

Data democratisation: empowering the environmental management community through data

Marsh, N¹, and Marsh, B²

1. Truii, PO Box 3134 West End, 4101. nick.marsh@truii.com

2. Truii, PO Box 3134 West End, 4101. Bonny.marsh@truii.com

Key Points

- Data driven organisations are more efficient and profitable.
- Data visualisations are more effective forms of communication than tables of numbers.
- Open data resources and online tools are opening up opportunities for natural resource management organisations to communicate through data visualisations.
- Data democratisation allows the broader natural resource community to contribute to natural system understanding through the contribution of data.

Abstract

We live in a world of ever increasing connectedness through the Internet Of Things (IOT). The IOT produces growing streams of data that are interconnected and cross-correlated to sell us stuff that we don't need. However, in the natural resource management sector we still operate on a more binary organisation by organisation and project by project piecemeal basis. We could embrace the data driven world by getting our data into the public domain and allowing data mashups, interesting visualisations and ultimately provide a more complete picture of all the aspects of the natural environment which we have been diligently measuring and modelling. In this paper we step through the value of a data driven approach to decision making, the increasing availability of open data and tools to interact with the data and create and publish data visualisations. We also present some recent examples of natural resource management organisations that have embraced data democratisation through novel online data visualisations.

Keywords

Open Data, Data democratisation, Data visualisation

Introduction

Data driven decision making is a relatively new phrase adopted by the broader business community, however it encapsulates the standard approach which has always been applied to natural resource management. The natural resource management industry tends to be science and data driven, we try and understand the natural system through data collection and modelling and then apply that knowledge to stretch the meagre resources for greatest effect.

In the past the data/knowledge custodians have been the natural resource management organisations. We argue that that it is possible to empower the broader natural resource community by putting the data in their hands. This approach allows them to explore the condition of the catchment, where the nutrient hotspots might be and where the opportunities for restoration are. Whilst it may be empowering to allow the broader community access to the underlying data, a further development in data democratisation is to allow the broader community to also contribute to the data (Figure 1). To transition from in-house data driven

Marsh, N. and Marsh, B. – Data democratisation: empowering the environmental management community

decisions to a broader data democratisation approach requires non specialist data management, analysis and visualisation tools (blue in Figure 1).

This paper will firstly present the case for the value of data democratisation, and why this should be based on data visualisation (not just tables of data). We present an overview of the open data movement, some examples of emerging non specialist data tools to illustrate how open data may be incorporated into the NRM workflow, and lastly we present some examples of where NRM organisations have embraced data democratisation.

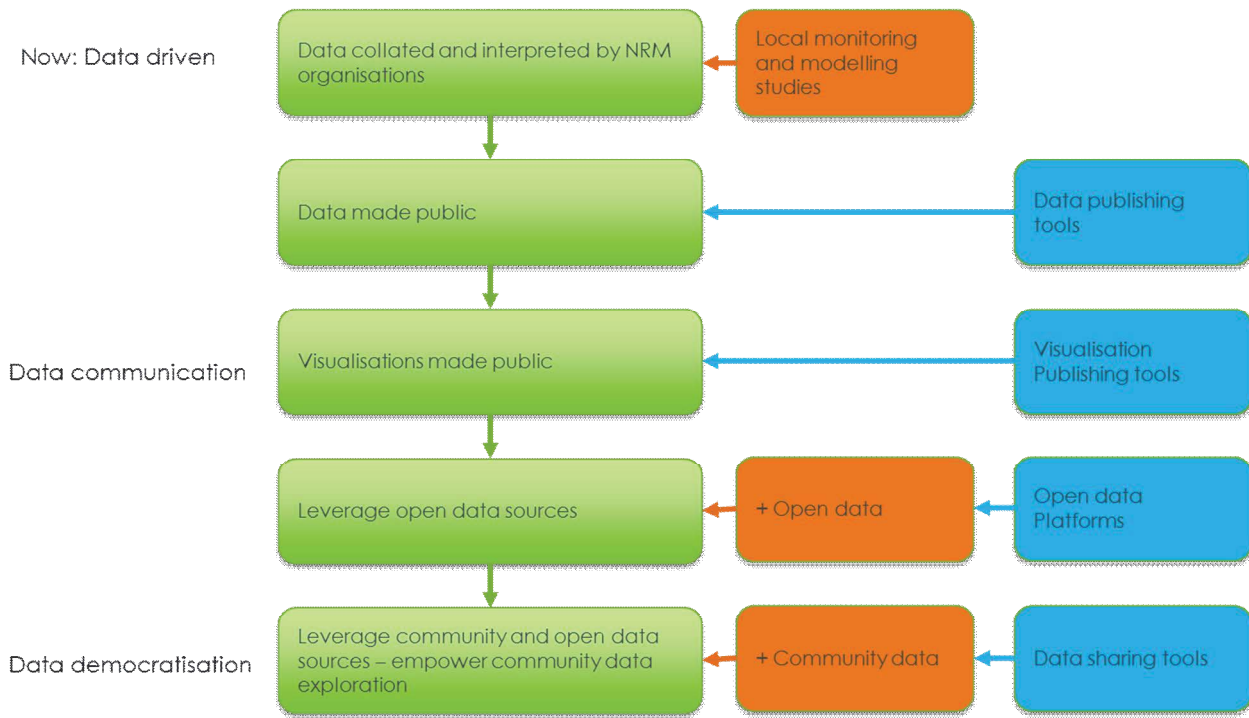


Figure 1: Moving from data driven NRM to data democratization (top to bottom in green, data sources in orange, and non-specialist tools required in blue).

The value of data democratisation

To illustrate the value of data driven decision making from a business perspective, a survey of 330 public North American companies showed a distribution of data driven management approaches. Those companies in the top third (i.e. the most data driven) were 5% more productive and 6% more profitable than their competitors (accounting for other factors) (McAfee and Brynjolfsson 2012). So using data to make decisions over and above raw skill and experience produces a measurably better outcome.

It is not simply the need to be data driven, but how data is presented that can have a significant impact. Intuitively, a great chart trumps a table of numbers at communicating a key idea or trend. Pandey et al (2014) tested this hypothesis by running three alternative experiments across 720 individuals to test the persuasiveness of charts versus tables. The results show that for people who were not strongly polarised in their views (tables nor charts persuade zealots) charts were 8% more effective at converting views than the same data presented as a table.

So being data driven is better than raw skill, and data that is well presented through visualisations is better again. Natural resource management organisations are by their very nature community organisations. They

Marsh, N. and Marsh, B. – Data democratisation: empowering the environmental management community

may be managed by paid staff but the success of their outcomes is entirely dependent on a broad number of stakeholders. Those stakeholders include the landholders who need to be convinced of the value of doing works on their land and the broader community who willingly donate their time and enthusiasm. We argue that NRM stakeholders are critical participants in the decision making process and should be empowered and engaged through effective communication. Our view of effective communication is not simply published reports and community workshops but the underlying data or data summaries that present competing maps and charts demonstrating the value of proposed catchment management strategies.

Big data and open data

The global economy is starting to operate in real time thanks to live streams of data showing where people are traveling, what they are buying and the flow of energy and money (McKinsey Global Institute 2013). Corporate decisions are being made in real time based on these data streams. This is big data analytics at work. There is a complimentary trend, 'open data', which has now gained a critical mass. Open data is simply the release of data from government (and to a limited extent private) institutions to allow its repurposing and leveraging for any use.

The basic premise of open data is to encourage government accountability by placing government collected data in the public domain. The open data movement which is premised on making government data "open by default" was supported by the 2013 G8 signing of the 'Open Data Charter'. In 2014, the G20 followed up by pledging to advance open data as a weapon against corruption, and the UN recognised the need for a "Data Revolution" to achieve global development goals (The World Wide Web Foundation 2015).

However the opportunities created by open data are far more than the initial intent of government accountability. There are significant economic benefits from the open data movement. A study into the broader economic benefits of open data estimates an annual benefit of \$US3.2-5.4 trillion economic potential across seven industry sectors (McKinsey Global Institute 2013). This equates to around 4-7% of gross world product (GWP). To put this in some perspective, the global financial services sector (retail banking, insurance) was estimated at US\$6.6 trillion in 2011 (Roxburgh et al 2011) (about 8.4% of GWP).

These estimates seem high, estimates of 'potential' are fraught with value judgments that can lead to wildly varying predictions. A study by the European Union to directly measure the actual realised benefits of open data, showed a direct and indirect value to the European Union of open data of 200 billion Euros (about 1.5% of GDP) (Carrer et al 2015). Given that the open data movement is still in its relatively early stages, and that the measured value is already significant, the predictions of the McKinsey Institute of an industry that rivals the size of the financial services sector may prove to be accurate.

Approaching critical mass for open data

For open data to realise these additional benefits, it has to meet some basic criteria such as being freely available and in formats that allow its reuse. Furthermore, there needs to be enough volume of data to generate the follow-on services built to use that data. The World Wide Web Foundation maintains and undertakes quantitative analysis of the direct and indirect impact of open data across 86 countries through ongoing surveys, review of the internet mentions and use of open data in scientific publications. Figure 2 summarises the impact of open data (y axis) based on the 'readiness' of open data by country (x axis). Where readiness is largely a representation of the volume and quality of open data availability. From Figure 2, there is clearly a two stage process in place, whereby little impact is realised until a readiness score approaching 40 is realised. Australia is illustrated by the green circle. We have a high degree of readiness (in the top 20% of countries), but our impact is still relatively low.

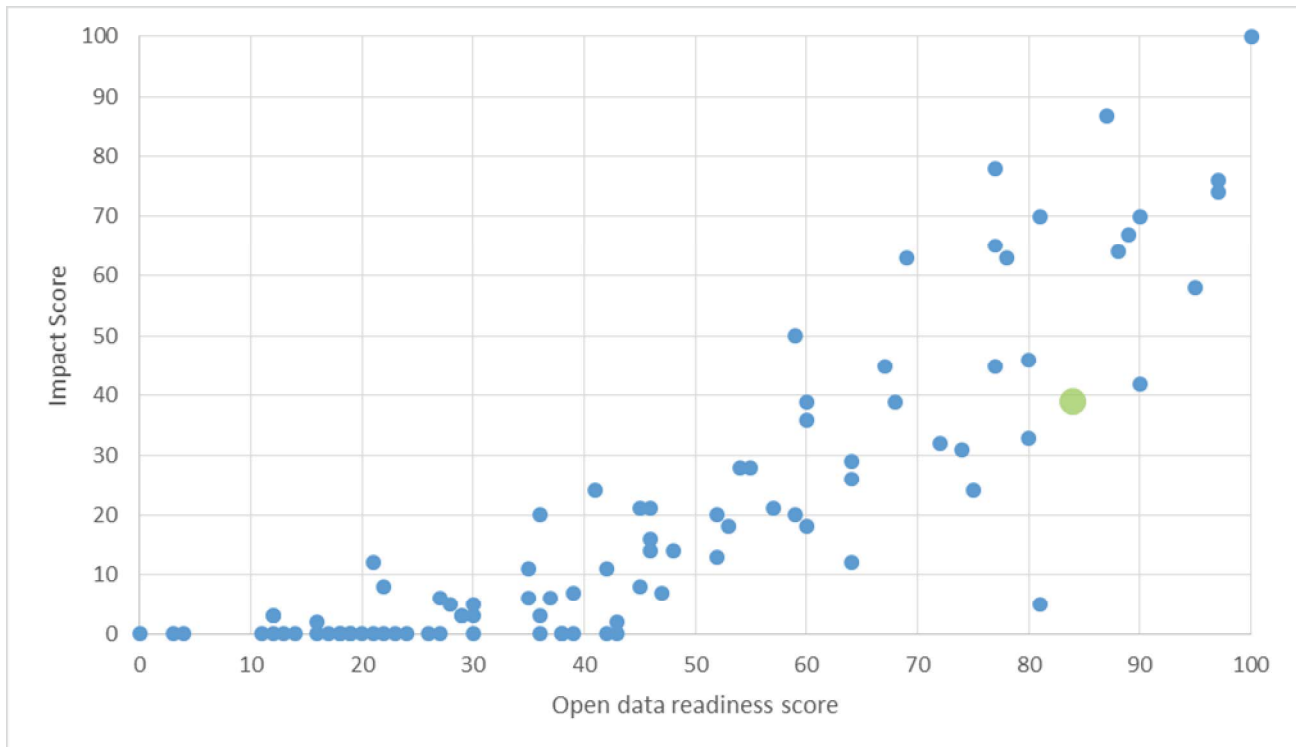


Figure 2: Impact of open data is not realised until a critical mass is achieved (Australia is green) (Source The World Wide Web Foundation 2015)

What does this mean for the environment sector?

The collective open data environment in Australia is relatively advanced, which means high volumes of good quality data are being made available. What is of particular interest to the environment sector is that the bulk of open data is environment related data. For example, in May 2016 the federal government’s open data portal (data.gov.au) had 7,846 published data sets, of these, 4,993 (64%) were directly from environmental and physical science organisations such as geosciences Australia. Whilst the open data movement was initiated to engender greater government accountability, the environmental sector can benefit greatly because the infrastructure and processes that are now in place to publish all government data disproportionately benefit the environment sector.

Tools to support data democratisation

In order to empower the community through data one needs simple fast and effective ways of creating and publishing data visualisations as well as access to worthwhile data.

Open data sources

The federal government maintains an open data portal (www.data.gov.au) which publishes not only federal government derived datasets, but also republishes syndicated data from other government portals. The Open Data Institute of Queensland (ODIQ) maintains a list of Australian based open data portals (<http://queensland.theodi.org/home/resources/data/>) which is an excellent starting point for finding local

8ASM Full Paper

Marsh, N. and Marsh, B. – Data democratisation: empowering the environmental management community

and relevant natural resource data sets. At the time of writing the ODIQ open data portal had 55 different Australian based open data repositories listed.

Publishing platforms

Natural resource management organisation are not only users of open data but can contribute back to the open data resources by publishing their data (Marsh and Marsh 2014). Most of the existing open data portals use the same free, open source data publishing infrastructure (CKAN.org). The platform is free to install and run for any organisation, hosting solutions are provided, and there are also established portals that will allow anyone to freely publish data. The most notable free portal for anyone to publish data is 'data hub' (<https://datahub.io/>). The great advantage of using the CKAN approach (other than being free) is that it is well supported and data searchers have a similar experience across the different platforms. The underlying CKAN database also allows the syndication of data, so that data aggregators such as the Australian Federal Government can link to a large collection of publishing houses to produce a mega data library.

Data visualisation

Data visualisation is an expanding field with many novel forms of data visualisation emerging. The great advantage of web based data visualisation is that dynamic and animated data visualisations are possible. For example, instead of a static chart or a static map, these visualisations can be animated to show how conditions change through time. Data exploration through 'progressive disclosure' is also a great feature possible through online visualisations. The concept of progressive closure is to present the simplest user experience, and allow the user to drill down to explore the data in ever increasing detail. There has been recent growth in the availability on non-specialist tools for data visualisation:

- www.plot.ly – create scatter, line, bar and pie charts online.
- www.visual.ly – graphic design firm who will create your visualisation.
- www.mapbox.com – create online maps.
- www.click2map.com – create online maps.
- www.truii.com – create data pages by combining maps and charts. Embed those maps or charts in your own web site.
- Microsoft Power BI – create dashboards.

Data sharing

To truly embrace data democratisation one needs an easy way to manage data from many contributors. The CKAN system for publishing data is an excellent repository for releasing cleaned and annotated data. However, there are limited online collaboration tools available that allow users to get their data into a tidy format. This process of creating tidy data accounts for around 80% of the time involved in data analysis (Wickham 2012). In our view, the data tidying or wrangling process should be a collaborative process and the data needs to stay in one place with tight version control and editing attribution. The Truii.com tool provides this level of data wrangling support as an online tool.

Data democratisation

There are no true data democratisation examples that we could find whereby community members are actively contributing to a public data resource for natural resource management. We have however identified several organisations that are well on the way by actively making the internal organisational datasets available online. Following are some brief examples.

South East Queensland Healthy Waterways

The South East Queensland Healthy Waterways Partnership (HWP) is a collaboration between 39 local state and industry groups in south east Queensland. The HWP produces an annual report card to rate the environmental performance of the catchments of the region over the preceding 12 months (Figure 3). The HWP has recently embarked on a major modelling exercise and will release the underlying data, initially to its partner organisations through web based data visualisations (using the Truii.com platform). This will be the first example that we know of where the results of an extensive catchment and waterway modelling program will be released through online data visualisations.

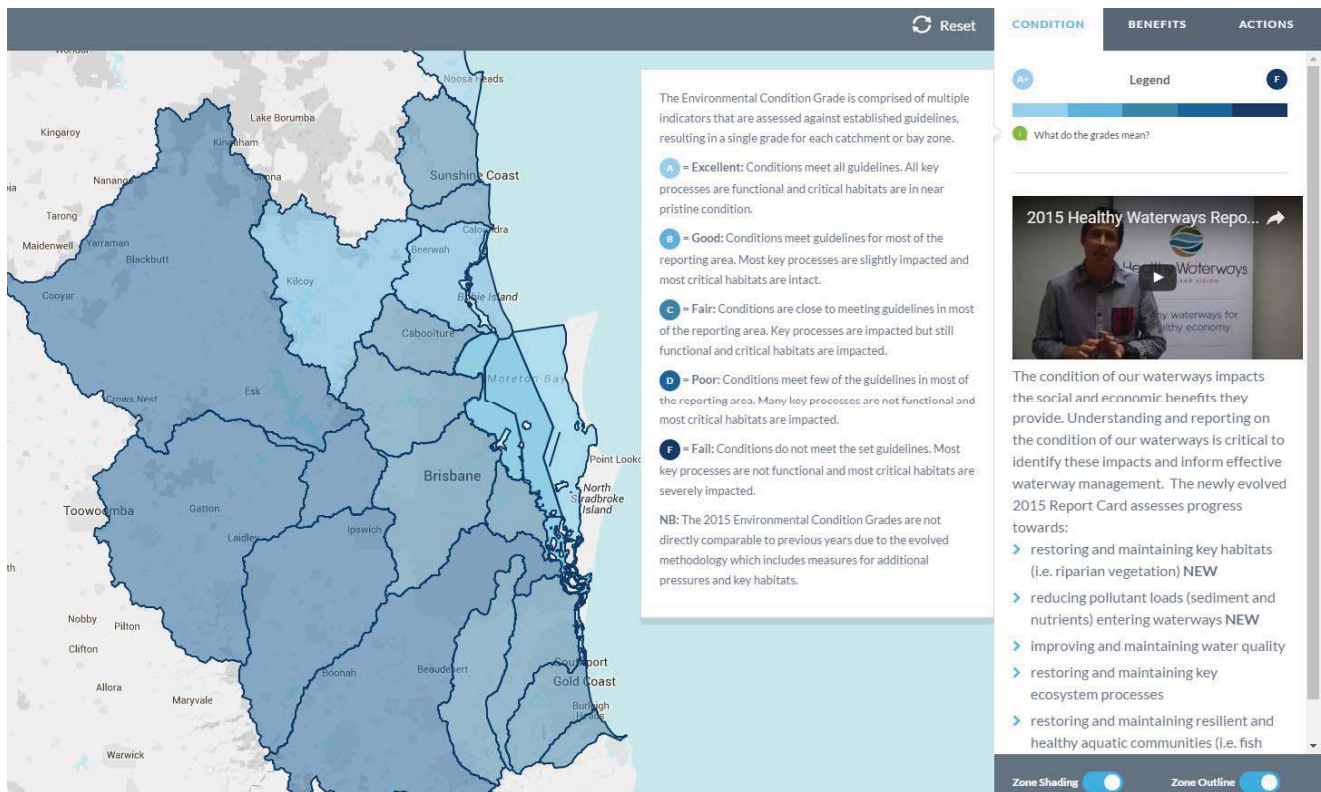


Figure 3: Healthy Waterways annual health reporting with progressive disclosure and several layers of detail that users can drill into (source <http://healthywaterways.org/reportcard/#/overview/condition>)

Murray Darling Basin Authority

The Murray Darling Basin Authority (MDBA) is a federal authority for water management within the Murray Darling basin. The MDBA has been very progressive in making data, maps, reports and images available on its website in recent years (Figure 4).

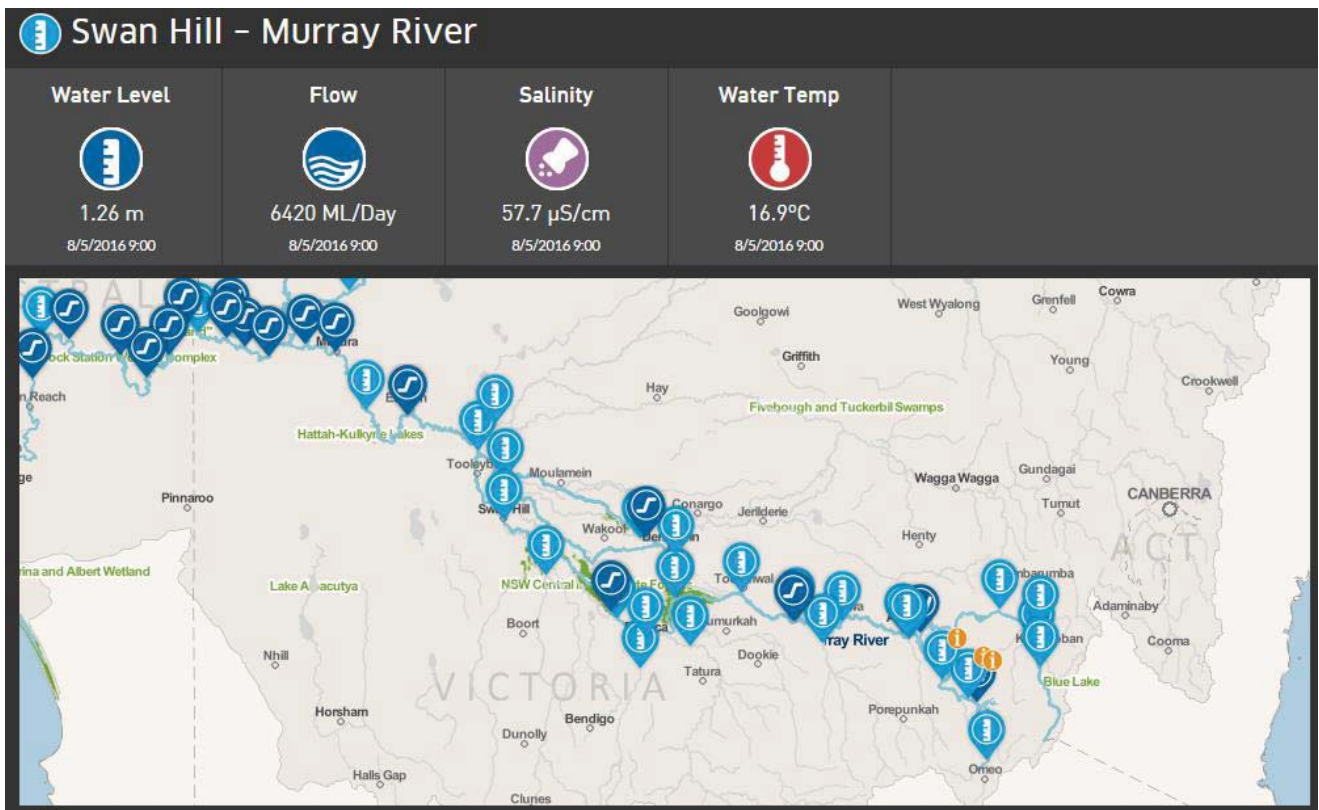


Figure 4: MDBA presents live river data for key locations on the main channels (source: <http://livedata.mdba.gov.au/>)

The Atlas of Living Australia

The Atlas of Living Australia is a fantastic resource of species observation data for Australia. Major contributors to the Atlas are the collective state based museums and the broader scientific and enthusiast community. The Atlas of Living Australia has several great products and ways to access and repurpose the underlying data as well as simple yet powerful exploration tools such as ‘explore your area’, whereby you can see all the species observations reported within a set radius (Figure 5).

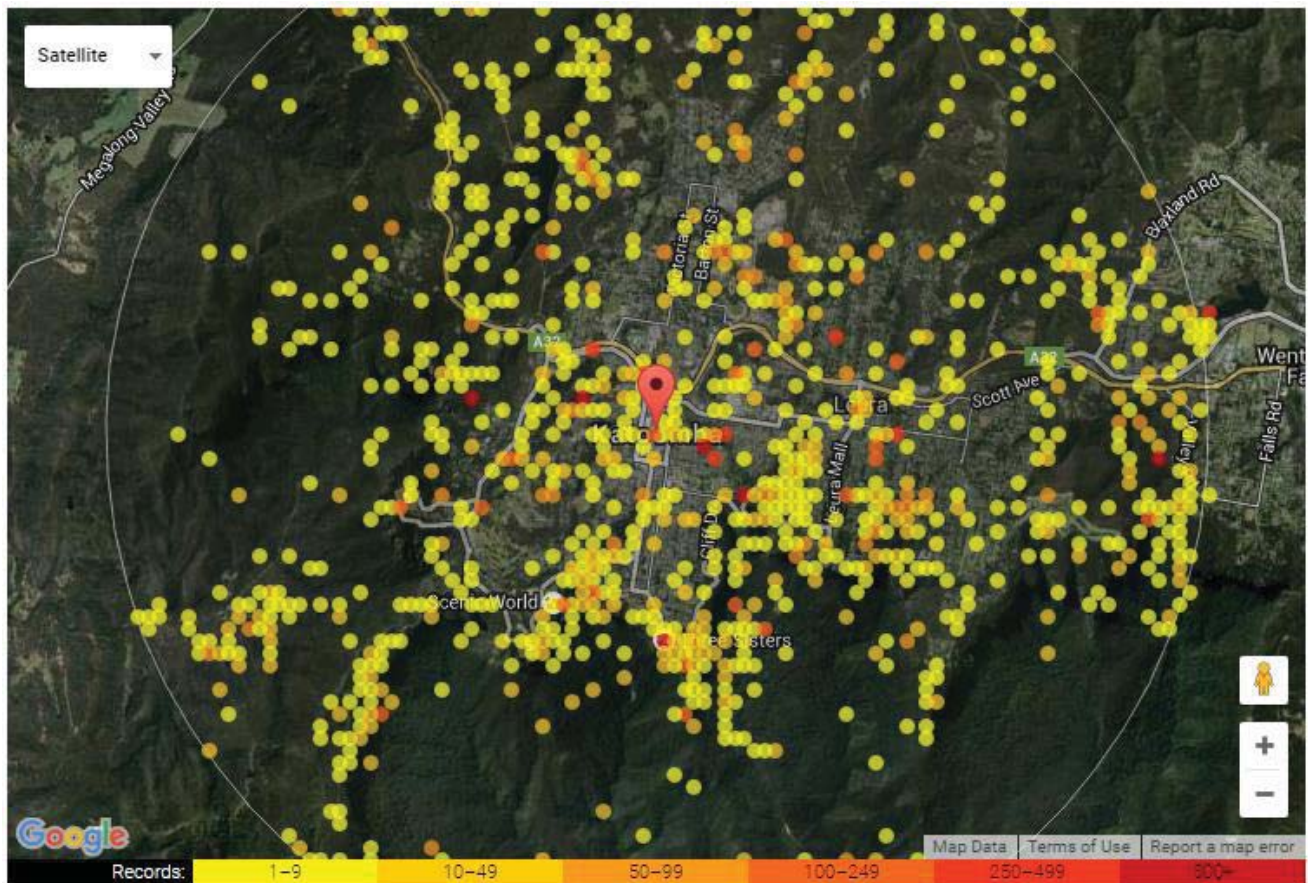


Figure 5: Atlas of Living Australia ‘explore your Area’ showing the number of observations of different species within a given radius (Source http://biocache.ala.org.au/explore/your-area#-35.28|149.12|12|ALL_SPECIES).

Conclusions

Organisations that make data driven decisions are more effective than those that do not. Data communicated through visualisations such as charts and maps are more persuasive than tables of numbers. Natural resource management organisations are data driven and are familiar with creating compelling data driven narratives through data visualisation. We would encourage these organisations to move to creating online versions of their data visualisations to further engage the broader community. We would further recommend a move toward empowering the broader community to contribute their own data to the collective catchment understanding to truly embrace the flow of data through data democratisation.

The explosion of freely available open data published in reusable formats along with the growing collection of non-specialist tools brings the opportunities of data democratisation to natural resource management organisations. The open data movement was initiated for greater government scrutiny, however 2/3 of published open data relates to environmental observations. This availability of data and tools presents an opportunity to natural resource managers to leverage their own local monitoring and modelling data with broader government collected data.

8ASM Full Paper

Marsh, N. and Marsh, B. – Data democratisation: empowering the environmental management community

References

- Carrer, W., San Chan, W., Fischer, S., Van Steenberg, E. (2015) *Creating Value through Open Data: European Data Portal*. European Commission.
- Roxburgh, C., Lund, S., and Piotrowski, J. (2011) *Mapping global capital markets 2011*, The McKinsey Global Institute.
- Marsh, N., Marsh, B. (2014) *Big Data concepts for natural resource management*, in Editors Names, Proceedings of the 7th Australian Stream Management Conference. Townsville, Queensland pp 156-161,
- McAfee, A, and Brynjolfsson (2012) *Big Data: the management revolution*, Harvard Business review, October 2012 issues (<https://hbr.org/2012/10/big-data-the-management-revolution/ar>)
- Open Data Charter G8 – 2013 <https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex>
- Pandey, A., Manivannan, A., Nov, O., Satterthwaite, M., and Bertini, E. (2014) The Persuasive Power of Data Visualization. *New York University Public Law and Legal Theory Working Papers*. Paper 474. http://lsr.nellco.org/nyu_plltwp/474
- The World Wide Web Foundation (2015) *Open data barometer*, The World Wide Web Foundation (access via www.opendatabarometer.org)
- Wickham, H (2014) *Tidy Data*. Journal of Statistical Software. 159, 10.