

Scenario Planning Case Studies Using Open Government Data

Robert Power, Bella Robinson, Lachlan Rudd, Andrew Reeson

► **To cite this version:**

Robert Power, Bella Robinson, Lachlan Rudd, Andrew Reeson. Scenario Planning Case Studies Using Open Government Data. 11th International Symposium on Environmental Software Systems (ISESS), Mar 2015, Melbourne, Australia. pp.207-216, 10.1007/978-3-319-15994-2_20 . hal-01328550

HAL Id: hal-01328550

<https://hal.inria.fr/hal-01328550>

Submitted on 8 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Scenario Planning Case Studies using Open Government Data

Robert Power, Bella Robinson, Lachlan Rudd, Andrew Reeson

CSIRO, GPO Box 664 Canberra ACT, 2601, Australia

{robert.power,bella.robinson,lachlan.rudd,
andrew.reeson}@csiro.au

Abstract. The opportunity for improved decision making has been enhanced in recent years through the public availability of a wide variety of information. In Australia, government data is routinely made available and maintained in the <http://data.gov.au> repository. This is a single point of reference for data that can be reused for purposes beyond that originally considered by the data custodians. Similarly a wealth of citizen information is available from the Australian Bureau of Statistics. Combining this data allows informed decisions to be made through planning scenarios.

We present two case studies that demonstrate the utility of data integration and web mapping. As a simple proof of concept the user can explore different scenarios in each case study by indicating the relative weightings to be used for the decision making process. Both case studies are demonstrated as a publicly available interactive map-based website.

Keywords: urban planning · environmental planning · decision support · government data · public transport

1 Introduction

This paper addresses the issue of characterising the demand for community services, and presents a framework for evaluating existing service locations in terms of their accessibility by public transport and ability to meet demand. To demonstrate the approach, we fit simplistic models to show the types of policy enhancing results that are possible. The main objective is to encourage government to push more data into open access; then more complex models could be fit to the actual datasets. The simplistic models we use show the results that are possible and highlight the power of the mapping tool for policy makers to interpret large amounts of data.

This approach can be applied to a range of services. Here we demonstrate its applicability using two case studies drawing on publicly available datasets. The first focuses on government shopfronts (used in Australia for the delivery of welfare and other government services), and the second looks at environmental services in the form of urban parks.

The framework allows users (modelers, decision-makers or other interested parties) to explore and evaluate scenarios using an interactive map-based website. This exploration allows the user to define appropriate parameters to the decision making model. The model results are then computed in a database to produce the final results which are made available on the map-based website.

The demand for services varies across the community. This demand can be categorised based on the expected community need. Different categorisations define scenarios to help assess services which form the basis of the decision making model. An example of a service is a school, shopping centre, swimming pool, playground, government service centre or shopfront. The community need for these services varies. Some services will be needed by the general public, but will be used more by some than others; for example a shopping centre or swimming pool. Other services will predominantly be used by specific community subgroups; for example schools are used by families with school aged children.

The aim of our investigation is to explore how publicly available datasets can be used to help assess services using a methodical, repeatable, transparent and evidence based process that makes use of an interactive map-based website to allow modelers to better explore the data and its interrelationships. The focus is on services provided at specific locations, and we consider that demand is determined by demographics, accessibility by public transport (since not all have access to private cars), and the availability of alternatives.

2 Related Work

Demand for services varies spatially, with population density, demographics and transport as key determinants of accessibility. Location-allocation models are used to optimise the locations of service centres. A specific review in the health sector, found that greater availability of geographic data (e.g. GIS) and enhanced computing power has greatly expanded the range of data applications [1]. More recent applications include using mathematical optimisation techniques to identify optimal locations for defibrillators in urban areas [2,3].

A decision support system was developed using a mixed integer programming model to determine the efficient location of government service delivery offices by identifying which offices could be closed with the least impact on customers [4]. Such models can be used to explore a number of different scenarios, and provides decision-makers with an objective method of making, and justifying, service delivery decisions. A review of the use of geo-information technologies to support collaborative and participatory decision-making found that this can help to engage a greater range of stakeholders and elicit more local knowledge [5].

The simplest approach to accessibility considers Euclidean distance from locations, while more detailed models include roads and other transport features [6]. Accessibility by private car can be estimated from traffic flow models, but public transport has proved more complex due to the vagaries of routes and timetables [1]. Spatial optimi-

sation has been applied in environmental management, including conservation [7] and natural resource management [8].

3 Methodology

The following process describes the steps taken to develop the interactive map-based website for use by a domain matter expert: the modeler. This process requires expertise in data management and website administration. While these skills are available in the Information Technology (IT) section of most organisations, they are also becoming a common skill set within reach of researchers. We believe the skills required to accomplish the tasks outlined below are increasingly within the capability of modelers themselves, especially as supporting software tools become available in the field of semantic web and linked data.

3.1 Problem Definition

The first step is to clearly define the problem to be solved. While this may seem obvious, it is sometimes assumed to be self-evident and therefore overlooked. By taking the time to define the problem to be solved using clear and unambiguous language, all participants in the task will share a common understanding of the goal to be achieved.

3.2 Obtain the Data

Central to the methodology is the need to understand the community for whom services are being provided (usually using census information) and the preferred means of accessing the services (in our case public transport). A summary of these two critical datasets are described in Section 4 below.

The other necessary dataset for the decision making process is the locations of the services being evaluated. This is expected to be available to the subject matter expert: the modeler. There may be other datasets necessary and again, it is the responsibility of the modeler to provide these and describe how they are to be used.

In some cases the required data will not be available and so proxies must be identified. This occurs when the necessary level of detail is not available and so approximations are required or where the specific data is not available, but known correlations can be found. An example is presented in the first case study where the home locations of customers for a government shopfront are not known, but are approximated using the Department of Social Security (DSS) payment recipients by postcode data from data.gov.au.

3.3 Load the Data

In order to load the data into a database, the data needs to be in a machine readable format. This process may require the data to be ‘cleaned’ (free of errors or inaccuracies), restructured (only relevant data extracted) and pre-processed (transformed to be

suitable for the database loading tools) before it can be loaded. In some cases the data may need to be modified to make it compatible with other datasets. For example, the identifiers used to reference common artifacts (such as postcodes, suburbs, street names and so on) may need to be modified to match other datasets. This is often necessary when combining data managed by different custodians.

3.4 Define the Decision Making Model Function

There are three elements to the definition of the decision making model, outlined below. Note that this is intentionally a simplistic approach that is quick to implement with publicly available datasets as a proxy for actual usage data. As actual data is released, more powerful models can be used.

1. Categorise the community to reflect different service demands.

For the example case studies below, we categorised the community into demand groups (e.g. levels of high, medium and low demand). The number of categories and how their membership is defined is dependent on the problem under investigation and at the discretion of the modeler. It is likely to include demographics (e.g. families with young children and families living in high density housing, are likely to have a higher demand for urban parks) and the availability of alternatives (e.g. households with broadband internet are more likely to access government services online, and so will have reduced demand for shopfronts).

2. Combine the different measures of service accessibility.

The service accessibility includes public transport travel time but can also include other measures (e.g. walking or driving distance).

3. Include parameters to account for relative competing service demands.

The contribution of each variable parameter to the overall result will change based on modeler supplied weights. This is how outcomes are prioritised. For example, decision-makers may assign a higher weighting to certain disadvantaged groups, or long travel times.

3.5 Update the Website

Geographic data can be explored visually in order to gain deeper insight into the available data. As well as making the input datasets available, the parameterised decision making model function is included as a derived data layer. This is how the modeler explores different scenarios – by providing different model parameters and viewing the results.

4 Gathering Data

4.1 The Australian Statistical Geography Standards (ASGS)

The Australian Bureau of Statistics (ABS) divides the country into a structure of hierarchical non-overlapping regions called the Australian Statistical Geography Standard (ASGS). These regions are used to present aggregate statistical information, such as population density, economic indicators, and so on. The main regions are termed Statistical Areas defined in four levels: SA1 (Statistical Area Level 1), SA2, SA3 and SA4. A collection of SA1 regions are grouped to form an SA2; the SA2s are grouped to form an SA3 and so on. There are 54,805 SA1s; 2,214 SA2s; 351 SA3s; 106 SA4s. SA1 data has been used in the case studies described below. These are the smallest regions for which the ABS releases census data and they vary in size to have a population of 200 to 800 people with an average of about 400 people.

4.2 Public Transport Travel Times

One of the measures of accessibility to a service is the travel distance to its location. A different accessibility measure is the travel time using public transport. Websites exist for people to plan travel on road networks, see for example Google Travel Planner¹. Similarly, public transport journey planners are available for all Australian capital cities and some regional areas. It would not be feasible to manually use these journey planner websites to calculate the travel time to the nearest service.

Software was used to perform this task. Public Application Programming Interfaces (APIs) are available that provide a software interface to public transport journey planners, such as Google's Directions API² and the SilverRail³ Journey Planner. However, these APIs have strict licensing conditions and/or only allow a small number of requests before a fee is required. For example, Google's Directions API only allows 2,500 requests per day and the results must be displayed on a public Google Map. An alternate open source solution was found along with timetable information, based on OpenTripPlanner and OpenStreetMap. By installing this software locally, we could make as many requests as necessary.

Software was written to systematically use OpenTripPlanner to generate travel plans for the 34,172 SA1s (62%) that were within the regions of Australia for which we had public transport data. For each SA1, travel plans were generated from the SA1 centroid to all service locations that were within a specific radius of the SA1. OpenTripPlanner can generate several alternate itineraries for each SA1 service location combination. Only the journey with the shortest duration was recorded in the database. Further details are provided in the case studies below.

¹ <http://maps.google.com.au/intl/en/landing/transit/>

² <https://developers.google.com/maps/documentation/directions/>

³ <http://www.silverrailtech.com/journeyplanner>

5 Government Office Locations Case Study

5.1 Description

Government services have traditionally been provided to citizens through face-to-face interactions at shopfronts or through telephone services provided by call centre staff. More recently, there has been a push to transition customers to self-management through digital services with the aim of reducing service delivery costs while maintaining customer satisfaction. Over time, face-to-face government services can be combined between departments and call centres consolidated providing efficiencies.

5.2 Process

For this investigation we worked with the Australian Government Department of Human Services (the department) to evaluate alternative options for service delivery to meet anticipated customer needs balanced against the department's resources. There are 304 existing government shopfront locations, referred to as customer service centres, or simply offices, available from data.gov.au. Another public dataset used was Detailed Payment Demographic Data which included payment recipient numbers by postcode for the March 2014 quarter⁴. There were 27 payment categories such as age pension, carer allowance, disability support, unemployment benefit and so on. This data was de-aggregated from postcodes to SA1 regions using ABS socio-economic indicators by matching the expected customer profile to the different benefit types. For example, areas of high socio-economic need are allocated more customers for specific benefit types, such as unemployment payments, compared to areas of lower socio-economic need [9].

In summary, the aim of the investigation was to evaluate the accessibility of existing offices in terms of public transport for the department's customers and to model the potential for customers to transition to self management through digital services.

The ability of citizens in the different geographic areas to transition to self management through digital services can be approximated by the broadband availability and usage in their local area. The idea being that digital services are best provided to areas that are currently well serviced and utilised by the community. This information was available from two sources: the ABS census data (which includes an indication of the Internet availability for dwellings for each SA1 region) and the Australian Government Department of Communications MyBroadband website⁵.

The demand for services was modeled by categorising the customers into three groups; high, medium and low, using the payment demographic data. Those customers with high need likely require regular and possibly intensive meetings with front line customer service staff. Those with low need are expected to understand their obligations and (mostly) meet these requirements using self service interactions, preferably online. Customers categorised as medium fall between these two and will have

⁴ <http://data.gov.au/dataset/dss-payment-demographic-data>

⁵ <https://www.mybroadband.communications.gov.au/>

some irregular face-to-face or phone contact with the government while also fulfilling some aspects of their interactions through self service mechanisms.

A demand (d) measure for each SA1 region is calculated by combining the customer category numbers (customers categorised as being high (h), medium (m) or low (l) need) with the services expected to support them. This is a weighted combination using parameters (w_h for high weightings, w_m medium and w_l for low) as follows:

$$d = w_h \cdot h \cdot t_i + w_m \cdot m + w_l \cdot l \cdot b_s$$

Here t_i is the travel interval in a range of 1-10 and b_s is the broadband scale. In both cases, 1 is good (short travel time or good broadband) and 10 is bad. The demand measures for the SA1 regions close to an office are then summed, normalised and scaled and the results ranked to allow comparisons between offices.

The rankings were used to divide the centres into deciles loosely based on customer need. If the need was low, policy makers could consider downsizing the centre and shifting more services online. Demand for a centre's services was defined as the sum of the SA1 rankings in the centre's catchment. This sum accounts for the volume of customers weighted by their service requirements (high, medium or low) and their access to public transport or broadband.

Once demand was calculated and centres were arranged into deciles, policy makers could look at the distribution of customers at any given service centre. Univariate statistics like the mean, median, skew and kurtosis were generated to help evaluate large groups of service centres within a decile and a histogram of customer need scores was generated for each individual centre.

Using the univariate statistics, policy makers are able to identify centres with concentrations of one particular customer type, or centres with skewed distributions indicating a wider mix of customer demand. The histograms of key centres can then be analysed prior to considering changes to the services offered by a centre. This allows the impact of change on the full spectrum of customer types to be considered.

The data is available to explore on the <http://rapt.csiro.au/rprc> website.

6 Urban Parks Accessibility and Demand Case Study

6.1 Description

The next case study examined the expected use of parks in the Brisbane metropolitan area. Urban parks offer a range of environmental services, which are likely to become increasingly important as housing density increases. The 2,042 park locations and descriptions of their facilities were obtained. The park facilities indicate, among other information, the presence of playgrounds, sporting areas (fitness exercise equipment, netball courts, soccer fields), recreation areas (bike paths, skating areas, bushwalking) and picnic facilities (barbeques and seating).

The aim of this investigation was to measure park accessibility and demand by using public transport. Some parks will be more attractive than others based on the facil-

ities they provide. This equates to an ‘attractiveness’ measure, defined by the number of facilities provided, ranging from 0 (no facilities), 1 (one of playgrounds, sporting areas, recreation areas and picnic facilities) up to 4 (includes all facilities).

6.2 Process

Two stages of data preparation were required: calculate the walking and public transport travel times to the various parks from the SA1 regions and identify members of the community most likely to want to use the different parks.

Parks with playgrounds were expected to be used by families with small children, recreation parks used by teenagers, parks with picnic areas by people who live in dwellings without a yard and so on. To reduce the calculations required, thresholds were placed on the distance people will travel to a park based on the attractiveness measure: people are expected to only walk up to two kilometers to a ‘0’ park (attractiveness measure of 0, i.e. a park with no facilities) and optionally use public transport to travel to the other parks. For a ‘1’ park, the maximum distance travelled is set to 3 kms, 5 kms to a ‘2’, 7 kms to a ‘3’ and 10 kms to a ‘4’. We are assuming that people will be willing to travel further on public transport to parks with more facilities (ie those with a higher attractiveness score). The travel planner software was configured to calculate public transport travel plans for each SA1 – park combination that meets these park attractiveness and distance constraints.

The model used is an arbitrary proxy to demonstrate the types of inferences that can be drawn from this type of data. More powerful models could be constructed using the local council’s actual usage statistics. Our model’s advantage is in using publicly available data and easy to implement methods for early stage exploratory analysis and to demonstrate what can be learned if more data is pushed into open access. Additionally the linking of public transport information and the readily interpreted map-based display of model outputs is of high utility to policy makers.

The data is available to explore on the <http://rapt.csiro.au/parks> website.

7 Discussion and Future Work

There have been a few lessons learned in undertaking these case studies. While there is an abundance of open government data published in Australia, the same data is available from different agencies, in different formats, with different structures and includes different content. The data.gov.au site is a central repository of authoritative government data, however state governments maintain the same data on their own data repositories or agency specific websites with content that is more up to date and more extensive. For example, the Brisbane park data was obtained from the Brisbane City Council website⁶ not data.gov.au.

Finding data relevant for a particular investigation has become easier and governments are more willing to describe and share their data. However, improvements in

⁶ <http://data.brisbane.qld.gov.au/index.php/datasets/>

data related outcomes can still readily be achieved if governments are able to interlink their releases with existing public and private datasets. Interactive map-based websites are becoming ubiquitous and user expectations of the features available from them are increasing. This is in our view an opportunity and was the motivation for our case studies: combining readily available open government data for new purposes and making it available in a publicly accessible and interactive website. However to achieve these aims we have had to exploit specialist skills in data management, software engineering, web application development and statistics.

While the short term aim was to demonstrate the benefits of combining these skills to deliver the solutions presented, this process is not sustainable. Instead, tools need to be developed to allow users to readily incorporate web accessible datasets of their own choosing. This integration needs to be more than simply making datasets appear as a new ‘layer’ on the website. It needs to be combined in the context of the datasets already present. This can be achieved by utilising emerging standards and supporting tools in the field of the semantic web. This is an ongoing area of our research.

Another area of further work is to move away from a data warehouse solution and utilise a services oriented architecture. This requires the data custodians to introduce suitable services which adopt appropriate standards. While these services are not currently universally available, it is expected they will be established over time.

Other future work includes automating the process of exploring and evaluating the scenario generation process, mechanisms to combine private data using authentication controls and to extend the public transport analysis. For example, to preference trains over buses, use different thresholds for the maximum number of transfers, varying the maximum walking times for different customer cohorts and including more timetable options, such as weekends and different times of the day.

8 Conclusions

We have presented a methodology for characterising the demand for community services, with a focus on evaluating existing service locations in terms of their accessibility by public transport and ability to meet demand. The methodology has been demonstrated by two case studies using publicly available datasets. It provides a methodical, repeatable, transparent and evidence based process for users to perform scenario planning. The scenarios are explored by users providing parameters to the decision making model with the results available as an interactive map-based website.

The aim of the first case study was to demonstrate what could be done without actual access to the required data. We weren’t subject matter experts, did not have access to the detailed customer data, and did not try to build the most accurate model. Our objective was to demonstrate what could be achieved if government were to open access to more data. We showed possible results and how they could be used by decision makers to enhance policy outcomes.

If the data was to become open, our model for demand would change drastically (mixed integer programming, approximate Bayesian computation, hierarchical regression models with spatial correlations). We could tailor the model to the data itself. In

this paper we are using a balance between simplicity and general applicability to make a case for releasing more data. The second case study examined the expected use of parks in the Brisbane metropolitan area based on park facilities, demand and accessibility by using public transport.

These examples show that the decision making process is available to non-experts such as members of the general public who have an interest in exploring such issues in their community. In order to fully realise this aim, the user needs to be able to integrate their own data and define their own decision model function. The availability of relevant public datasets goes some way to achieve this aim. The next stage is to develop advanced tools based on semantic technologies and linked data to allow users to integrate content, visually explore data and to model relationships. The result would be a general purpose web accessible 'spatial spreadsheet', combining features found in web mapping, databases, Geographical Information Systems and spreadsheet tools.

Acknowledgements

This work was funded under the Human Services Delivery Research Alliance between the CSIRO and the Australian Government Department of Human Services. Thanks go to David Lovell (CSIRO) who provided helpful advice during the project.

References

1. Tanser, F., Gething, P., and Atkinson, P. (2009). Location-allocation Planning. In "A Companion to Health and Medical Geography", pp. 540-566. Wiley-Blackwell.
2. Tsai, Y. S., Ko, P. C. I., Huang, C. Y., and Wen, T. H. (2012). Optimizing locations for the installation of automated external defibrillators (AEDs) in urban public streets through the use of spatial and temporal weighting schemes. *Applied Geography* 35, 394-404.
3. Chan, T. C. Y., Li, H., Lebovic, G., Tang, S. K., Chan, J. Y. T., Cheng, H. C. K., Morrison, L. J., and Brooks, S. C. (2013). Identifying locations for public access defibrillators using mathematical optimization. *Circulation* 127, 1801.
4. Narasimhan, R., Talluri, S., Sarkis, J., and Ross, A. (2005). Efficient service location design in government services: A decision support system framework. *Journal of Operations Management* 23, 163-178.
5. McCall, M. K., and Dunn, C. E. (2012). Geo-information tools for participatory spatial planning: Fulfilling the criteria for 'good' governance? *Geoforum* 43, 81-94.
6. Noor, A. M., Amin, A. A., Gething, P. W., Atkinson, P. M., Hay, S. I., and Snow, R. W. (2006). Modelling distances travelled to government health services in Kenya. *Tropical Medicine & International Health* 11, 188-196.
7. Moilanen, A., K. A. Wilson, and H. P. Possingham. 2009. *Spatial conservation prioritization: quantitative methods and computational tools*. Oxford University Press, United Kingdom.
8. Bryan, B. A., Crossman, N. D., King, D., and Meyer, W. S. (2011). Landscape futures analysis: Assessing the impacts of environmental targets under alternative spatial policy options and future scenarios. *Environmental Modelling & Software* 26, 83-91.
9. Power, R., Robinson, B, Rudd, L (2014). *Planning the Future Face to Face Service Delivery Footprint*. CSIRO Australia.