Big Data and Architectural History

How to Characterise Architects Using Big Data

Adrian Humphris

Geoff Mew

Abstract

Analysing large sets of data is often used to uncover hidden patterns, unknown correlations, market trends, customer preferences and other potentially useful information. It is increasingly popular and much more relevant as data has become more widely accessible and computing technology has advanced.

Could this approach tell us more about the characteristics of New Zealand architects? In researching our book 'Raupo to Deco' we created a large set of data about architects and the buildings they designed. Local authority permit records, newspaper tender notices, and many other sources were incorporated - this allowed us to identify architects, the time frames they worked in, the types and numbers of buildings they designed, and the connections and relationships between individuals and practices. We also used genealogy techniques to work out birth and death dates and to track where they practiced over time.

We have continued to build this dataset, broadening it to cover the whole of New Zealand. Currently we have more than 25,000 individual building tenders for buildings between 1840 and 1940, and information on more than 1,500 'architects' over the same period.

This paper illustrates some aspects of how we can present a view of 'an architect' over time. How many were there? How old were they and how long had they been practicing, and did this change over time as the industry developed? Did architectural output match the building cycle, or were architects insulated from regular booms and busts? And how specialised were architects in different areas of design, and did this also change over time? We also discuss some of the advantages and pitfalls of dealing with a large data set, and explore how we can ensure the validity and accuracy of the results.

Background

New Zealand architecture in the broadest sense has only been described in three major works, *Colonial Architecture in New Zealand*[1], *Looking at the Architecture of New Zealand*[2], and *A History of Architecture in New Zealand*[3]. The latter has been progressively revised in an attempt to keep up with current trends. All the authors were or are architectural historians who have used conventional research techniques to compile their works and have concentrated on the built environment with little emphasis on the characters of building designers.

Our book, *Raupo to Deco: Wellington Styles and Architects 1840 - 1940*, was an attempt to look more closely at the personalities of Wellington architects as well as the buildings they designed. However we also largely used conventional methods of research for data accumulation and assembly. These included reviewing prior literature, both published and unpublished on the individual architects selected for inclusion, summarising their work from sources such as their practice records, tender notices and building permit records and consulting available definitions of different architectural styles over time. Unlike the authors of detailed biographies of some New Zealand architects, such as those on William Mason, Benjamin Mountfort and Frederick de Jersey Clere[4], we did not attempt to portray our architects in comparison with their contemporaries, nor did we attempt to exhaustively list all their individual works.

A further project, currently underway, is a more in-depth study of a selection of nationally recognised architects who practised between 1840 and 1940. In order to do this, we perceived a need to expand our Wellington database to cover New Zealand in the most efficient and objective way possible. In this paper we describe some of the methods used and illustrate some of the results achieved from compiling and working with large data sets.

Compiling a big data set

During the writing of *Raupo to Deco* a new potential tool that could be adapted for architectural research was introduced by the National Library of New Zealand. This was the online search engine Papers Past, which was first made available in 2001.[5] Papers Past provides access to millions of pages of digitised New Zealand newspapers dating back to the 1840s, and thanks to the use of optical character recognition allows researchers to search via keywords and phrases to find specific references.[6] At first the number of newspapers and geographical area they cover was small, but this has increased over time

so that now most national and many of the regional papers are now available online, although for varying time period prior to the mid-1950s.

We have used a range of search terms and innovative approaches to find the information needed to compile our data set. Of primary interest are the tender notices posted by architects; however many other articles and adverts relate to or feature information about architects. Searching ranged from broad; “tender” and “architect” (preferably with a name for the latter), through to specific searches by names and initials in the case of particular architects. Collating the results has allowed us to build a list of all the tenders placed by a particular architect as recorded in the papers.

Additional information gained relates to the types of buildings designed, their general location, approximate date of construction, and possibly the qualifications of the architect. Using this approach we have obtained information on over 25,000 New Zealand buildings and 1,500 architects or architectural firms in practice during the 1840 to 1940 period. Although the numbers appear large, they only just qualify as big data.

Big data and other research approaches

Manovich, in Gold (2012) defines big data as “data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time.” The size and inter-related nature of this information means traditional research methods and tools are unable to deal with them. New computer-based ways are needed to analyse and understand them. The big data approach, that is using a large data set to define concepts and detect otherwise less obvious trends, is well accepted in digital humanities, but has not been applied to architectural research to any great extent[7]. The main reason for this is the general lack of sufficient quantities of objective raw data needed to build up a big enough database.

Manovich (2012:466) contrasts two separate approaches to research; using either surface data (for example studying summary information concerning a huge group of people) or deep data (which would be detailed in-depth research about a few individuals or small groups). He questions the outputs of both approaches, and whether the same insights and results can be obtained from both approaches.

Traditional research methods and studies seem to have followed the deep data approach, researching a single architect or firm, or a small group of architects. This approach appears

to be well practiced, particularly in tertiary research, with post-graduate theses exploring the lives and architecture of specific architects. Some work does look at change over time, for example Peter Richardson's 1997 thesis on New Zealand government architecture[8], however these works tend to be tightly defined around a specific theme or subject.

In our research we are specifically concerned with surface data, and believe this allows us to be more objective in our analyses and outcomes, but more importantly allows us to answer broader questions about the development of the architectural profession that may not be possible were we taking a deep data approach.

We would argue our approach is quantitative rather than qualitative. This helps eliminate any subjectivity that may arise. Amassing a data set of tenders by architect is entirely objective in the raw state and is not subject to any biases of the collectors. A definition of architectural research by James Snyder is systematic research directed to the creation of knowledge. But it is accepted that all research is reductionist in some form.[9] The idea of objectivity is to keep potential bias or interference by the researcher out of the process of data arrangement and manipulation - but inevitably some generalisation will occur.

If one could collect every plan by William Gray Young, for example, one could then make definitive, objective statement about them/him - but few architects have left complete records of their works or lives. Most architectural researchers do the best they can with all the resources they can find. The end product reflects combinations of their writing skills, the time and money available for their project, the amount of material available, and their research ability - all largely subjective. Clearly an objective database should be a fundamental starting point, particularly if it can be boosted with further objective material.

Thus we have not stopped at just amassing the above-described data from Papers Past but have gone on building a more extensive dataset using additional information both obtained from Papers Past and other sources.

The wider data set

Our data is split into three related sets; building information [linked to architect], then information about the architects and architectural firms themselves. The key attributes captured are shown in Figure 1 below.

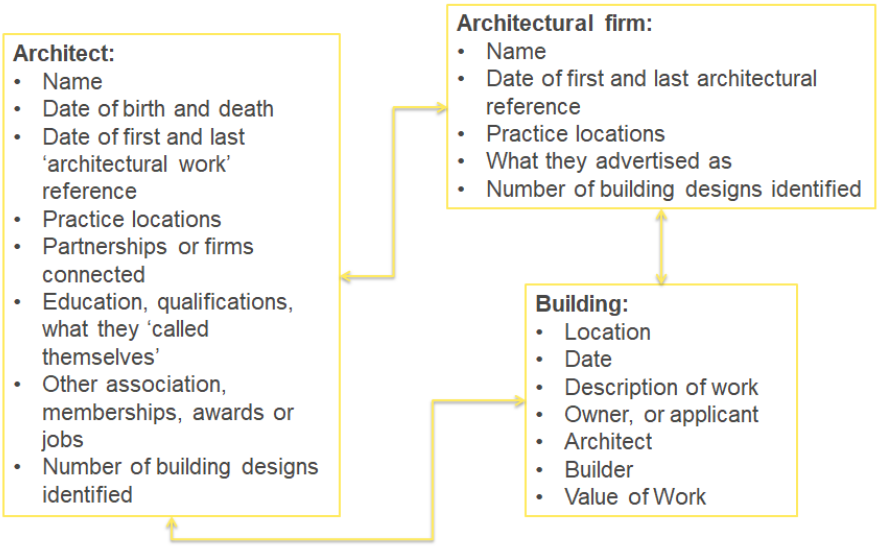

Figure one: Data attributes and relationships captured. This is the complete sets of attributes we are using, and the relationships they share.

For a number of these attributes we have had to consult sources such as the Department of Internal Affairs Births, Deaths and Marriages website, various directories and almanacs, New Zealand Institute of Architects records and local authority records. Depending on our source, not all the fields may be populated - for example newspaper tenders rarely provide the owner/builder or value of the work, whereas building permit applications generally do. Similarly it is sometimes not possible to find precise dates of birth and death for individual architects. Because of this we aim to use different subsets of data to answer different questions. Thus we may only calculate average ages using the architects we have specific birth dates for. It is hoped that as our research progresses, data capture will be increasing complete, but we have to assume currently our sample size is large enough to be representative of the whole.

We are also in the midst of moving our data to a platform that better allows the manipulation and analysis that we want to carry out. Due to the original research needs our data is assembled in Microsoft Excel, as this was more than suitable for our original purposes. However Excel has limited capability for the more complex sorting and pattern identification we now want to do; similarly it cannot easily be linked to other data sets (for example GIS or economic data) to allow us to visualise our results or extend our research. The current plan is to create an SQL database structured to allow us to easily interrogate our data, and integrate with other data sets.

One other assumption we have made is that construction of new buildings will give a suitable representation of an architect's output, so we have not been capturing every single record of alterations and additions (although we have included 'significant' alterations, such as an additional wing to an existing building, or a complete rebuild of a structure). Therefore we have assumed the number of architects who spent their careers doing alterations only was negligible.

Examples of data manipulation

One of the strengths of our approach is the ability to do analyses across the whole time period our data set covers. Rather than having to identify the year of interest then research and collate information, we can interrogate and combine our data for any time period. In the following example we have calculated some basic data about architects between 1876 and 1936, looking at 20-year intervals.

Year	Architects	Practices	Architect average age	Youngest architect	Oldest architect	Average length practice
1876	144	5	38.6	17	71	10.0
1896	259	17	43.5	19	83	14.7
1916	387	42	40.6	18	77	13.1
1936	304	38	52.8	20	89	23.6

Table one: Summary data for New Zealand architects, 1876-1936, at 20 year intervals

The figures in this table however still require extra traditional research to explain the results. For example does the average age of architects decrease as architectural courses become available locally and increasing numbers of students graduate from them, then increase as this new generation of architects settled in to long and established careers? And is there a move from small partnerships or individual architects practicing to the establishment of larger firms over time?

As well as comparative analysis we can also investigate the relationship between individual architects and architectural firms. In this case we have used Frederick de Jersey Clere as an example. The following table gives the number of designs identified either for Clere himself, or for architectural partnerships where Clere was a partner.

Architect or Partnership	Date range	Total designs
Frederick de Jersey Clere	1872-1952	167
Atkins & Clere	1883-1888	68
Clere & Richmond	1891-1895	35
Clere, Fitzgerald & Richmond	1895-1899	64
Clere & Swan	1900-1902	27
F de J Clere & Son	1911-1920	34
Clere & Busby	1913-1917	4
Clere & Williams	1918-1923	32
Clere & Clere with J F Munnings, Joint Architects [one-off design]	1923	1
Clere & Clere	1923-1962	126
Clere & Clere & Hill	1937-1939	6

Table two: Number of designs attributed to Frederick de Jersey Clere (either individually or where Clere was in partnership).

We can then take this data and analyse it over time, as shown in Figure two.

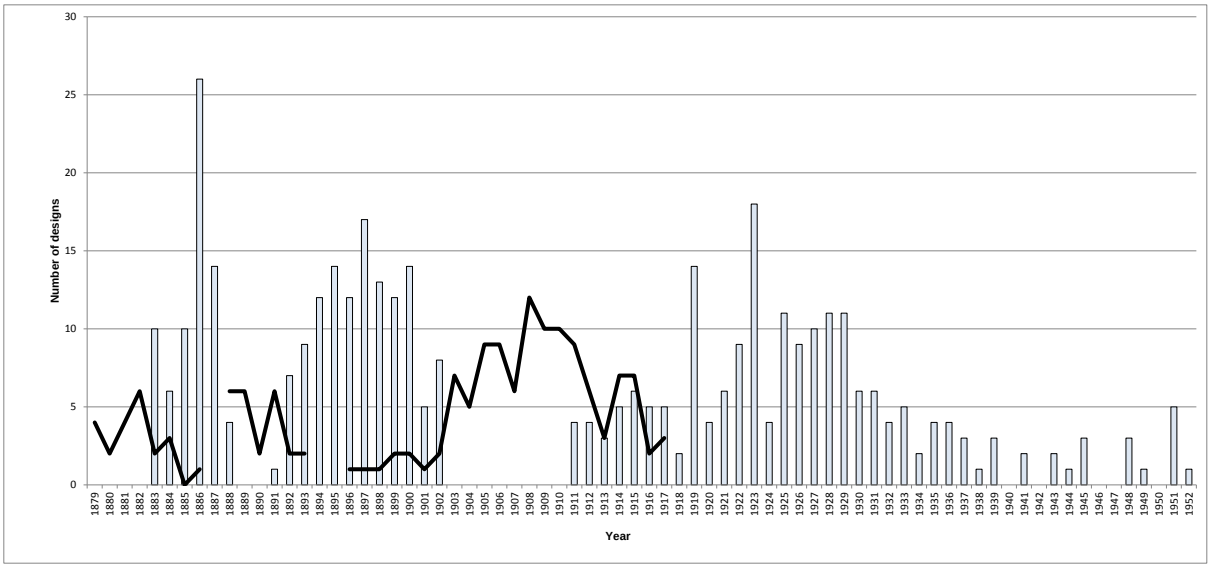

Figure two: Output for Clere over time, either tendering individually (line graph) or in partnership (bar graph). Note in particular the 1910s, where Clere was tendering individually as well as involved in multiple partnerships concurrently. Other than the early 1900s the data suggests Clere spent the majority of his career in partnership with other architects.

A consistent problem with architects in partnerships is determining which of the partners actually designed specific buildings. Two solutions are commonly adopted; either the total number of buildings is attributed to the senior partner, or the number is attributed to the

generic partnership. Ideally the plans for all buildings designed by a partnership should be individually examined for authorship but this is seldom practicable and in many instances impossible due to lack of records.

Our third example of data manipulation is comparing types of buildings designed by different architects over their careers. In this case we are comparing Clere with Herbert Thomas Barnes, an architect who practised from the 1910s, mainly in Wellington. The following two Figures look at both the annual output of designs as well as the types of buildings.

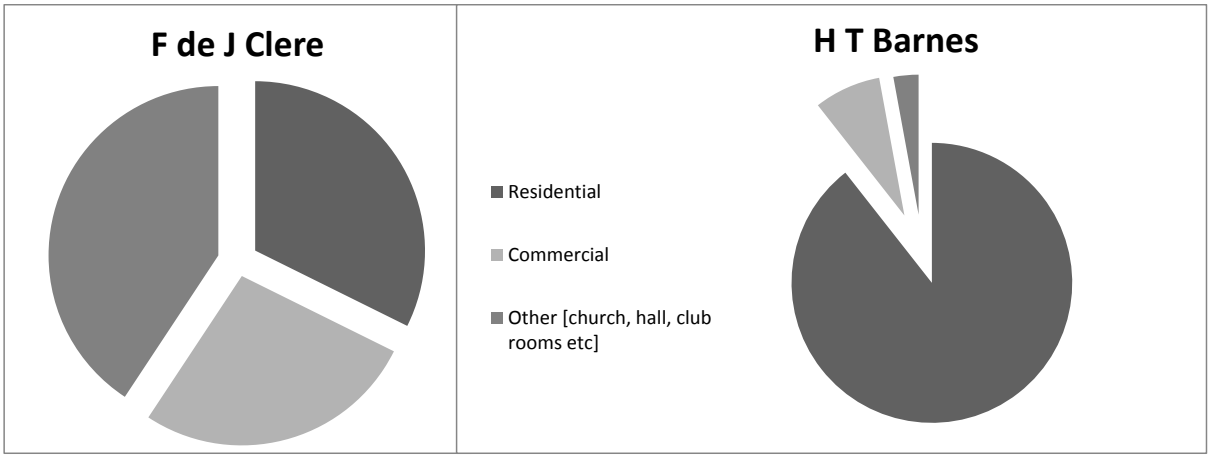

Figure three: The pie diagrams show that Clere was involved designing a broader range of buildings types, particularly church buildings, compared with the output of HT Barnes. Barnes is shown to have concentrated on the design of residential buildings.

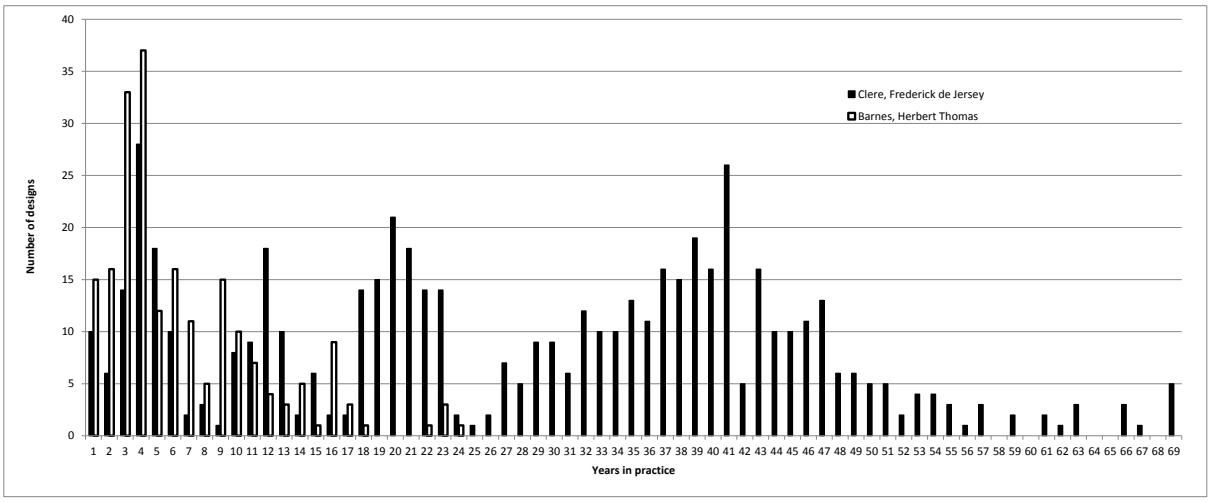

Figure four: This graph aligns the date Clere and Barnes started in practice (although Clere was active from the 1870s to1950, whereas Barnes started in the 1910s and stopped in the late 1930s). Clere's longer practice can clearly be seen; the first few years for both were busy before tailing off somewhat. Clere has regular peaks; possibly due to his involvement in partnerships?

Our final example focuses on where architects practiced, and whether this changed over time. For example does the distribution of architects correlate with population growth across the country? Does it relate to other events? For example did numbers of architects in Wellington increase once Wellington became the capital in 1865, and there was a subsequent relocation of government as well as business head offices? And how settled were architects; did they set up practice and stay in one place, or did they move around the country to find work? And did this change over time?

Table three shows the proportions of architects in practice by decade. This is a high level summary; we are hoping to also use our data to analyse at a much more specific level as well. While we have presented the data here in a table, our aim is to map data to show geographical trends.

	Distribution of architects (%)										
Decade	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940
Auckland	50.0%	44.4%	29.6%	29.0%	13.9%	17.6%	15.2%	16.3%	17.3%	22.2%	18.6%
Wellington	50.0%	44.4%	18.5%	18.8%	14.6%	24.2%	28.9%	33.0%	37.7%	42.5%	50.6%
Christchurch	0.0%	11.1%	25.9%	15.9%	17.2%	11.5%	10.2%	9.2%	9.1%	5.2%	6.4%
Dunedin	0.0%	0.0%	7.4%	11.6%	17.9%	7.9%	9.1%	3.9%	5.2%	6.3%	6.4%
Other centre	0.0%	0.0%	18.5%	24.6%	43.0%	38.8%	36.5%	37.6%	30.7%	23.8%	17.9%

Table three. Distribution of architects by decade, 1840-1940. The table shows the proportion of architects practising in the four main centres, as well as outside of those cities.

Conclusions

The current state of our project indicates that the assemblage of big data on architects and their buildings in New Zealand will provide a new, objective, kind of research tool for investigating both general and specific historical trends.

Taken over time, a portfolio of the types of buildings designed by a particular architect can be assembled with a reasonable degree of accuracy - although checking that a specific tender resulted in a completed building by other means is recommended. Thus variation in output over time can also be assessed and related to that of other architects from the same or different regions.

The work to date shows that, although tender notices provide objective data on architects and their buildings, more information from other reliable sources must be added to form a comprehensive database.

[1] Stacpoole, John (1976). *Colonial Architecture in New Zealand*. Wellington: AH & AW Reed
[2] Hodgson, Terence (1990). *Looking at the Architecture of New Zealand*. Wellington: Grantham House Publishing
[3] Shaw, Peter (2003). *A History of New Zealand Architecture*, Third Edition. Auckland: Hodder, Moa, Beckett Publishers Ltd
[4] Stacpoole, John (1971). *William Mason. The First New Zealand Architect*. Auckland: Auckland University Press
McLean, Susan (2003). *Architect of the Angels, The Churches of Frederick de Jersey Clere*. Wellington: Steele Roberts Ltd.
Lochhead, Ian (1999). *A Dream of Spires, Benjamin Mountfort and the Gothic Revivial*. Christchurch: Canterbury University press
[5] https://paperspast.natlib.govt.nz/about
[6] Papers Past provides similar access and functionality to that provided by Trove, the National Library of Australia online search; https://trove.nla.gov.au/
[7] One study we are aware of was carried out by Juan Pablo Bonta in the mid-1990s.He analysed the indexes of nearly 400 architecture books and journals, spanning 150 years, analysing the frequency and subjects of architects and articles to interpret architects' standings within the profession over that period. Bonta, Juan Pablo (1996), *American Architects and Texts*. Cambridge Ma: MIT Press.
[8] Richardson, Peter (1997), "Building the Dominion: Government Architecture in New Zealand, 1840-1922." A Thesis Submitted for the Degree of Doctor of Philosophy in Art History. Christchurch: University of Canterbury
[9] Groat, Linda and David Wang. 2013. *Architectural Research Methods*. New York: Wiley.