (QV) and its commercial partner CoreLogic have assembled a substantial national dataset of residential property information and transactions in New Zealand such as actual selling prices, rating valuations, and the characteristics of buildings (e.g. floor area and number of bedrooms). QV and CoreLogic also sell services that make use of this data, such as property valuation services, rating valuation services for local councils, and analysis of property market trends.

75. We understand that residential property transaction data is originally collected by territorial authorities (TAs) who obtain data about transactions from conveyancing lawyers. TAs are permitted to sell this data under the Rating Valuations Act and we understand that QV/CoreLogic has agreements with all TAs to obtain property

market data from them, which it then processes and combines into a national dataset. The commercial relationships between the TAs and QV/CoreLogic are not transparent.

76.   Data including historic selling prices, the rating valuation, and characteristics of an *individual* property is readily available. QV sells such information on its website.[9] Other websites including homes.co.nz and TradeMe Property make some data on prices, valuations, and characteristics available for free, for one property at a time. Auckland Council provides a website where the rating valuation of any individual property can be looked up for free, but it has refused to provide open access to bulk ratings data and it prohibits users from scraping large amounts of data from its website.[10]

77.   The real value of property data comes from being able to analyse patterns and trends in prices, e.g. to show how these are changing in different areas over time, or to develop valuation models that can estimate the value of a property depending on its characteristics and location. Such analysis requires data on the prices and other attributes of many properties in different areas and at different points in time, i.e. a comprehensive nationwide property dataset.

78.   A limited nationwide property *titles* dataset is freely available from Land Information New Zealand (LINZ), but this only contains information about the geographic boundaries of each property and a list of its current and former owners. LINZ also publishes a comprehensive dataset of street addresses. However, the LINZ property datasets do not include information about the characteristics of buildings, actual selling prices of properties, and other attributes. It is not possible to analyse drivers of or trends in property prices using LINZ data alone.

79.   A nationwide property dataset that includes rating valuations and/or market prices is the second most requested dataset on data.govt.nz, where it was first requested in 2011. The requestor noted that release of such data would allow companies to "build innovative services utilising this data, or interesting views on this data – especially given that it is relevant to a large proportion of home-owning New Zealanders". QV responded that such data is "not suitable for release" due to existing commercial arrangements.[11] QV responded to a similar request on data.govt.nz for bulk property sale price data by saying that "The data you have requested belongs to the Territorial Authorities that we contract with. Under the contracts QV has with Territorial Authorities we have no rights to supply the data to any third party either free or for financial consideration. We can only use sales data to complete rating valuations on their behalf."[12]

80.   This situation therefore appears to have arisen from a combination of:

---

[9] https://www.qv.co.nz/valuations-and-reports/pricing-and-property-packs. A "comprehensive report" for a single residential property currently costs $79.95.

[10] https://fyi.org.nz/request/bulk_machine_readable_property_a and https://fyi.org.nz/request/2169-ytd-2014-total-revenue-generated-from-rating-valuation-data-sales.

[11] https://data.govt.nz/datasetrequest/show/3

[12] https://data.govt.nz/datasetrequest/show/203

- TAs being the primary source of property market transaction data and having the ability to sell this data under the Rating Valuations Act;

- The investment that QV/CoreLogic has made in obtaining property market data from multiple TAs and combining it into a single dataset, and the commercial agreements that it has reached with TAs to achieve that; and

- The fact that QV/CoreLogic also competes with other actual and potential users of property market data and does not have an incentive to negotiate arrangements that would enable it to make property data available to its competitors.

Potential solutions

81.   It may be difficult to adjust the existing commercial arrangements between QV/CoreLogic and TAs or to compel QV (as an SOE) to make its property dataset available on reasonable terms. The better solution may be to change the incentives faced by TAs with respect to property data and/or to make the data available via other channels. This could include:

- 1) Removing the ability of TAs to sell property market data, and requiring them to make it available to anyone who asks for it.

- 2) Requiring TAs to also provide property market data that they collect to another central government agency such as LINZ or Statistics NZ for free or on a cost-recovery basis, and allow this data to be made public.

- 3) Requiring conveyancing lawyers to provide data about property market transactions to an agency (such as LINZ), in addition to the information that is already provided to TAs, and allow this data to be made public.

82.   To properly resolve the problems described above, either option 2) or 3) needs to include a requirement for the agency (in our example, LINZ) to publish unit record data, i.e. prices and other characteristics for individual properties. Given that anyone can already purchase extensive data for any residential property from QV, there should not be any privacy or confidentiality issues that prevent the publication of unit record data.[13]

Benefits and costs

83.   In the time available for this project we have not done a full analysis of the benefits and costs of the three options above relative to the status quo. The following points are relevant for such a cost-benefit analysis:

- The benefits of making property market data freely available could include:
  o   Reduced barriers to entry and increased competition in services that make use of this data, such as property valuation and analysis of property market trends.

  o   Innovation and development of new products and services that make use of this data. This could include generating advice and reports on

---

[13] We understand that the Statistics Act prevents Statistics NZ from publishing data about identifiable individuals, which would prevent Statistics NZ from filling this role.

property trends using automated algorithms. People will be able to make better decisions when buying and selling property.

- o Improved efficiency of the property market if better information is available to property buyers, sellers, developers, urban planners, and policymakers.

- The potential benefits are significant. Reserve Bank figures show that New Zealand households own housing and land worth $720 billion as at June 2016,[14] and the Real Estate Institute of New Zealand reports total residential property sales of $56 billion in calendar 2016.[15] For most people, buying and selling property is a significant financial event that may occur only a few times in their lifetime.

- Under option 1), anyone who wants to analyse a national dataset must obtain data from each of the 67 TAs and will incur costs to combine and standardise the data provided by them. Thus, while this option potentially makes property data more widely available than the status quo, it will create some duplication of costs, and may lead to little additional use of the data if few users are willing to pay these costs.

- Under option 2), the costs of combining and standardising data remain but would be incurred once only, by the central agency tasked with collecting and publishing the data. The relevant agency would obviously need funding to cover these costs, if the data is made freely available.

- Option 3) could potentially avoid most or all costs of combining data from TAs if data from conveyancing lawyers is obtained in a standardised format directly. However, there would be costs associated with setting up and maintaining a data collection system.

- Under all three options, TAs would lose any existing revenues from selling property market data. Assuming TAs operate under constrained budgets, these revenues will have to be replaced with other sources. However, these revenues do not appear to be very large – in 2014 Auckland Council reported that it earned $423,000 from three sales of rating valuation data.[16] We do not know how much revenue was earned by other TAs, but scaling up Auckland's revenue by 2014 population, TAs earn annual total revenue from selling rating valuation data of at most $1.25 million.

- It is likely that wider availability of property market data will lead to increased competition for QV and CoreLogic. This may lead to a loss of profits for QV/CoreLogic. Such lost profits are not usually included as costs in a cost-benefit analysis if they result from increased competition or innovation (since they are a transfer from QC/CoreLogic to other parties, including consumers). The extent to which the solutions proposed above would affect QV's profits are unclear as QV offers a wide range of services

---

[14] http://rbnz.govt.nz/statistics/c21

[15]

https://www.reinz.co.nz/Media/Default/Statistic%20Documents/2017/Residential/December%202016/Market%20Facts/R100%20-%20New%20Zealand%20-%20December%202016.pdf
[16] https://fyi.org.nz/request/2169-ytd-2014-total-revenue-generated-from-rating-valuation-data-sales

and may have other sources of competitive advantage aside from its contracts with TAs. On average over the past five financial years, QV reported net profit after tax of $5.8 million per year.[17]

- We note that the combined profits of QV and estimated revenue earned by TAs from rating valuation data are around 0.01% of the total annual value of residential property sales. Open property market data only needs to enable a very small improvement in the efficiency of the residential property market to outweigh any losses experienced by QV and the TAs.

A way forward

84. The solutions proposed above would likely require legislative change. That is a slow process and will require an extensive case for change. One way to start might be a project that explores this issue more fully, discussing the problem definition and potential solutions with relevant organisations, and undertaking additional work on the benefits and costs. Greater transparency about the commercial arrangements between QV/CoreLogic and the TAs, and the implications of these agreements, would also be useful.

85. More generally on the subject of business models for government agencies, work could be done to raise awareness of the barriers to the use of data that can be created by public sector business models based on selling data.[18] One other case we are aware of is NIWA's business model based on selling weather and climate data that appears to inhibit competition in related services such as weather forecasting.[19]

An alternative strategy

86. In this last section, we discuss a different strategic approach to open data, which might be a way to address many of the barriers that hold back progress on open data at the same time for the most valuable datasets.

Our understanding of the issue

87. New Zealand's approach to open data relies on compliance by government agencies with their obligations under the Declaration on Open and Transparent Government. Open data efforts are not specifically funded. As noted above, agencies see their work as one more compliance exercise among many. And the obligations extend only to a small number of government agencies.

88. For areas where agencies have incentives to share data to improve their own operational performance, there has been substantial progress. Data sharing

---

[17] https://qvgroup.qv.co.nz/annual-reports. This average is slightly distorted upwards by a $13.5 million capital gain on sale of investments reported in 2016. Profits between 2012 and 2015 ranged from to $0.3 million to $5.2 million.

[18] There is extensive experience with this issue in Europe as well, as we explain in the literature review we report on in our first report in this project.

[19] MBIE is currently investigating open access to weather data; the terms of reference for its study are at: http://www.mbie.govt.nz/about/whats-happening/news/2016/open-access-to-weather-data-report.

between social sector agencies to support the investment approach is work of this nature. Agencies whose job is data publication also do a better job.

89. For other areas, progress in data sharing is much more limited. There are thousands of datasets available on government agency websites that are not published in compliance with the Declaration, including not being listed on the register at https://data.govt.nz. Most agencies publish no data. There are thousands more datasets held by agencies that are not released. The distributed approach to data sharing means that costs are low, but it also limits progress. As we know from our previous work, the ability to integrate data across datasets is central to generating value. Data release also needs to be authoritative and regular: you can't build a business based on OIA requests.

90. The basic argument behind this solution is that adding additional resources to the existing decentralised system is not the best approach: a centralised solution for specific datasets would be better.

Potential solutions

91. A different approach would involve a single organisation assembling and publishing high-value data for government agencies and the public to use, within a framework of appropriate safeguards. This could operate in addition to the existing agency-by-agency approach.

92. We can see this sort of approach in practice in the Danish government's Basic Data initiative.[20] The Danish government assembles and releases authoritative Registers of businesses, land and residents' details for government agencies and others to use. The data is free and made available from a single system (the Data Distributor). The data involved is listed in Table 1.

Table 1: Data released to be released as Basic Data (as at 2012)

Register	Contents
Central Business Register	Information on Danish businesses, including registration number, legal form, date of start-up, legal name and address, owners, sector code, number of full-time employees, and credit status
Company Register	Information on all registered companies. Some overlap with Central Business Register plus information on the management and auditors
Cadastral Register and Cadastral Map	Information about all land parcels in Denmark. Each parcel is identified by a cadastral number and its location is identified on the Cadastral Map. There are plans to expand it to include other real property including individually-owned apartments, and buildings on leased land.
Building and Dwelling	Information about all buildings and dwellings in Denmark,

---

[20] Thanks to the person who gave feedback on the DFP's Diagnose and Fix workstream in response to its online invitation and pointed us in the direction of this work.

Register	including age, usage, area, number of floors, heating, technical installations, fresh water supply and drainage, and annual rent, plus the complete register of all addresses in Denmark, including geographic positions
Register of Property Owners	Information on the actual owners of all real property in Denmark
Map data	Descriptions of landscape forms and special characteristics like towns, roads, cliffs, tracks, watercourses, and wind turbines
Administrative and geographical boundaries	Details of the geographical demarcation of municipalities, regions, parishes, judicial districts, postal codes and more
Danish Elevation Model	A digital elevation model of the terrain and its elevation above sea level
Place Name and Information Register	Lists approximately 200,000 place names
Civil Registration system	Basic data on individuals, including civil registry number, name, address, date of birth, marital status, religion, kinship and nationality, both current and historic

93.   Basic Data is now a very substantial programme involving a large number of government agencies. The initiative is run by the Ministry of Finance because that is also the home of the government's Agency for Digitisation.

94.   There are initiatives to provide authoritative Register data in other countries too. Just focusing on address datasets, for example:

- The UK government announced funding of GBP 5m in the 2016 Budget to "develop options for an authoritative address register that is open and freely available" to "unlock opportunities for innovation".[21]

- There was also a previous initiative funded by the UK government to look at options. Open Addresses was created by the Open Data Institute but eventually shuttered partly because of legal issues created by the sale of address database when the Royal Mail was privatised in 2013.[22]

- The French government has released a public national address database.[23] The Australian government released a national address file and boundaries datasets openly from February 2016.[24] Efforts in the United States seem to be rather further back.[25]

---

[21]   https://gds.blog.gov.uk/2016/03/23/an-open-address-register/

[22]   https://alpha.openaddressesuk.org/blog/2015/07/27/a-time-for-going-to-bed

[23]   http://www.modernisation.gouv.fr/salle-de-presse/communiques-de-presse/open-datathe-first-french-collaborative-national-address-database-now-online

[24]   https://blog.data.gov.au/news-media/blog/geocoded-national-address-data-be-made-openly-available

[25]   http://fedscoop.com/national-address-database

95. The MyInfo service in Singapore is a different approach again, just in relation to data about residents. It is a government service that lets users register for government services more easily and keep track of who has access to data about them.[26] The Singaporean model is user controlled, i.e., users register and can see the data that is held about them by different agencies. Users specifically authorise use of the data when registering for services that are enabled to work with MyInfo.

96. There is a wide range of data available in MyInfo including name, age, date of birth, address, contact details, nationality, income, employer, occupation, family status, car registration, and property ownership. Most data is sourced from (and verified by) public sector agencies. Users can readily see what this data is and engage with those agencies to ask questions or to update it. Other data is provided by users themselves.

Benefits and costs

97. One direct benefit of this effort would be cost reductions for those that are using this data today. They would save on the costs to find, secure and transform the data into something they can use for their own internal processes. Operational cost savings for agencies could be an important part of encouraging agency co-operation with the project.

98. By increasing publication of usable data, this initiative would generate the same benefits as publication of open data in general. We reviewed these in our first report. They include greater economic and social value derived from the use of data that can arise in many ways, including an increase in economic activity, better information for people that enables better decisions, and more effectiveness and efficiency in government services. Although uncertain in scale, the benefits are thought to be very large.

99. There would be some direct costs, although these would not be expected to be large, especially when compared with the benefits available. We can get some sense of the scale of benefits and costs from an economic assessment done by the Danish government of Basic Data.

100. The precursor to the Basic Data programme was the release of address data from the Building and Dwelling Register in 2002. An economic assessment in 2010 showed substantial economic gains from the data release. The total costs in the first five years were assessed at €2 million (to compensate the municipalities that collected the data and pay for the costs of data distribution) while the benefits added to €62m. The benefits counted were only the direct financial benefits to the 1,200 parties that accessed the data. Other uncounted benefits included an increase in the number of digital services and websites that enhanced transparency and competition in the Danish property market. The assessment foresaw substantial additional value available from application of the data in new areas, including navigation and mobile applications, logistics and transport, and utilities.

101. The Danish government has also assessed the costs and benefits of the wider Basic Data initiative. Agencies expect to secure significant fiscal savings from not

---

26   https://www.singpass.gov.sg/myinfo/intro

having to collect and store the data themselves, as well as benefits for service delivery from having better data and integrated data rather than just a partial view of a user. There are also broader policy benefits from a better understanding of the status and activities of the Danish population. For users, the programme is expected to improve the quality and efficiency of public sector services, as well reducing form-filling and making authentication easier. Wider economic benefits for firms and individuals from the use of the data are also expected to be very substantial.

102. Together the Danish government estimates societal benefits of about NZD 160m a year from the Basic Data programme once it is fully implemented in 2020. Private-sector revenues will be up about 70 per cent of this total. The real estate, insurance, finance, telecommunications and logistics sectors are expected to be among the most substantial beneficiaries.

A way forward

103. If something like this Basic Data initiative were of interest, a next step would be a more detailed investigation. An early decision to focus on a particular data domain (e.g., data about businesses and companies), to define the data in detail, and to form a government and private sector team to estimate the costs and benefits of open publication and work through the details could be helpful.

104. Learning from Basic Data, an effort could start with addresses (building on the addressing data already published by LINZ[27]) and then move to company data, and have a structure that enables more data to be added over time. It will work best if it delivers value immediately to those that collect the data, so that they will have the desire to share data with it.

105. There will doubtless be much discussion about where such an initiative should live. A central agency makes most sense to us because it is most influential and would have a funded mandate. As we understand the Statistics Act, it could not be led by Statistics NZ without a change in the law since released data cannot identify anyone (and a substantial benefit of the initiative is the provision of identified data to government agencies).

Recommendations

106. In this final section, we isolate some proposed actions for the DFP from our project.

107. Our thinking was guided by our understanding of the levers that are within the DFP's influence:

- Making observations and recommendations to Ministers on how to make faster progress with data sharing (including suggesting changes to the DFP's own mandate).

- Encouraging organisations (in any sector) to do a better job of data sharing, through conversation, analysis and education.

---

[27] Since December 2015, LINZ has published a comprehensive dataset of all "addressable locations" in New Zealand – see https://data.linz.govt.nz/set/87.

- Providing a place and a context for useful conversations about data sharing to happen.

- Engaging with the public directly to build awareness, understanding and support for data sharing done well.

- Encouraging data sharing activity directly by providing seed money to qualifying catalyst projects.

108. In our view the best place to start with boosting data sharing remains with data collected by government agencies. This is large and still relatively untapped opportunity. New Zealand has done an okay job on open government data so far, given the lack of dedicated resources, the lack of motivation for agencies to share data, and the use of a highly decentralised model. Much more is possible. Open government data done better has widespread acceptance as a concept, substantial economic and social benefits, and high-level government support.

109. The main change that we would recommend would be to exploring making one agency responsible for assembling, integrating and sharing a subset of defined datasets to all others and to the public, in a form similar to the Danish Basic Data initiative.

110. To get started and demonstrate usefulness, the initiative could begin with a small subset of very useful data that is already available but needs some tidying up, e.g., address and business data.

111. In time, data could cover the physical environment (e.g., topography, land cover, real property ownership), the built environment (e.g., buildings, addresses, place names, urban and suburb boundaries), businesses (e.g., company names, locations, owners and performance) and individuals (e.g., names, addresses, age).

112. There would be controls on some data to protect confidentiality and personal privacy. Non-confidential, non-personal data would be publicly available for free.

113. This initiative could sharply increase the availability of high-value data, reduce costs for government agencies in storing the same data more than once, simplify data analysis by integrating the datasets with beneficial effects across the economy, and resolve some of the challenges for data sharing created by government agency business models.

114. The next action for DFP would be to support a project to define in more detail what this Basic Data approach could look like, engaging with potential users to ensure that the product design fits with their needs. A project could be completed by October this year in time to include in the DFP's report on its progress so far to Ministers.

115. The DFP could also:

- Encourage government agencies to put together and publish a list of data assets they hold, what they contain, and what condition they are in, and support this encouragement with practical assistance and public praise for agencies that are doing this well.

- Engage with agencies that publish data using non-standard geographic definitions to understand their reasons for doing so, and support and encourage them to adopt consistent geographic standards (and standards for other characteristics too).

- Dig in more detail into the costs and benefits of opening up data on the residential property market, and developing a preferred solution for that work for Ministers to engage with.