

The Big Data Analysis Challenge for Landscape Architecture

Don Royds

Lincoln University, Canterbury/NZ · don.royds@lincoln.ac.nz

Abstract: Big data is growing in volume, variety, and velocity and is becoming more available and significant. Big data, if used as part of the design process by landscape architects, has the potential to broaden and inform site understanding. However, working with big data and analysing the diverse range of datasets is currently the main challenge. It is unclear whether the tools used in landscape architecture are sufficient to work with the volume, variety or velocity of big data. Through a descriptive critique and a matrix evaluation, a clearer understanding of the current tools used in data analysis approaches, are presented. Three data analysis approaches are applied to a case study site and reveals that there is a clear gap between analysing geospatial data and non-spatial data. The implications are that if landscape architecture is to take full advantage of working with big data, new data analysis tools need to be developed.

Keywords: Big data, overlay method, Geodesign, data visualisation, data analysis

1 Introduction

Landscape architecture is a complex, interdisciplinary profession which takes into account many different streams of information using a broad scientific knowledge whilst applying an artistic creativity at the same time (GAZVODA 2002) with a wide range of tools (GONOT 2013). The profession explores ways of interpreting and representing attributes of the landscape including how data “can visualize the intangible and often invisible forces that shape our environment” (AMOROSO 2015, 4).

Through web-based portals and live streams, fed by a growing, connected and mobile culture, data is growing at an exponential rate. This data is not only growing in volume but is also becoming increasingly diverse and more available to not just businesses but to the wider public. This is often referred to as ‘big data’. MAYER-SCHÖNBERGER & CUKIER state that “there is no rigorous definition of big data” (2013, 6) and after surveying the definitions that have previously been presented for big data, DE MAURO, GRECO & GRIMALDI (2015) proposed three V’s to define it: Volume, Velocity and Variety. Volume refers to the quantity of the data (which can range from a text file of a few kilobytes to a cache of terabytes of CCTV footage), velocity refers to the speed at which the data can arrive (especially human-sourced streaming data), and variety refers to the increasingly diverse formats that the data is available as (SOUBRA 2012).

Big data can help us to see relationships between disparate pieces of information and point us toward a deeper understanding to derive correlations (MAYER-SCHÖNBERGER & CUKIER 2013). This understanding can inform decisions based on verifiable data using datasets which are no longer samples but that are inclusive of an entire layer, population, scenario, or narrative.

Big data is being explored within landscape architecture. For example, individual social media datasets, such as Flickr with their geotagged photos, have been worked with to reveal

how people interact with the environment (LEE 2013). Even prominent landscape architecture theories, such as Landscape Urbanism, have underlying principles that benefit from the use of big data (WELLER 2007). So what data analysis tools are currently being used by landscape architects, and how well do these work with analysing big data?

2 Case Study: Testing the tools

2.1 Selection of data analysis tools

First, a literature review was undertaken to identify which tools were available, including those currently used within landscape architecture. These tools were evaluated using qualitative criteria with three being selected as a feasible number for testing applicability with big data using a case study method.

The literature review (major sources include DANGERMOND (2010), GAZVODA (2002), LEE (2013), LOHR (2015), SIMON (2014), STEINITZ (2010), TURNER (2014), WALLISS & RAHMANN (2016)) identified seven data analysis tools, and these were then qualitatively evaluated against criteria which relate to big data, along with author-created criteria based on common practice requirements. The first three criteria were Volume, Variety and Velocity, and the last four criteria were Cost (affordability), Use (easy to learn), Geo- (ability to work with site data) and Scale (small site scale to regional scale). One of four scores for each criterion were possible: -2 (does not meet the criterion at all), -1 (meets the criterion somewhat), +1 (meets the criterion well) and +2 (meets the criterion very well). 0 was not used.

The scores from the qualitative evaluation were tabulated (see Table 1.) with the first three being the more frequently-used data analysis tools within landscape architecture.

Table 1: Qualitative Criteria Matrix showing resulting scores from the evaluation of the data analysis tools identified in the literature review

	Big Data Criteria				Professional Criteria					
	Vol.	Var.	Vel.		Cost	Use	Geo-	Scale		
SAD method	-2	-1	-2	-5	+2	+2	+1	-1	+4	-1
Overlay Method	-1	-1	-1	-3	+2	+1	+1	+1	+5	+2
Geodesign	+2	-1	+1	+2	-1	-1	+2	+2	+2	+4
Data Augmented Design	+1	-1	-1	-1	+1	-2	+2	+2	+3	+2
Data Visualisation	+2	+2	-1	+3	+1	+1	-1	+2	+3	+6
Data Storytelling	+2	+2	+2	+6	-2	-2	+1	+1	-2	+4
Data Sensory Perception	+2	+1	+2	+5	-2	-1	-1	-2	-6	-1

The Overlay Method, Geodesign and Data Visualisation were selected based on their scores, with the first two being the highest scoring tools within the top section and the last tool being the highest in the bottom section. The rationale for this was to ensure that the current tools used within landscape architecture were tested (the SAD method was excluded as it was the lowest scoring data analysis tool).

Before outlining the application of these three tools, it is appropriate to define each one (including any assumptions made or limitations).

The Overlay Method is a common practice used by landscape architects as it has been used for teaching landscape architecture students (TURNER 2014). It relies on maps which show discrete datasets or ‘values’ to be physically overlaid above each other, with the result showing areas where there are overlaps between them or gaps where the ‘values’ are not present. The method itself contains no mechanism to analyse data but does give the data a hierarchy – which could later be used in analysis. This lack of a mechanism also means that there is a simple linear process that is followed and McHarg himself acknowledged this with his statement: “... where any man, assembling the same evidence, would come to the same conclusion” (TURNER 1996, 146).

Geodesign is currently not clearly defined and there are several definitions under discussion (FLAXMAN 2010) but the common attributes of the varied definitions require the use of digital tools and begins with identifying layers, or geographic features, to evaluate. The geospatial analysis tools, such as Geoprocessing Tools, require a predetermined set of land uses or land-use management strategies in order to assess or create an output (MILLER 2012). These outputs are either geographic displays such as maps, or scalar values (working with a shape). For this case study, the first phase in the Geodesign workflow (Design Instantiation) as described by FLAXMAN (2010) was tested as further phases require a design proposal to continue.

Data Visualisation spans across many fields including science and medicine and has been shaped by many individuals including Minard, Tukey, Tufte, McCandless and Yau. Common attributes of data visualisation include the simplifying data, use of software, and cognitive processing. There is a reliance on the ‘designer’ to look at the data and question it, investigating which visual methods would aid its processing. Academics and scientists in the area of information design are now being joined by Internet users who also appreciate their effectiveness (LANKOW et al. 2012).

The case study site was selected based on the criteria that it was: a significant civic space (such as a park); had a high-profile within the landscape architecture profession (such as a sustainable design exemplar); and had the potential for collecting a large amount of data (such as a well-established site within an urban area). Based on these criteria, Waitangi Park (Wellington, New Zealand) was selected.

2.2 Data Collection

Data was collected over a three-month period, from known online sources, searchable online sources and were all freely available for download. The collection of the data was directed by searching databases, such as NZStats, LINZ, Twitter and YouTube, using keywords relevant to the site, and then expanded to searches using search engines (Google, Bing, etc.). A rich data source is of course Social Media and the following social media sources were ex-

amined: Twitter, YouTube, Facebook, Tumblr, Flickr, Instagram, Google, Picasa, Vimeo and Blogs. There was the expectation that this would yield a 'flood' of data but only a limited amount of data was readily available. There were several reasons for this, for example Facebook allows access to a user's posts only if they permit it, however by using the hashtag filter #waitangipark, posts with that hashtag applied were visible and could then be exported as a PDF. Another reason was that mining the data from these was time consuming. In this research, no freely available tools were found to make the task more efficient, and relied on the use of a spreadsheet or PDF export.

In summary, the data gathered during the data collection phase for the case study method comprised of: 1.74 GB total folder size of raw, collated data; 111 individual datasets; 49 available as geospatial data; 17 PDFs; and 45 Excel spreadsheets. Significant datasets included: WELLINGTON CITY COUNCIL (2015), ATKINSON, J. et al. (2013), STATISTICS NEW ZEALAND (2013), LITMUS (2011), WELLINGTON CITY COUNCIL (2009).

2.3 Application of Data Analysis Tools

In applying the Overlay Method, software was used (ArcMap) only to construct basic maps with the focus of this tool's applicability through physical media. Georeferenced shapefile layers were added into ArcMap, their transparency was reduced to 50 % to aid the 'layering' process, and were then printed onto overhead projector transparency film. They were then overlaid into one 'stack'. This 'stack' contained 30 layers and to ensure that the integrity of the film itself was not a factor it was placed on top of a light table. These layers were then separated out into three groupings: line layers, point layers and area layers, and examined as individual groups.

As Geodesign is synonymous with GIS, ArcMap was used for the application of this data analysis approach. Compatible shapefiles from the data collection phase were added in ArcMap and the layers which had subcategories (such as values 1-9) were identified and given a colour gradient ramp to display these values in more detail.

To begin the data analysis, a series of questions that related to the data and the site were formulated to query the data, then the relevant geoprocessing tools and workflow were identified in order to achieve an answer to the question proposed. Figure 2 shows the result from the question "What is the density of the buildings near the case study site?"

In applying the Data Visualisation approach, all the available data was examined by reading the rows and columns of data or text from Excel files, Database files and PDFs. Then all the terminology and attributes were reviewed and investigated further from their sources to gain a deeper understanding of what these values meant. Next, Excel and Tableau were used to interrogate the data to identify patterns or relationships of interest.

3 Results

3.1 Overlay Method

Figure 1 shows the result from applying the Overlay Method after the 30 layers of film were grouped into point, line and area layers. The original ‘stack’ of 30 layers showed a result that was ‘muddy’, confusing, and did not appear to reveal anything of use.



Fig. 1: Composite groupings using point, line, and area layers of data for the Overlay Method

3.2 Geodesign

The results from applying the Geodesign approach were six maps which provided a geographical understanding for the topics covered by the original six questions, along with a working GIS file which could be further scrutinized. Figure 2 shows the result from the questioning of data relating to building density near the case study site.



Fig. 2: A map produced using Geodesign to answer the question: ‘What is the density of the buildings near the case study site?’

3.3 Data Visualisation

The Data Visualisation approach produced statistics, charts and graphics throughout the process which, along with other patterns of interest, were summarised and combined into one final data visual (Infographic) using a template developed by YAU (2011). The result is shown in Figure 3. It provides the viewer with a ‘dashboard’ of sorts, with a variety of information available in one glance, which could then be investigated further.



Fig. 3:

The Data Visualisation approach produced many statistics and graphics throughout its application which can then be incorporated into one large infographic

4 Discussion

The Overlay Method, especially when using physical media, was not suitable for big data. This was not improved when data was coordinated into grouped maps and did not facilitate a robust understanding of the data. As this tool is reductive i. e. aims to identify a single or series of 'suitable' areas, interaction between datasets is discouraged. This was demonstrated by adding only 30 layers of information with no usable outcome, being obscured through physical stacking, and when the layers were grouped into point, line and area layers it was marginally more legible.

Geodesign was suitable for only the geo-spatial datasets and has no current way to display data that has no coordinates. A challenge when using Geodesign with big data is that for any analysis to begin, a set of features or layers needs to be identified to direct the process (along with a design). While this does provide a starting point, it can overlook other datasets that are not initially apparent or significant and narrows any capturing of big data too early.

Out of the three data analysis tools tested, Data Visualisation was the most suitable data analysis approach for working with big data but does rely heavily on cognitive processing. It is an exploratory process which can convey large amounts of data quickly, or just the 'juicy bits', and has the potential for misrepresenting data through individual interpretation (TUFTE 1983). It has some scope for working with both geospatial and non-spatial data and was the only tools that could incorporate data such as the average age for visitors, topics of YouTube videos, whether tweets were commercial or personal, or descriptors of visitors. This data analysis approach proved to be the most challenging to apply – not just because of the complexity of big data but because it used a new data analysis tool to the author. Despite experts writing extensively on this topic, there was little guidance to suggest a standard process to follow across all the datasets and this meant that the process was an organic, explorative one. This approach relies on the cognitive processing ability of the brain to spot and associate patterns or trends in data that computational algorithms can't (MAZZA 2009), but in order to do this, a great deal of time must be invested in becoming familiar with the data. Part of this time challenge is not only going through line by line in spreadsheets, but also deciphering and decoding technical classifications.

While Geodesign does call for visualisation of data, it would be hoped that the assistance of new available features such, as the GeoEnrichment service and 'Insights for ArcGIS', different results could be produced in future testing. Still, a different or hybrid data analysis tool is needed to incorporate both non-spatial and geo-spatial data.

Communicating the data through visual representation was however a common component of the three data analysis tools tested.

5 Conclusion

There is more scope for the use of interactivity and real-time input for bridging this gap, but this could dilute the efficacy of data visualisation which works best when all of the summary data visuals are displayed in one combined area – as opposed to hiding away waiting for a mouse click or mouse hover. CHESHIRE & UBERTI discuss interactivity when they looked at making their data visualisations interactive and their conclusion was "you can simply see

more at a glance on a printed page than you can pinching, tapping and scrolling on a smart-phone screen” (2014, 28).

There is an increasing list of software programs and online tools available for data analysis and these could be evaluated further on their potential to bridge this gap between geospatial and non-spatial data, along with the balancing of machine and cognitive processing.

We are moving away from *just* maps and it is clear that the current data analysis tools within landscape architecture are not sufficient for working with big data, particularly with the increasing variety of geo-spatial and non-spatial datasets available.

Data Availability Statement

This article uses a large number of datasets from publically available third-party sources. Datasets heavily used in this article have been cited in the reference list. Details of other datasets can be obtained on request from the author.

References

- AMOROSO, N. (2015), Representations of the Landscapes via the Digital: Drawing types. In: AMOROSO, N. (Ed.), *Representing Landscapes: Digital*. Routledge.
- ATKINSON, J., SALMOND, C. & CRAMPTON, P. (2013), NZDep2013 Index of Deprivation [Dataset]. <http://www.otago.ac.nz/wellington/departments/publichealth/research/hirp/otago020194.html>.
- CHESHIRE, J. & UBERTI, O. (2014), London – The Information Capital. Particular Books.
- DANGERMOND, J. (2010), Geodesign and GIS – Designing our Futures. In: BUHMANN, E., PIETSCH, M. & KRETZLER, E. (Eds.), *Peer Reviewed Proceedings Digital Landscape Architecture*, Anhalt University of Applied Science, Germany. Wichmann, Berlin/Offenbach, 502-514.
- DE MAURO, A., GRECO, M. & GRIMALDI, M. (2015, February), What is Big Data? A Consensual Definition and a Review of Key Research Topics. In: *AIP Conference Proceedings*, 1644 (1), 97-104.
- GAZVODA, D. (2002), Characteristics of Modern Landscape Architecture and Its Education. *Landscape and Urban Planning*, 60 (2), 117-133.
- GONOT, J. (2013, July 26), Part Three: Big Data to Big Design [Blog Post]. <http://land8.com/profiles/blogs/part-three-big-data-to-big-design>.
- LANKOW, J., RITCHIE, J., & CROOKS, R. (2012), *Infographics: The Power of Visual Storytelling*. John Wiley & Sons.
- LEE, L. (2013 JANUARY), The View from the Crowds. *Landscape Architecture Magazine*, 103 (1), 35-40.
- LITMUS (2011), *Wellington's Waterfront: Are we on the right track?* <http://wellington.govt.nz/~media/about-wellington/research-and-evaluation/built-environment/2011-waterfront-right-track.pdf>.
- LOHR, S. (2015), *Data-ism: Inside the Big Data Revolution*. London: Oneworld Publications.
- MAYER-SCHÖNBERGER, V. & CUKIER, K. (2013), *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Houghton Mifflin Harcourt.

- MAZZA, R. (2009), *Introduction to Information Visualization*. Springer Science & Business Media.
- MILLER, W. R. (2012), *Introducing Geodesign: The Concept*.
- SIMON, P. (2014), *The Visual Organization: Data Visualization, Big Data, and the Quest for Better Decisions*.
- SOUBRA, D. (2012, July 5), *The 3Vs that Define Big Data*.
<http://www.datasciencecentral.com/forum/topics/the-3vs-that-define-big-data>
- STATISTICS NEW ZEALAND (2013), *Meshblock 2013 [Dataset]*.
<https://datafinder.stats.govt.nz/layer/8347-meshblock-2013/>.
- STEINITZ, C. (2010), *Landscape Architecture into the 21st Century – Methods for Digital Techniques*. In: BUHMANN, E., PIETSCH, M. & KRETZLER, E. (Eds.), *Peer Reviewed Proceedings Digital Landscape Architecture*, Anhalt University of Applied Science, Germany. Wichmann, Berlin/Offenbach, 2-26.
- TUFTE, E. (1983), *The Visual Display of Quantitative Information*. Graphics Press.
- TURNER, T. (1996), *City as Landscape: A Post Post-modern View of Design and Planning*. Spon, London.
- TURNER, T. (2014), *Landscape Design Theory & Methods: Modern, Postmodern & Post-postmodern, including Landscape Ecological Urbanism & Geodesign [Google books]*.
<http://books.google.com>.
- WALLISS, J. & RAHMANN, H. (2016), *Landscape Architecture and Digital Technologies: Re-conceptualising design and making*. Routledge.
- WELLER, R. (2007), *Global Theory, Local Practice*. *Kerb*, 15, 66-71.
- WELLINGTON CITY COUNCIL (2009), *Wellington City 1m Contours (2009) [Dataset]*.
<https://koordinates.com/layer/1479-wellington-city-1m-contours-2009/>
- WELLINGTON CITY COUNCIL (2015), *Rating Valuations [Dataset]*.
<https://wellington.govt.nz/services/rates-and-property/property/property-search>.
- YAU, N. (2011), *Visualize this: the FlowingData guide to Design, Visualization, and Statistics*. Wiley Publishing, Indianapolis.