

Awareness and Use of the New Zealand Web Archive

A survey of New Zealand academics

Harriet Riley
Summer Research Scholar
Victoria University of Wellington

Mark Crookston
Alexander Turnbull Library
National Library of New Zealand

March 2015

Contents

1. Introduction	3
2. Definition of Terms	3
3. Summary of Findings	4
4. Survey Population and Methodology	5
5. International Context	7
5.1 Netherlands 2007	7
5.2 Portugal 2010	8
5.3 United Kingdom 2010	8
5.4 France 2012	9
6. The New Zealand Web Archive	10
7. Survey data and Analysis	11
7.1 Demographics	11
7.2 Awareness of Archived Websites	12
7.3 Use of Archived Websites for Research	13
7.4 Searching and Accessing Archived Websites	17
7.5 Using Archived Websites for Teaching	21
7.6 Content of the New Zealand Web Archive	23
7.7 The Importance of Social Media	24
7.8 Domain Harvests	28
7.9 The Research Value of Archived Websites	29
8. Conclusions	31
9. Recommendations	32
10. Bibliography	33
11. Appendices	35
Appendix 1: Survey Email	36
Appendix 2: Questionnaire	37

1. Introduction

The purpose of this research project was for the National Library of New Zealand (NLNZ) to establish a greater understanding of the use of archived websites by university academics, and to explore the future direction for web archiving in New Zealand. The Library has been archiving New Zealand websites since 1999 but to date there has been no research completed on the extent to which this collection is known and used.

The hypothesis was that there was little knowledge of the existence of web archives by university academics in New Zealand, and therefore, little use of this resource for research purposes. However, the Library was aware of several New Zealand tertiary courses that use archived websites as a learning resource.

The research sought information regarding how archived websites are used as a teaching and learning resource and what more the Library could offer New Zealand teachers at tertiary level to assist with providing access to archived websites for educational purposes.

2. Definition of Terms

Archived website

A copy of a website or blog that is preserved and made accessible in an archive. Most websites selected by NLNZ are collected regularly to capture new content. For example, the Library has taken a copy of the website *Pacifica Inc* every year since 2008 and these copies are available here:

<http://ndhadeliver.natlib.govt.nz/content-aggregator/getIEs?system=ilsdb&id=1293916>

Domain harvest

A regular harvest of the publicly available web content from the New Zealand .nz, .net, .org, and .com domains. The 2013 domain harvest collected 150 million urls and consists of 14TB of data.

3. Summary of findings

A full overview of the findings and analysis are provided in section 7 of this report. Below is a brief overview of the main findings of the research:

- The majority of respondents (211, 73%) believe that it is important for New Zealand websites and blogs to be archived.
- 147 respondents (51%) indicated that our collection of archived websites is important for their current research within the next five years.
- Many respondents demonstrated confusion between an *archived website*, and *archived items that are available via the web*. With filters applied, 21 respondents (7%) can be reliably found to be aware of and have used the New Zealand Web Archive.
- In total 128 respondents (44%) had at some point used one of the following international sources of web archives; the Internet Archive, National Library of Australia Web Archive, UK Web Archive and the US Library of Congress Web Archive.
- 88 respondents (30%) used the Internet Archive at some point for their research, including 14 (5%) using it often.
- Respondents want full text search access to both web archives and domain harvests. 210 respondents (72%) preferred access to web archives by a full text search, and 183 respondents (63%) preferred access to the domain harvests by a full text search. No one with experience using the New Zealand Web Archive disagreed with a full text search to either web archives or the domain harvests.
- URL searching was preferred by only 81 respondents (28%). It was rejected the most strongly out of the search options and notably was rejected more strongly by those who had experience searching the New Zealand Web Archive.
- 43 respondents (15%) use archived websites or blogs from international sources, as a resource in their teaching. The most popular method of using archived websites or blogs in teaching is by providing a link to the archived site; this method is used by 31 respondents (11%).
- Government websites were considered the most important subject area to be archived for the research of 191 respondents (66%).
- 114 respondents (39%) believed archives of social media would be useful to their current or future research. While 105 respondents (36%) did not know, this section engaged the most respondents with 77 adding comments to their answers.

- Video channels were considered the most important medium of social media to be archived by 166 respondents. The second most important medium of social media to be archived was discussion forums which was agreed and strongly agreed upon by 124 respondents.
- Respondents were evenly divided in their opinion of the importance of personal identity and microblogging channels to research.

4. Survey Population and Methodology

The research was undertaken in partnership between the National Library of New Zealand and Victoria University of Wellington who provided the researcher. The survey was sent to 2470 university academics of the humanities and social science disciplines, from seven New Zealand universities and one wānanga (public tertiary institution providing education in a Māori context). Other disciplines were excluded in order to keep the survey to a more manageable scope and scale.

The invitation bounced from 96 of the respondents who had previously indicated they no longer wished to receive SurveyMonkey requests. In total the survey was sent to 2374 recipients and it was completed by 290 recipients, making a 12.2% response rate. Of the 290 survey responses gathered 257 were fully completed and 33 were only partially completed. Partially completed surveys were included in the final results.

The survey population was chosen because of its likelihood of using archived websites and their relationship to the Library's resources. To allow the Library to focus on their commitment to New Zealand universities; Polytechnics and other tertiary institutions were excluded from the survey population. Unintentionally New Zealand's two other Wananga were excluded from the survey population due to the inaccessibility of their staff email addresses. Survey recipients were contacted solely by email and all email addresses were sourced from publicly available websites.

The survey consisted of twenty-four questions based on the following themes: career demographics; awareness of archived websites; use, access and value of archived websites for research and teaching; subject content of web collections; the importance of archiving social media; and domain harvests. The questions contained a mix of dichotomous and trichotomous questions, Likert scales, and free text sections which allowed respondents to elaborate on their answers.

The survey questions were constructed in consultation with Library staff and Gillian Oliver at the School of Information Management, Victoria University of Wellington. The survey was piloted amongst a select group of staff members during November 2014. The survey was distributed on Thursday 4th December using the online survey tool SurveyMonkey. The majority of university academics have finished their teaching by December, although the Library acknowledge that this is still a busy period and the distribution timing of the survey likely affected the response rate. The survey was open for six weeks, closing on the 14th of January. The survey was completed anonymously and was granted approval by the Victoria University of Wellington, School of Information Management's Human Ethics Committee.

5. International Context

The National Library of New Zealand would like to thank the international archiving community for what little has already been published on the scholarly usage of web archives and acknowledge the influence that it has had on this research. Publications on web archiving are split into either a technical archivist discussion or are projections of the generally unknown user of web archives. There is still little in the way of scholarly articles and research on the use of web archives by researchers.¹ However, the following research projects have provided some insight that was drawn upon in our research.

5.1 Netherlands 2007

The 'Web Archiving User Survey' was undertaken by Marcel Ras and Sara van Bussel at the National Library of the Netherlands (KB) in an attempt to establish the potential users of a permanent archive of selected Dutch websites.² They consulted sixteen existing web archives around the world and concluded that generally the end users of web archives fall into the following categories:

- Historians
- Sociologists
- Linguists
- Journalists
- Website owners
- Public institutions
- Members of the public.

A user test was then conducted on fifteen participants drawn from the above categories. They completed a short survey and were observed searching and accessing web archives. In this study users ranked full-text searching as the most important condition to be satisfied by a web archive, this was followed by URL searching and also the importance of making a clear distinction between the presentation of an archived website and a live website. The most important subjects for archiving mentioned by the participants in the Netherlands' user test were; news sites, weblogs, cultural websites, government websites and scholarly websites. One of the conclusions drawn from the Netherlands' research was the expectation that the main use of web archives will continue to be for research purposes.

¹ Brenda Reyes Ayala. *Web Archiving Bibliography*. Denton: UNT Libraries, 28 June 2013. p. 2

² Marcel Ras, and Sara van Bussel. *Web Archiving User Survey*. The Hague: National Library of the Netherlands (Koninklijke Bibliotheek), 2007.

5.2 Portugal 2010

The report 'Understanding the Information Needs of Web Archive Users' was undertaken by Miguel Costa and Mario Silva in 2010, for the Portuguese Web Archive (PWA).³ The aim of the report and the research was to discover the intentions of a web archive user and gather more information regarding which subjects should be archived. The research for this report was threefold and included: analysing 400 search logs; an online questionnaire, which had nineteen respondents; and a laboratory study of twenty participants. Generally their conclusions were in keeping with other trends from research on the use of web archives.

The Portuguese Web Archive can interpret full-text queries and URL searches from the same search box. Their research showed users preferred full-text over URL searches; although examination of the search logs showed that URL searches made up 20.96% of all of the searches submitted, so they are still a popular means of accessing the Portuguese Web Archive. The most frequent use of their web archive was to find a page that was already known to the user, the second was to collect information written in the past, and lastly a small number used the Web Archive to satisfy a transactional need such as downloading an old file. Interestingly, this research revealed a slight tendency for users to prefer older incidences of archived websites than newer ones, which indicates that the importance of an archived website increases with age.

5.3 United Kingdom 2010

The most similar research to the NLNZ project was completed in the UK in 2010 by the Joint Information Systems Committee (Jisc).⁴ Their extensive report entitled: 'Researcher Engagement with Web Archives' included qualitative research conducted with seventeen individuals from the web archiving community and academics of the humanities and social sciences discipline. The report was primarily focused on answering the following two questions:

- How are researchers in the humanities and social science disciplines currently making use of web archives?
- What sort of technical and policy infrastructures will researchers need in the future in order to facilitate their work?

Their report discusses the gap between the potential and the actual number of researchers using web archives, expressing the general trend of a lack of awareness about web archives. Indeed Jisc recognise that they too know little about the users of their web archive. Overall, their research raised concerns about the rapid turnover of web content and the resulting loss of web data. A concern shared amongst web archivists around the world.

³ Miguel Costa, and Mario J Silva. "Understanding the Information Needs of Web Archive Users." *Proceedings of the International Web Archiving Workshop IAWAW 2010*. Vienna, Austria, 2010. 9-16. Portuguese Web Archive (PWA), 2010.

⁴ Meghan Dougherty, Eric Meyer, Christine Madsen, Charles van den Heuvel, Arthur Thomas, and Sally Wyatt. *Researcher Engagement with Web Archives: State of the Art*. London: JISC, 2010.

Their report discusses the difficulties of deciding what to archive, and how to best provide access to the masses of information held on the internet. Firstly, decisions about what to archive have to be made. Interviewee and European Archive director, Julien Masanès, explains the impracticality of librarians and archivists replacing the publisher's filter at the magnitude of the internet. The selection process creates a tension between libraries, who want to build large multi-purpose web archives, and researchers, who want deep, project specific archives. Interviewee Dr Kristen Foot refers to this relationship and expresses that there should be some level of university commitment and discussion with researchers to support web archiving. Jisc's report also discusses the relevance and difficulty of archiving social media which is of interest to the National Library of New Zealand who is gauging the researcher interest for this data in New Zealand.

The findings and report by Jisc were important in establishing the NLNZ research into the 'Use of Web Archives by Researchers.' There has been a consistency with the findings of Jisc and the results of our research showing how few people are aware that archived websites exist. Jisc present strong evidence for the need to archive the web for academic and cultural heritage reasons and they encourage library and research communities to develop and share best practises for web archiving to increase accessibility and awareness.

5.4 France 2012

A qualitative study entitled 'Web Archives for Researchers: Representations, Expectations and Potential Uses' was undertaken in France by the Bibliothèque nationale de France (BnF) in 2012.⁵ The aim of the project was to assist planning for the future of their Internet legal deposit. There were three user groups and a total of fifteen interviews conducted. This study focused on the first user group consisting of five researchers. The intention was to analyse the subject's practice in web research and their perception of web archives. All of their five researchers recognised the value of web archiving but none had in fact used the web archives of the BnF.

The BnF's research, like that of Jisc in the United Kingdom, grappled with the debate around archiving social media. The BnF researchers cited the difficulties in determining the value of social media stems from the difficulties distinguishing between the public, or 'published' domain, and the private personas and conversations of individuals, with the latter having questionable value for archival purposes, and raising ethical questions that should be addressed before collecting.. However, the BnF also acknowledge that archives of social media could be a great resource for future researchers, but that it is impossible to predict what material will interest researchers in the future. The BnF conclude that promotion and communication of web archives is necessary to engage researchers and improve their service for the future.

⁵ Peter Stirling, Philippe Chevallier, and Gidas Illien. "Web Archives for Researchers: Representations, Expectations and Potential Uses." *The Magazine of Digital Library Research* 18, no. 3/4 (March/April 2012).

6. The New Zealand Web Archive

The New Zealand Web Archive forms part of the Alexander Turnbull Library research collections. The Library began selecting websites in 1999 and the collection has continued to grow with active development since 2005, reflecting New Zealand's growing online cultural and historical presence. The selected websites in the collection cover, over 22,000 sites at the time of writing, a diverse range of subjects and significant events, and is strong in the following areas:

- government
- politics, including blogs, and general and local body elections
- Māori, including iwi and Treaty of Waitangi
- community and ethnic groups
- music, including labels, organisations, artists and directories
- sport
- the arts
- the environment

The collection also provides a visual history of how websites change over time and most of the websites in the archive are collected at regular intervals to ensure new content is captured. This content includes web pages, images, multi-media, and publications, such as journals, that are made publicly available.

Link to more information about the New Zealand Web Archive:
<https://natlib.govt.nz/collections/a-z/new-zealand-web-archive>

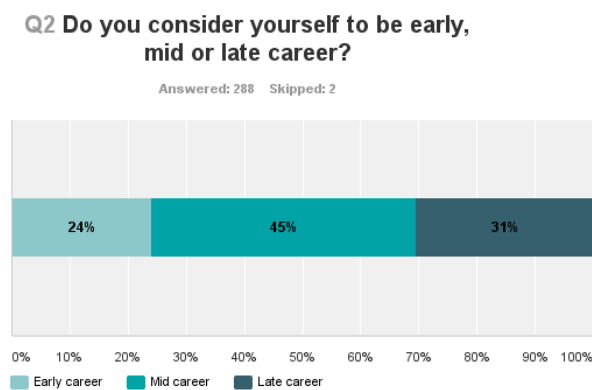
7. Survey data and analysis

All percentages given in the following discussion have been calculated from the total number of survey respondents (290) regardless of the number of respondents who answered the questions and rounded to the nearest whole number. The percentages shown in the graphs and charts have been generated from the number of respondents who answered the specific question and this varies as respondents were able to skip questions. In most cases the difference is slight.

7.1 Demographic

The purpose of these questions was to establish the respondent's subject interest and career position in order to explore trends that may develop later on in the survey. The survey was sent to disproportionately more individuals in education and this is reflected in the response rate (nearly a quarter of the total survey recipients were from education).

Figure 1: Overall respondent career stage



The highest number of respondents was from the disciplines of:

- Education (22%)
- Literature (10%)
- Economics (7%)
- History (6%)
- Other, specified by comments (15%)

The remaining respondents were scattered across disciplines. The full list can be found in appendix 2 on page 40.

The Library was curious to see whether there was a relationship between the career positioning of academics and the answers to the survey. The survey tool provided the ability to filter results based on these demographics. Figure 1 shows an even spread of career stage across the respondents. This was reflected in the results across the survey, suggesting there is little difference between awareness and use of web archives and career positioning.

7.2 Awareness of Archived Websites

International vs local awareness

The purpose of these questions was to get a general indication of the awareness levels of web archives by New Zealand researchers.

Figure 2: Awareness of international web archives

Q3 Did you know that there are international initiatives to archive copies of websites and blogs to support research?

Answered: 287 Skipped: 3

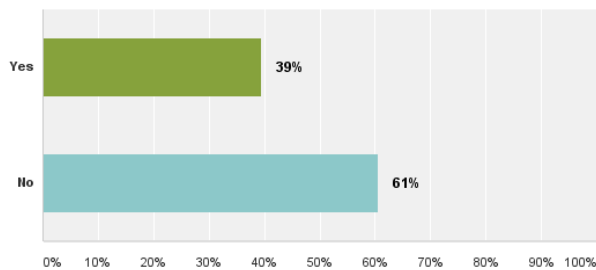
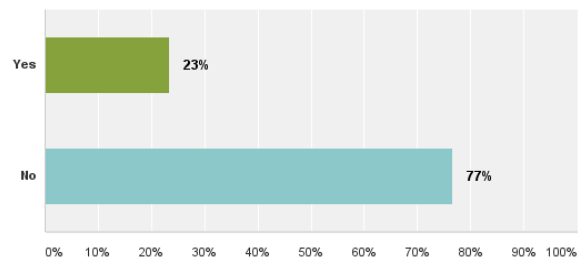


Figure 3: Awareness of local web archives

Q4 Did you know that copies of New Zealand websites and blogs are archived by NLNZ and made accessible to support research?

Answered: 286 Skipped: 4



Figures 2 and 3 show the relationship between awareness of international and New Zealand web archiving initiatives.

- 113 respondents (39%) were aware of the international initiatives concerning web archives
- 67 respondents (23%) were aware of the New Zealand Web Archive.
- 59 of the respondents (20%) were aware of both.

The gap between awareness of international and New Zealand web archiving initiatives is possibly due to a combination of factors:

- The relative high profile of the Internet Archive⁶
- NLNZ has done little to promote its web collection
- The New Zealand Web Archive is integrated into the search and discovery of all other collection items held by the Library and does not have a separate profile or branch, like the collections in Australia, the U.K., and U.S.A do.

Despite the lack of promotion or branding on the New Zealand Web Archive, it is encouraging that 23% are still aware of the collection. The 77% percent gap represents an opportunity for promotion to potential users of the web collection.

⁶ Dougherty et al. *Researcher Engagement with Web Archives: State of the Art*. p. 14

There was a slight discrepancy between the number of respondents who indicated that they were aware of international web archiving initiatives (113) and those discussed in section 7.3 who had used an international source of archived websites (128). This discrepancy is likely due to the broad wording of question three, and it does not affect the subsequent survey results.

7.3 Use of Archived Websites for Research

The Library knows that there is academic use of archived websites. Our analytics revealed that in 2013 we experienced a significant spike in access to certain archived websites. So much so that for that calendar year, three archived websites were the most highly accessed collection items in the entire Library. Further research revealed that these three archived websites were included in the reading lists for a range of university courses. The research attempted to get quantitative data on the numbers and percentages of academics that had used the New Zealand Web Archive for their research.

Some issues with the data

However, much of the data for the ‘have you used?’ questions is inconsistent and therefore inconclusive, but not all of it. Before discussing the valuable data from this section, it is first necessary to explain the inconsistencies and issues.

- 30 respondents (10%) answered yes to question 5: *‘Have you used the NLNZ website or catalogue to find archived websites?’*
- 116 respondents (40%) indicated they had used NLNZ as a source for archived websites, when answering question 6 *‘how often do you use the following sources of archived websites as part of your research?’*, and being presented with options for NLNZ, Internet Archive, National Library of Australia, U.K. Web Archive, and the U.S. Library of Congress Web Archive (figure 6).
- Comments from question 6 included references to Papers Past (New Zealand’s online newspaper archive), Google scholar and other non-web archive resources, suggesting that respondents got confused between an *archived website*, and *archived items that are available via the web*.
- The Library recognised the possibility of this confusion and provided definitions and links to the external sources within the questions to try and reinforce what an archived website is.

Possible reasons for the apparent confusion over what an archived website is, when applied to these two questions, include:

- The paucity of promotion to academics on what the New Zealand Web Archive is; how it can be accessed; and the possibilities for using archived websites as a research resource.

- The Library does not provide its web collection as a separate resource, like other countries.⁷ Rather, access to the New Zealand Web Archive is made through the catalogue search along with almost all of the other resources of the Library, via the main Library website. It is perhaps understandable that researchers are confused between archived websites and archived collections available via the website.

The difficulty in defining and explaining a web archive was also encountered by the BnF in their research in 2012 and the participants of the Jisc research in 2010. The BnF found that the confusion amongst their researchers was due to the existence of an ‘archives’ section in many websites or blogs and the history of changes on websites such as Wikipedia. This research has perhaps uncovered that there is a case for a separate ‘New Zealand Web Archive’ presence on the web, where the collection can be promoted, accessed and used.

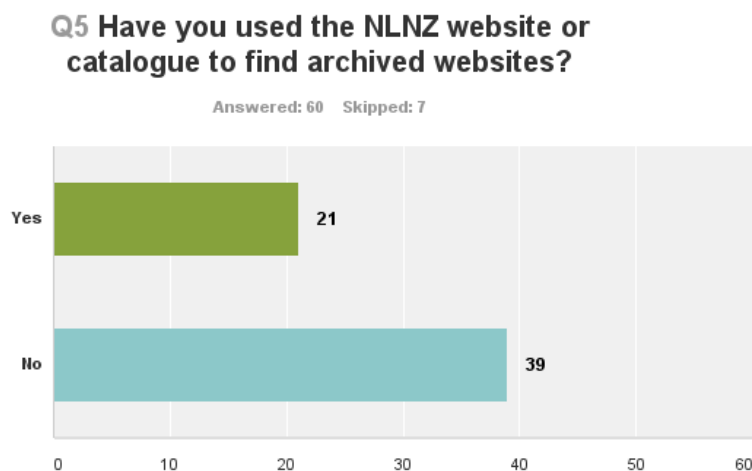
Useful data on use of the New Zealand Web Archive

Despite the inconsistencies and confusion articulated above, there is still some useful analysis and conclusions that can be made from the data.

Figure 4 is filtered to show the usage rates of the New Zealand Web Archive by those 67 respondents who had previously stated in question four, that yes they were aware of the New Zealand Web Archive. We used the filters to remove answers from respondents who appeared confused and came up with 21 respondents who knew about the New Zealand Web Archive AND had used it. This represents only 7% of the total survey population. A further 39 respondents (13%) knew about the New Zealand Web Archive but HAD NOT used it.

- Many respondents demonstrated confusion between an *archived website*, and *archived items that are available via the web*.
- With filters applied, 21 respondents (7%) can be reliably found to be aware of and have used the New Zealand Web Archive.

Figure 4: Usage of the New Zealand Web Archive by those who knew about it



⁷ See Australia: <http://pandora.nla.gov.au/>; the U.K. <http://www.webarchive.org.uk/ukwa/>; and the U.S. <http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html>

Q.4 Did you know that copies of NZ websites and blogs are archived by NLNZ and made accessible to support research? (Answered: 286 Skipped:4)	
Yes	67
No	219
Total	286

Q.5 Have you used the NLNZ website or catalogue to find archived websites? (Answered: 60 Skipped: 7)	
Yes	21
No	39
Total	60

New Zealand use of international web archives

The Library was also interested to find out the extent to which researchers use web archives outside of New Zealand. Figures 5 and 6 show an active use of web archives from sources other than the New Zealand Web Archive. The NLNZ stats for figure 5 are not useful, given the confusion over accessing archived content via the National Library's website. However, the other sources cited provide no confusion. While the Internet Archive has some additional content, the others are exclusively sources for archived websites.

- 88 respondents (30%) stated that they used the Internet Archive at some point, including 14 (5%) using it often
- 66 respondents (23%) stated that they used the National Library of Australia's Web Archive at some point
- 60 respondents (21%) said they used the US Library of Congress Web Archive
- And 44 (15%) said that they used the UK Web Archive at some point.

In total 128 respondents (44%) indicated that at some point they used one of the international web archives that we provided as options. This figure is very encouraging and gives an indication of the potential number researchers who could use the New Zealand Web Archive.

Figure 6 further demonstrates this point by showing that of the 161 respondents (56%) who have never used the NLNZ web collection; there are still a number of academics exclusively sourcing archived websites for their research, outside of New Zealand.

- 29 of these respondents had used the Internet Archive (10%)
- 16 had used the US Library of Congress Web Archive (6%)
- 11 had used the National Library of Australia (4%)
- And 7 had used the UK Web Archive (2%)

Again, there can be no confusion for this data, as they relate to sites that almost exclusively provide archived websites. These results are encouraging. It shows some level of demand and use for archived websites as a resource for research.

Figure 5: Total usage by all respondents

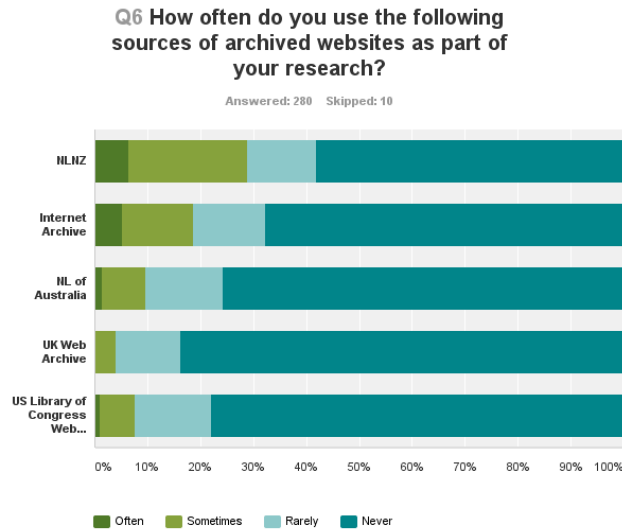
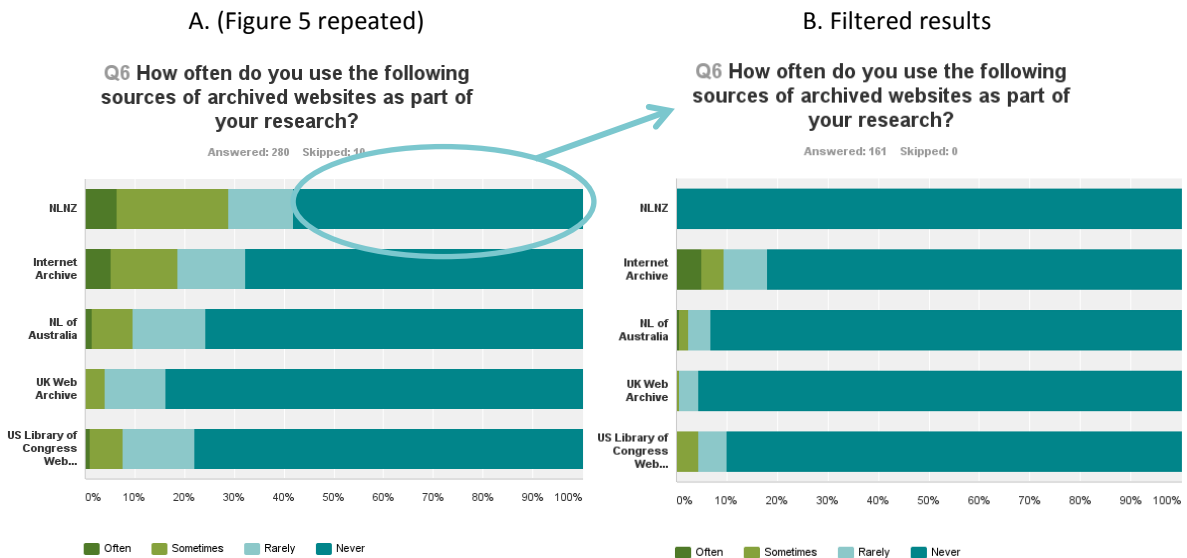


Figure 6: Filtered responses showing those who have never used the NLNZ website as a source of archived websites



What is not clear is whether academics are going to places like the Internet Archive for content they could otherwise get from the NLNZ web collection. While there is some duplication of content between NLNZ and the Internet Archive, NLNZ selective harvesting tends to be a deeper, and is therefore more likely to be a complete harvest of the website than the shallower, broader harvesting approach of the Internet Archive.

These results again point to a lack of awareness of the New Zealand Web Archive, or alternatively show that there are gaps in New Zealand Web Archive that are preventing researchers from using it.

The results also raise the question of whether there is value in combining, through aggregation or some other model, the web collections of various jurisdictions. Further research could identify the information seeking behaviour of academics looking for web archives and whether the same search is being conducted across all available web archive services.

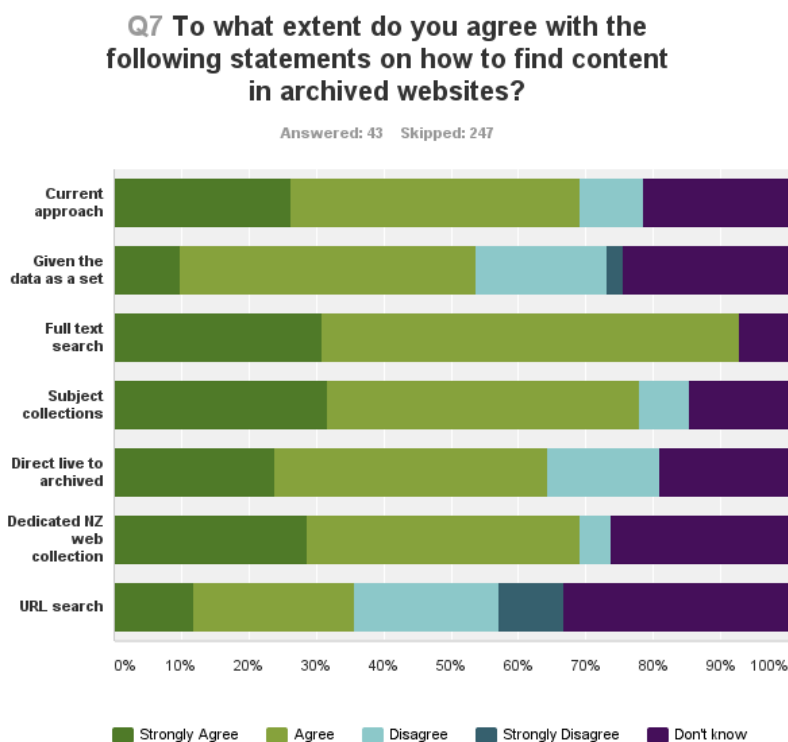
“Now I am aware of the NLNZ resource and the LOC and UK ones, I may use them, though the Internet Archive seems to go back further in time and be the most comprehensive.”

7.4 Searching and Accessing Archived Websites

This is the first piece of research investigating how researchers want to access the New Zealand Web Archive. Currently, websites selectively archived by the Library can be searched in the catalogue record by keyword, title, subject or name searches. The archived website is then viewed in a browser via a link in the catalogue record.

For this survey section the respondents were automatically separated based on their answers to question five (section 7.3) regarding whether they had previously used the Library website or catalogue to find archived websites, or not. This was in order to identify any differences between the needs of actual and potential users of the web collection. Those who had not used the web collection before were understandably more uncertain than those who had. However, on the whole the difference between the two user groups was not compelling; this is shown below in figures 7 and 8.

Figure 7: Respondents with experience using the New Zealand Web Archive

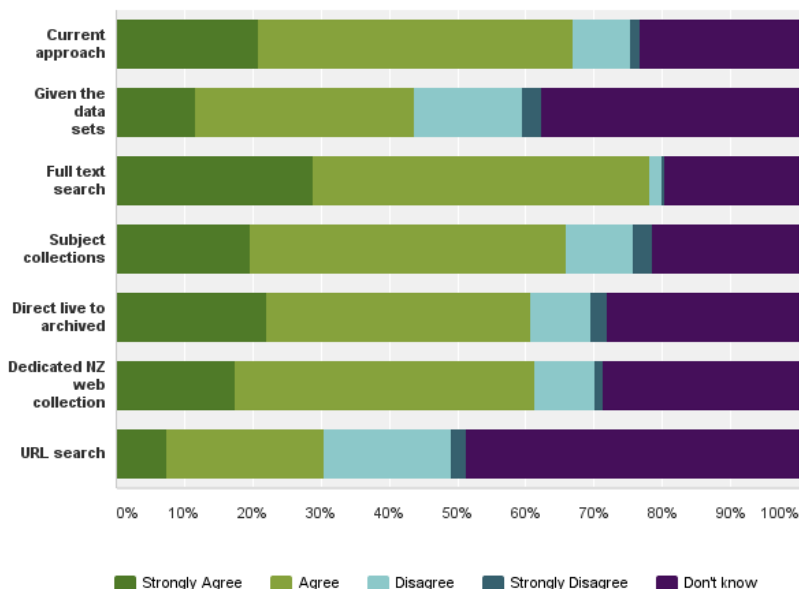


Q.7	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know	Total
I would prefer to find archived websites as part of my search for any item in the NLNZ collections (current approach)	26% 11	43% 18	10% 4	0% 0	21% 9	42
I would prefer to be given archived websites as a data set that I could search using my own research tools (e.g. data mining)	10% 4	44% 18	20% 8	2% 1	24% 10	41
I would prefer a full text search across all archived websites (e.g. like Papers Past)	31% 13	62% 26	0% 0	0% 0	7% 3	42
I would prefer subject collections of archived websites to be available for my specific research or teaching area	32% 13	46% 19	7% 3	0% 0	15% 6	41
I would prefer to go directly from a live website to archived versions of that site	24% 10	40% 17	17% 7	0% 0	19% 8	42
For items collected by NLNZ, I would prefer to find them all together in a dedicated NZ archived website collection	29% 12	40% 17	5% 2	0% 0	26% 11	42
I would prefer to use a URL search in the NLNZ catalogue to find archived websites	12% 5	24% 10	21% 9	10% 4	33% 14	42

Figure 8: Respondents who have not used the New Zealand Web Archive before

Q9 If you were to access content in archived websites at NLNZ, to what extent do you agree with the following statements?

Answered: 222 Skipped: 68



Q. 9	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't Know	Total
I would prefer to find archived websites as part of my search for any item in the NLNZ collections (current approach)	21% 45	46% 99	8% 18	1% 3	23% 50	215
I would prefer to be given archived websites as a data set that I could search using my own research tools (e.g. data mining)	12% 25	32% 69	16% 34	3% 6	38% 81	215

I would prefer a full text search across all archived websites (e.g. like Papers Past)	29% 63	49% 108	2% 4	0% 1	20% 43	219
I would prefer subject collections of archived websites to be available for my specific research or teaching area	20% 42	46% 99	10% 21	3% 6	21% 46	214
I would prefer to go directly from a live website to archived versions of that site	22% 47	39% 83	9% 19	2% 5	28% 60	214
For items collected by NLNZ, I would prefer to find them all together in a dedicated NZ archived website collection	18% 38	44% 95	9% 19	1% 3	29% 62	217
I would prefer to use a URL search in the NLNZ catalogue to find archived websites	7% 16	23% 50	19% 40	2% 5	49% 105	216

Most popular search method: Full text

Both user groups were strongly in favour of using a full text search across all archived websites.

- 171 respondents who *had not* used the New Zealand Web Archive agreed or strongly agreed to a preference for a full text search.
- 39 respondents who *had experience* using the New Zealand Web Archive agreed or strongly agree to a preference for a full text search.
- In total 72% of the survey respondents showed a preference for full text searching
- Overall, only 5 respondents disagreed with a full text search, making up 2% of the total survey population.
- No one with experience using the New Zealand Web Archive disagreed with a full text search.

The respondents also indicated a desire for full text searching later on in the survey regarding access to the domain harvests (see section 7.8). The favour of full text search access to web archives is also reflected internationally. The participants in the Netherland's study preferred full text searches and regarded searches restricted to URL alone as a limitation.⁸ While the Portuguese study held that the prevalence of Google has resulted in an expectation for the availability of full text searching on the internet.⁹

In the live web, search engines rank and display results according to specific algorithms and assumptions about the user's needs, however, a full text search in a library catalogue does not (and should not) make these assumptions. This difference can result in 'conflicting expectations' about the presentation of full text search results in a web archive.¹⁰

The findings show that there is demand for full text searching and that the current access the Library provides is limiting our researchers. This is an opportunity for the Library to modify the current catalogue record which does not reflect the wants of researchers who desire full text indexing of the website itself.

⁸ Ras and van Bussel. *Web Archiving User Survey*. p.14,18

⁹ Costa and Silva. "Understanding the Information Needs of Web Archive Users." p.9-10

¹⁰ Andrew Jackson. "Building a 'Historical Search Engine' is no easy thing." *UK Web Archive Blog*. 19 February 2015.

Least popular search method: URL search

The least popular method of finding content in archived websites is by a URL search in the Library catalogue.

- 66 respondents who *had not* used the New Zealand Web Archive agreed or strongly agreed to a preference for a URL search.
- 15 respondents who *had experience* using the New Zealand Web Archive agreed or strongly agreed to a preference for a URL search
- In total 28% of the survey respondents showed a preference for a URL search.
- URL searching also had the most disagree or strongly disagree responses answers from of all of the given options.
- Proportionally the URL search method was rejected more strongly by those who had experience searching the New Zealand Web Archive.

These finding showed similarities to the Portuguese study which also found full text searching to be preferred over URL, but with URL searching still considered a somewhat popular means of accessing their web archives. 28% of respondents are still a moderate number of researchers who would prefer this method of access. This was also the response that had the largest number of respondents who did not know. Perhaps the findings show that researchers do not want to be limited by the search and access methods, or that some are simply used to the Internet Archive which requires a URL search.

“The more ways of finding and using these resources the better.”

7.5 Using Archived Websites for Teaching

The Library's analytics show that there are several tertiary courses using archived websites as a teaching resource and, as a result, the library sought more information about how and why archived websites are being used in a learning environment. The Library also wanted to know what barriers are being faced by those wanting to use archived websites in their teaching in order to assist the use of their web collection in teaching and learning in New Zealand.

Some issues with the data

As with section 7.3, there are also some inconsistencies in the data for this question due to confusion around using the NLNZ website to access archived websites. The Library therefore decided to filter out any data concerning the New Zealand Web Archive, and make the focus of this question to focus entirely on the use of international web archives in teaching and learning in New Zealand.

- 128 respondents indicated that at some point they used any one of the international web archives that we provided as options (National Library of Australia, U.K. Web Archive, U.S. Library of Congress Web Archive and the Internet Archive).
- While the Internet Archive has some additional material the other options are exclusively sources of archived websites and therefore we are confident of the results that they show regarding the use of archived websites in teaching.

The use of international web archives in teaching in New Zealand

As previously mentioned in this report, the three highest access items of the Library's entire collections in 2013 were archived websites. Furthermore, the Library is aware that archived websites are being used in teaching in New Zealand. However even when disregarding all responses for the New Zealand Web Archive, we discovered that:

- 43 respondents have used archived websites or blogs from *international sources* as a resource in their teaching.
- This means that 34% of the respondents who use international web archives are using them as a resource in their teaching.
- In total these 43 respondents represent 15% of the total survey population.

The Library is pleased with this finding and is encouraged to see such a high level of engagement with web archives in a learning environment. The Library also was interested to find out how archived websites are used as a resource in teaching. The 43 respondents who had used international web archives in their teaching were asked two further questions regarding how and why they use this resource.

- 31 respondents said that they use archived websites in their teaching by providing a link to archived websites
- 26 respondents said they used screenshot of archived websites in their presentations

- And 16 respondents said that they include archived websites in the reading lists that they provide to their students

These findings show quantitative evidence of the use of archived websites in New Zealand classrooms. These 43 respondents were also asked why they used archived websites in their teaching.

- 30 said that they use archived websites in their teaching because they want their students to use a wide range of resources
- 24 said it was because archived websites contain the only content they can find on a topic (as the content is no longer available in print or online)

Some of the comments from this question were particularly insightful:

“Important for students to find ways to access ‘invisible’ materials, to see what has come before, and to recognise that failing to store or archive materials used in research may be affected when such pages disappear (often). Also useful to recognise sites that change or delete information that may be unflattering.”

“To show how available information changes given political and economic contexts; to demonstrate the fluidity and malleability of online content, and to encourage students to consider how information placed online is both ephemeral and long-lasting; to show how (for instance) a particular organisation might change its content to reflect popular or institutional attitudes towards things or in reaction to criticism; to show how information is sometimes ‘hidden’ or taken out of the public domain”

“Sometimes an abandoned site offers a particularly dated - but still popular – interpretation of a specific data set. This helps students hone their analytical skills by letting them sift out the nonsense.”

The answers from these 43 respondents regarding how and why they use archived websites in their teaching are enlightening and certainly of interest to the Library. The respondents express a real value in the role archived websites play in their teaching and the learning of their students.

All of the respondents were asked what the barriers and frustrations were to using archived websites in their teaching. The Library has filtered these responses to show only the responses of the 128 respondents who had used international web archives.

- Uncertainty regarding copyright and what to do with an archived website was the biggest barrier to using archived websites, it was agreed upon by 53 respondents.
- The next biggest barrier or frustration, agreed by 52 respondents was that archived websites relevant to their course are difficult or time consuming to find.
- Two comments mentioned that university processes requiring copyright clearances were a barrier to their use.

Overall these results show that there are some significant barriers to the use of archived websites in teaching and there are some improvements that could be made in delivering the service.

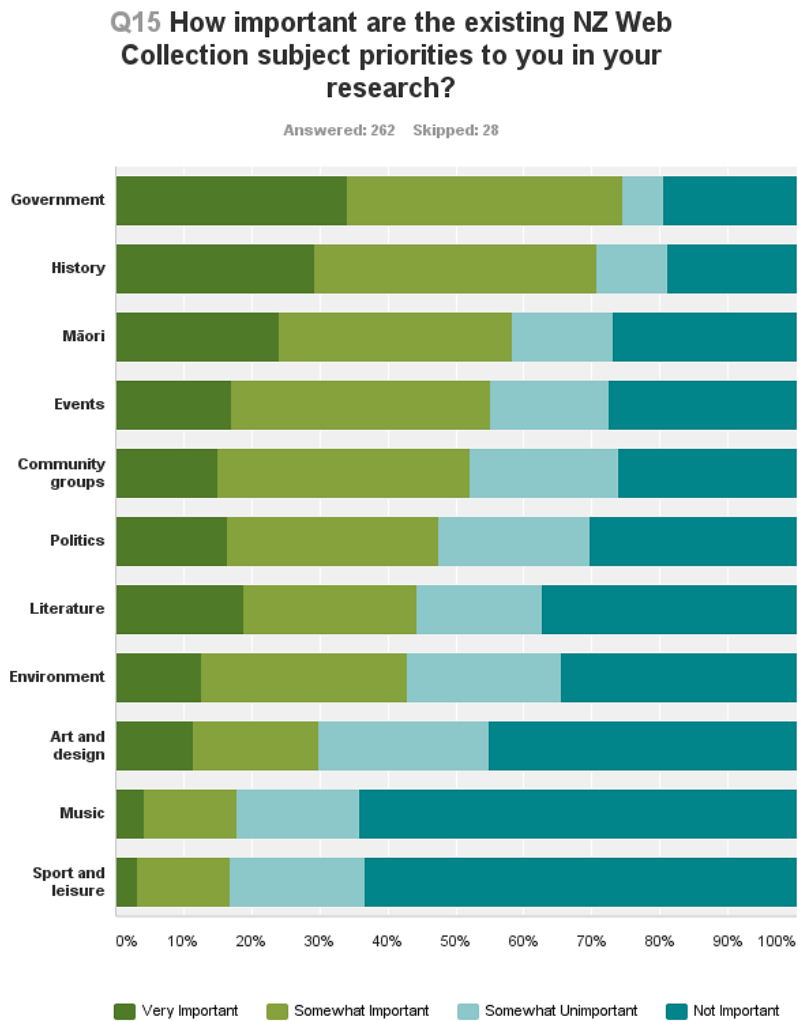
When considering that 174 of the total survey respondents had no idea of the international initiatives concerning web archives, and 219 respondents did not know that the Library archives New Zealand websites, it is clear that the biggest barrier to the use of archived websites in teaching is awareness of their existence.

The use of international web archives as a resource in teaching in New Zealand is indicative that the New Zealand Web Archive has the potential to be of as much or even more use in a New Zealand learning environment.

7.6 Content of the New Zealand Web Archive

It is currently not possible or practical to collect every website, blog or social media account relevant to New Zealand, so the Library prioritises certain subjects or themes. The purpose of this section was to seek feedback on the Library’s priorities and identify any gaps.

Figure 9: Respondent ranking of the New Zealand Web Archive subject priorities



All subject priorities received some acknowledgement of importance from respondents. Comments highlighted the potential gap in the collection to support research relating to commercial activity, e.g. online purchasing or commercial websites. The Library does collect some this web content as part of the bi-annual domain harvests (see section 7.8), however these finding suggest some demand outside of current collecting activity.

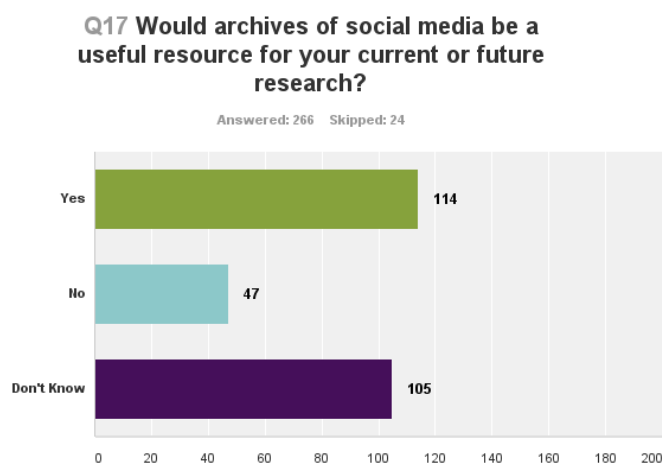
It is interesting that 191 respondents (66%) considered government archived websites either somewhat or very important to their research. Information online from government is created as a record, covered by the Public Records Act 2005, and is likely to surface in 20-25 years' time at Archives New Zealand, via the public archive process – as a record, not as a web archive. However, it is clear there is a research demand for such information now. Does this research suggest that the web is shortening expectations of availability of public information through archives? Further research and analysis is required to explore the extent to which web archives are shifting the relationship between research libraries and public archives through and what is collected when, by whom and for what purpose.

“The 2008 national led government removed a number of significant websites in their first year. These websites held research and other valuable documents that don’t exist elsewhere. Even the TEAC site was archived by the labour led government. As policy changes it is important for the government to present themselves as they want. But key decisions will be lost to the mists of time if these websites are not preserved somewhere.”

7.7 The Importance of Social Media

An increasing amount of social, political, and cultural interactions are occurring on public social media channels. The Library is currently investigating methods for collecting social media and was seeking an indication of the research value of archived social media.

Figure 10: Respondent acknowledgment of the importance of social media archives



114 respondents believed that archives of social media would be useful to their current or future research. This figure is much higher than anticipated, and provides confidence that the Library can proceed with its social media archive plans with knowledge of a known demand from the academic community.

This section overwhelmingly had the most comments of the entire survey (77) which reveals the interest in this topic by researchers. The general trend of the comments was positive and encouraging or else reflected uncertainty about the process. There were few comments suggesting that social media content is not useful for research. Representative examples include:

“Social media is increasingly a research focus and to be able to access material that is no longer available, especially on controversial topics (such as Treaty issues, racism, sexism, etc.) would be very valuable.”

“I am interested in how social media are used by Māori, and the impact of this on tradition tikanga. I am also interested in how social media is used to comment on news.”

“From my perspective it would certainly be very useful to archive the social media activities of public figures, such as politicians, ministries, political parties, lobbyists, NGOs, etc. as this forms an increasingly important part of their strategy of public communications and dissemination of information.”

“There is some potential for value here, but also great potential for a load of rubbish to be collected, and searching through that would be fraught with problems of (a) finding stuff, and (b) knowing whether it had any credibility.”

International research on archiving social media has been limited. The BnF research stated that material shared on social media “leaves the domain of publication and becomes more that of conversation,” and that it is ‘improper’ to archive much of the information held on social media as often it is the trace of actions could be easily performed in a street or shop.¹¹

It is possible that researcher awareness of the value of social media archives has changed in only three years. It is also possible that the New Zealand context, where evidence of social media interaction had significant media coverage in a recent general election, is different to the French context of 2012. Differences in contexts and methodologies make it difficult to compare researcher demand across the two studies; however the data in New Zealand reflects an encouraging demand for archives of social media as a research resource.

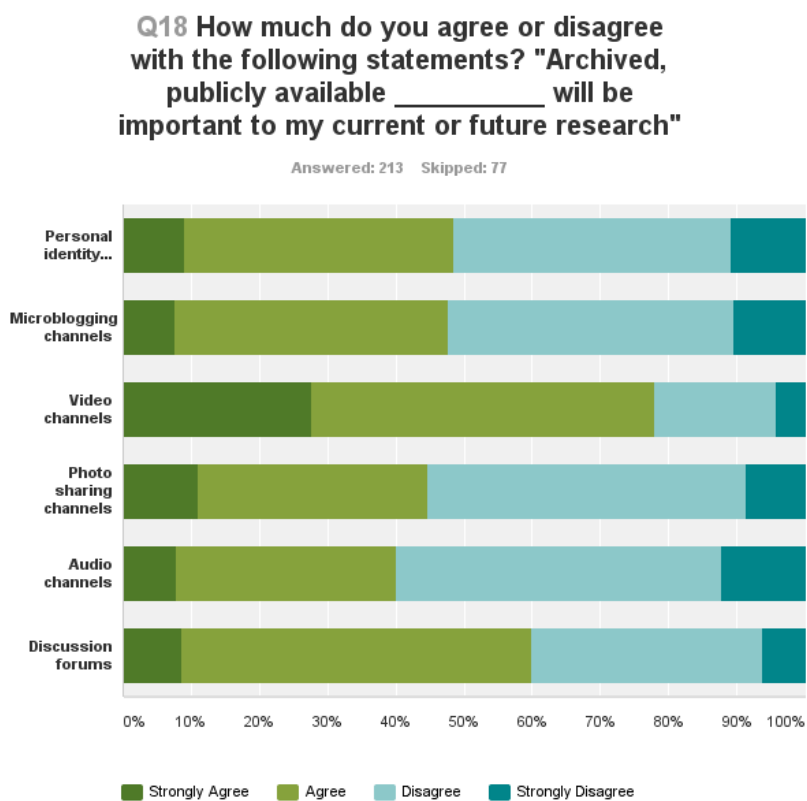
Demand for different types of social media

Respondents who answered that an archive of social media will be useful to their current or future research, and those who did not know, were prompted to an additional question on the specific types of social media that they would find valuable to be archived.

¹¹ Stirling and Chevallier. “Web Archives for Researchers: Representations, Expectations and Potential Uses.” Section 3

- Video channels were clearly considered the most important medium of social media to be archived, with 166 respondents agreeing or strongly agreeing.
- The second most important medium of social media to archive was discussion forums which was agreed and strongly agreed by 124 respondents.

Figure 11: Respondent preference for archives of specific forms of social media



Q.18	Strongly Agree	Agree	Disagree	Strongly Disagree	Total
Personal identity channels (e.g. Facebook, Myspace, LinkedIn)	9% 19	40% 84	41% 86	11% 23	212
Microblogging channels (e.g. Twitter)	8% 16	40% 84	42% 88	10% 22	210
Video channels (e.g. YouTube, Vimeo)	28% 59	50% 107	18% 38	4% 9	213
Photo sharing channels (e.g. Instagram, Flickr)	11% 23	34% 70	47% 97	9% 18	208
Audio channels (e.g. Bandcamp, Soundcloud)	8% 16	32% 66	48% 98	12% 25	205
Discussion forums (e.g. 4chan, Reddit)	9% 18	51% 106	34% 70	6% 13	207

The Library was slightly surprised that Video channels such as YouTube and Vimeo clearly outranked all other mediums in terms of importance to researchers. Collecting online video content comes with significant resource and legal challenges that can prohibit collecting, however this demand by researchers shows that the Library are justified in their attempts to work through these challenges.

Another interesting finding from this section was how highly the research community valued discussion forums. Discussion forums were ranked the second most important after video channels with 124 respondents agreeing or strongly agreeing that they will be important to their current or future research. Currently the Library is collecting discussion on sites such as theatre reviews, specialist sport and recreation communities (e.g. climbing magazines) and gamer forums. These communities are primarily online only communities and therefore, are difficult to document in any other way.

Aside from documenting communities, discussion forums offer the opportunity to document contemporary social phenomena, like trolls or shaming among others, for researchers to further understand the impact of the Internet on New Zealand culture. From a technical perspective, discussion forums are difficult to harvest due to the dynamic nature of the content. It is very encouraging to receive quantitative evidence that discussion forums are considered to be of such potential value to researchers. This is an area the Library will explore in more detail.

Responses regarding the value of micro blogging and personal identity channels were almost evenly divided between those for and those against. The Library is unsure why half the respondents were opposed to the idea of archiving micro blogs and personal identity channels. Possible reasons discussed in the research team included; little precedent of academic uses of these channels; the tools are not yet easily accessible for the computational methods required to research these sources; or maybe, contemporary uncertainty on the relationship between access to perceived personal correspondence and privacy concerns – even for publically available information - are being reflected in these answers.

Archiving social media is a new activity, underpinned by only a couple of known research projects. In their 2012 project, the BnF questioned the value of archiving social media, citing difficulty distinguishing between the public, or ‘published’ domain, and the private personas and conversations of individuals, with the latter having questionable value for archival purposes due to it resembling more of a personal conversation.¹² However, the 2010 Jisc paper promotes the value of social media archives, mentioning the U.S. Library of Congress Twitter archive and an add-on which can create a personal Facebook archive.¹³

Collecting social media is a new area, with only a limited number of known programmes in place. While the Library previously considered that archiving the social media of today was for the benefit of future researchers, there’s some evidence by our respondents to suggest that the demand is now.

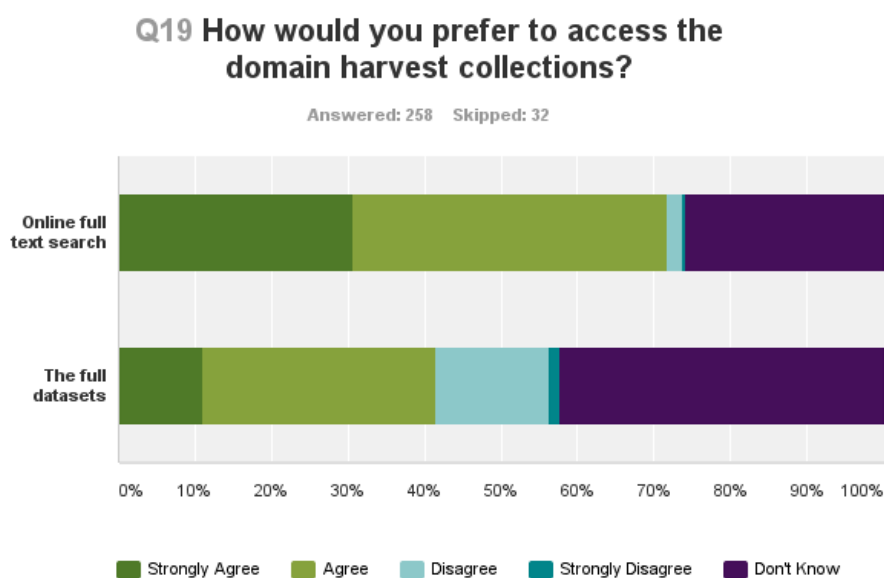
¹² Stirling and Chevallier. “Web Archives for Researchers: Representations, Expectations and Potential Uses.” Section 3

¹³ Dougherty et al. *Researcher Engagement with Web Archives: State of the Art*. p.14, 20

7.8 Domain Harvests

NLNZ regularly collects copies of the New Zealand internet from the .nz domain as part of the Whole of Domain web harvest programme. The Library has sets collected from 2008, 2010, 2013 and the 2015 set is being collected January 2015. Each set contains at least 150 million pages, and the whole collection comprises over 30TB of data, approximately 1 billion files, and over 600 million URL captures. The Library wants to provide access to these data sets online and sought information in this section that would assist this process. Generally we were pleased that so many people answered the domain harvest question considering the technicality of it.

Figure 12: Respondents preferred access to the domain harvest collections



	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't Know	Total
I would prefer access through an online full text search	31% 78	41% 105	2% 5	0% 1	26% 66	255
I would prefer to access the full datasets	11% 25	30% 69	15% 34	1% 3	42% 96	227

The popularity of full text searching was demonstrated here as well.

- 183 respondents (63%) indicated that they would prefer access to the domain harvests by a full text search
- While only 6 respondents disagreed or strongly disagreed with this method.

Full text search capability has become an established norm in research and something that academics have come to expect. This feedback expresses the view that when the Library is able to make this information available, it needs to be in a database that people can actually use.

- 96 respondents stated that they did not know if they would prefer to access the domain harvest collections by the full data sets.

This number is unsurprising given that computational analysis is new in the humanities and social science disciplines, and this question was targeted at individuals (such as those in digital humanities) who would understand the potential of access to the full datasets.

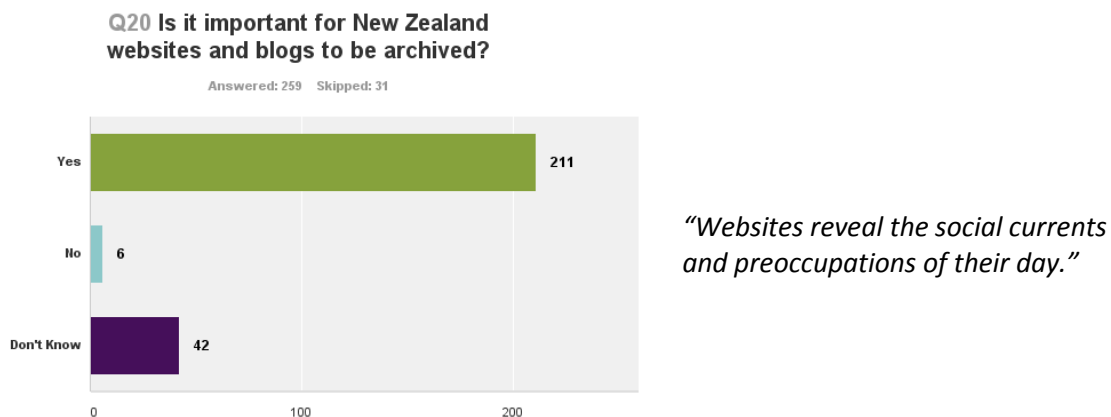
Overall, what this section showed was that the New Zealand domain harvests hold an unrealised potential for New Zealand researchers and the Library needs to put in some work to make the domain harvests accessible.

7.9 The Research Value of Archived Websites

Generally the Library archives items for long term future use. However, the results of this section show that researchers want immediate access to archived websites, blogs, social media, domain harvests and other archived web content and, researchers see this as being a valuable and important current and future resource.

- 211 respondents (73%) believe it is important for New Zealand websites and blogs to be archived.
- 6 respondents (2%) disagree
- And 42 respondents (14%) do not know

Figure 13: Acknowledging the importance of archiving New Zealand websites and blogs



Throughout the survey researchers from some disciplines (such as ancient history) noted that their subject area could not benefit from the use and availability of archived websites. The data from this question shows that the respondents value archives of New Zealand websites, regardless of whether they are personally useful for their research or not.

Our research shows that there is presently a need by New Zealand for the Library’s collection of archived websites.

- 147 respondents (51%) indicated that our collection of archived websites is important for their current research within the next five years.

This number is in contrast to our expectations and the BnF research which found that while their researchers did not have an immediate need for web archives themselves, they thought it would be of use to their students.¹⁴

Figure 14: The importance of archived websites to research over time

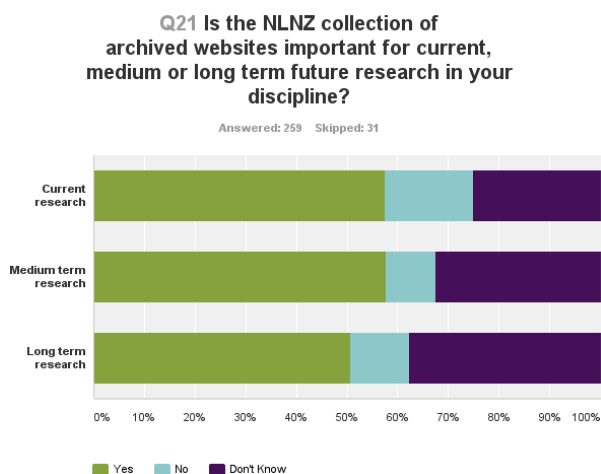
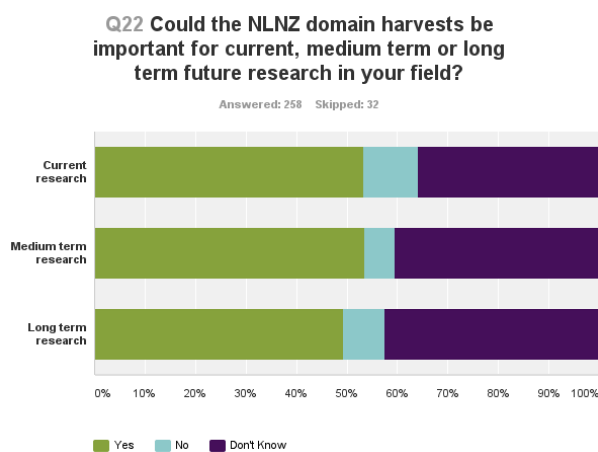


Figure 15: The importance of domain harvests to research over time



The final comments and feedback that we received at the end of the survey were largely very positive. Most respondents were grateful to have been made aware of this service and many said that they will think about using the Library’s Web Archive in the future:

“I did not know this resource existed, but I am excited to delve into it.”

“I think it’s important to get the word out to researchers that these services exist. I wasn’t aware of them prior to this survey.”

¹⁴ Stirling and Chevallier. “Web Archives for Researchers: Representations, Expectations and Potential Uses.” Section 2

8. Conclusions

This research has explored the use and awareness of web archives by New Zealand tertiary academics in the humanities and social science disciplines. New Zealand now has a baseline of data to be able to track and compare the uses and value of this unique collection over time.

The data shows that there is a large lack of awareness of web archiving amongst the majority of researchers in New Zealand's tertiary institutions. This is expressed in the results of the initial survey questions and the error respondents made interpreting the definition of a web archive.

- 23% of respondents believed they were aware of the New Zealand Web Archive.
- 7% of respondents are reliably aware of the New Zealand Web Archive.
- 39% of respondents were aware of the international initiatives regarding web archiving.

A lack of awareness was expected by the Library. These results confirm our hypothesis and show that there is an opportunity to promote the collection as a research and teaching resource.

There were, however, a small number of respondents who were very well informed about web archiving. These respondents tended to use international sources of archived websites, particularly the Internet Archive.

- 7% of respondents knew about AND had used the New Zealand Web Archive.
- 44% of respondents had used international sources of archived websites.
- 15% of respondents use archived websites as a resource in their teaching

The Library is committed to learning and development and is pleased with this finding as it shows there is a place for archived websites in the research and learning environment and bodes well for the potential future use of the New Zealand Web Archive.

The current search method for accessing the New Zealand Web Archive is not meeting the needs of New Zealand researchers. The respondent's opinions regarding searching and accessing archived websites varied little from those who had used a web archive before, and those who had not. Full text searching was by far the preferred option; this is an opportunity for the Library which does not currently offer full text searching of the New Zealand Web Archive. URL searching was the least preferred and although this is predominately used by the Internet Archive, they are signalling a move away from this means of access.

The New Zealand domain harvest holds an unrealised potential for researchers who also want access to this content via a full text search. The Library is now fully aware of the researcher desire for this information and is looking into the options for making the domain harvests available.

The majority of respondents believed archives of social media would be useful to their research at some point, or else did not know; unsurprising given this is uncharted territory. Only a small minority disagreed with the usefulness that archives of social media would be for research. Video channels and discussion forums are considered the most valuable areas of social media to be included in an archive. While archiving video content embedded on a website is currently beyond the capabilities of the Library, there is clear evidence of researcher demand for this content that the Library needs to respond to.

The lack of precedent for archives of personal identity channels and microblogs is likely a factor why these were not considered much value to researchers. Furthermore, recent cyber security attacks and online privacy concerns have likely influenced the lack of enthusiasm for this content. The Library needs to improve its social media archive capability as the data shows that the researcher demand for this information is now, rather than in the future.

Perhaps the most compelling finding is that *51% of researchers indicated that the New Zealand Web Archive will be important for their research within the next five years*. During the initial years of developing New Zealand's web archive capacity and capability (1999-2010), it was often stated that the activity was not for current researchers, but for our future generations to use as a resource to understand the present. As such, the emphasis has been placed on collecting and preserving, over improving discovery, access and usability. The results of this survey not only show that there's a demand that is current, but more is wanted and expected from this unique resource in the very near future. When it comes to the history of the New Zealand web, the academic community are telling us that the future is now.

9. Recommendations

- **Develop an active promotion strategy for the New Zealand Web Archive.** Promotion should be pushing the use of the New Zealand Web Archive for both research and teaching purposes. To date there has been no targeted promotion of the New Zealand Web Archive, and respondents have shown considerable confusion regarding access to the New Zealand Web Archive and access to other items in the National Library's collection.
- Clarity and consistency within the Library on the **definition of an archived website** compared to other archived items available online. Particular care needs to be given when distinguishing the difference between a web archive (of websites) and an archive of items that have been digitised and are available on the web. Internal clarity and consistency in terms will flow on to how we talk about the collection externally.
- **Full text search** capability of archived websites to be implemented in the New Zealand Web Archive. Full text searching was clearly the most preferred access method across all respondents for accessing both archived websites and the domain harvests.
- Provide assistance with **copyright information** for researchers wanting to use archived websites in their research and teaching. This information is currently available on the Library website but access to it could be clearer.
- **Further research** and exploration on the relationship between libraries and archives in terms of web archiving. The National Library has an existing MOU with Archives New Zealand on collecting government websites. However further research on how the different purposes of a government archive (evidence and accountability perspectives) and a national library (cultural heritage perspectives) view the uses, value and purposes of web archives could influence what is collected and how it is presented to researchers.
- Experiment with **archiving and providing access to archives of publically available social media.** There is now quantifiable evidence of a demand from the New Zealand academic community for archives of social media. Investment is required to build capacity and capability in this area. Working with collegial institutions (like the British Library, the Library of Congress and the Internet Archive), could limit the necessary investment while still achieving the desired outcomes.
- **Focus on discussion forums** as part of the Library's priorities. The Library has recently expanded its collecting of discussion forums for the online gaming community, however there's now evidence of should expand to discussion forums and discussion content for other community and social phenomena.
- The **domain harvests to be made accessible online**, with full text search capability.

- **Develop a strategy and implementation** plan for the Library to be prepared for the expected increased demand and use of the New Zealand Web Archive over the coming five years.
- **Repeat the study**, preferably through collaboration with Victoria University of Wellington, in 3-5 years.

10. Bibliography

Costa, Miguel, and Mario J Silva. "Understanding the Information Needs of Web Archive Users." *Proceedings of the International Web Archiving Workshop IAWAW 2010*. Vienna, Austria, 2010. 9-16.

Dougherty, Meghan, Eric Meyer, Christine Madsen, Charles van den Heuvel, Arthur Thomas, and Sally Wyatt. *Researcher Engagement with Web Archives: State of the Art*. London: JISC, 2010.

Jackson, Andrew. "Building a 'Historical Search Engine' is no easy thing." *UK Web Archive Blog*. 19 February 2015.
<http://britishlibrary.typepad.co.uk/webarchive/2015/02/building-a-historical-search-engine-is-no-easy-thing.html> (accessed February 24, 2015).

Ras, Marcel, and Sara van Bussel. *Web Archiving User Survey*. The Hague: National Library of the Netherlands (Koninklijke Bibliotheek), 2007.

Reyes Ayala, Brenda. *Web Archiving Bibliography*. Denton: UNT Libraries, 28 June 2013.

Stirling, Peter, Philippe Chevallier, and Gidas Illien. "Web Archives for Researchers: Representations, Expectations and Potential Uses." *The Magazine of Digital Library Research* 18, no. 3/4 (March/April 2012).

11. Appendices

Please note that the following are text only versions of the email and survey that was sent out through the platform SurveyMonkey.

Appendix 1: Email sent to survey population

Dear member of university staff,

This survey is about the role of archived websites as a resource for academic research.

The National Library of New Zealand collects NZ websites and blogs. They are archived and made accessible for research use now and in the future, long after the original website has gone or been changed.

The New Zealand Web Archive is very strong in areas of the humanities and social sciences.

This brief survey seeks to understand:

- The awareness of archived websites and blogs as a resource for the humanities and social sciences
- How and why archived websites and blogs are used or not used as part of research and teaching in the humanities and social sciences
- What is wanted from a NZ web archive service

Please assist the National Library of New Zealand by completing the survey below.

To complete the survey please click here: www.surveymonkey.com

The survey should take 5-10 minutes to complete.

The last date for receiving responses will be 16/01/2015. Completion of the survey is voluntary and you can withdraw at any time.

Submitting this survey will constitute formal consent to participate in this study.

Confidentiality

This study is being conducted according to VUW's Human Ethics Committee guidelines and has received their approval. Your confidentiality will be kept at all times. All opinions and data will be reported in an aggregated form, so that individuals will not be identified.

Please use the contact details below if you would like further information or clarification about this research.

Contact Details

Supervisors

Mark Crookston
Digital Collection Strategy Leader
Alexander Turnbull Library
National Library of New Zealand
mark.crookston@dia.govt.nz

Researcher

Harriet Riley
Summer Research Scholar
National Library of New Zealand
harriet.riley@dia.govt.nz

Dr Gillian Oliver
Senior Lecturer
School of Information management
Victoria University of Wellington
gillian.oliver@vuw.ac.nz

Thank you in advance for contributing to this research on New Zealand's Web Archive.

More information about the National Library's Web Archive can be found here: www.natlib.govt.nz/collections
If you do not wish to receive further emails from us, please click [here](#) and you will be automatically removed from our mailing list.

Appendix 2: Questionnaire

Introduction Page

Survey: Use of Archived Websites for Research

This survey is about the role of archived websites as a resource for academic research. The National Library of New Zealand collects NZ websites and blogs. They are archived and made accessible for research use now and in the future, long after the original website has gone or been changed.

The New Zealand Web Archive is very strong in areas of the humanities and social sciences.

This brief survey seeks to understand:

- The awareness of archived websites and blogs as a resource for the humanities and social sciences
- How and why archived websites and blogs are used or not used as part of research and teaching in the humanities and social sciences
- What is wanted from a NZ web archive service

Definitions of terms

Archived website: A copy of a website or blog that is preserved and made accessible in an archive. Most websites selected by NLNZ are collected regularly to capture new content. For example, NLNZ has taken a copy of the website Pacifica Inc every year since 2008 and these copies are available here.

Domain harvest: A regular harvest of the publicly available web content from the New Zealand .nz, .net, .org, and .com domains. The 2013 domain harvest collected 150 million urls and consists of 14TB of data.

NLNZ: National Library of New Zealand

Thank you in advance for contributing to this research on New Zealand's Web Collection

Harriet Riley
 Summer Research Scholar
 National Library of New Zealand
 harriet.riley@dia.govt.nz
 Introduction

Demographics

These questions allow us to explore any trends from the rest of the survey across disciplines and career positions.

1. In which discipline do you work?

- Architecture
- Art history
- Classics
- Communication
- Economics
- Education
- Environmental Studies
- Geography
- History
- Information Studies
- Language and Linguistics
- Literature
- Māori and Pacific Studies (including Te Reo)
- Media Studies
- Music
- Philosophy
- Political Science and International Relations
- Religious Studies
- Sociology
- Visual Arts
- Other (please specify)

2. Do you consider yourself to be early, mid, or late career?

- Early career
- Mid career
- Late career

Awareness of Archived Websites

A significant amount of published output and documentation of how society communicates, is on the Internet. Furthermore, research suggests the average age of a website is around 100 days, meaning they are vulnerable to loss soon after they are created. In response, the international library and archive sector collects websites. NLNZ has collected websites since 1999.

3. Did you know that there are international initiatives to archive copies of websites and blogs to support research?

- Yes
- No

4. Did you know that copies of New Zealand websites and blogs are archived by NLNZ and made accessible to support research?

- Yes
- No

Using Archived Websites for Research

Archived websites are increasingly being used in scholarly research. In 2013 the National Library noticed the three highest access items for our entire collections were archived websites. The purpose of this section is to learn about the use and sources of archived websites in New Zealand.

5. Have you used the NLNZ website or catalogue to find archived websites?

- Yes
- No

6. How often do you use the following sources of archived websites as part of your research?

(Links will open in a new window)

	Often	Sometimes	Rarely	Never
National Library of New Zealand http://natlib.govt.nz/	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Internet Archive https://archive.org/index.php	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
National Library of Australia http://pandora.nla.gov.au/col/11341	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
UK Web Archive http://www.webarchive.org.uk/ukwa/	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
US Library of Congress Web Archive http://lcweb2.loc.gov/diglib/lcwa/html/lcwa-home.html	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Searching and Accessing Archived Websites

Websites selectively archived by NLNZ can be searched in the catalogue by keyword, title, subject or name. The archived website is viewed in a browser via a link in the catalogue record.

For example, Elections New Zealand's website. Click on see this item online to see old versions of the site going back to 1999.

7. To what extent do you agree with the following statements on how to find content in archived websites?

	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
I would prefer to find archived websites as part of my search for any items in the NLNZ collections (current approach)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer to be given archived websites as a data set that I could search using my own research tools (e.g. data mining)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer a full text search across all archived websites (e.g. like Papers Past)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer subject collections of archived websites to be available for my specific research or teaching area	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer to go directly from a live website to archived versions of that site	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
For items collected by NLNZ, I would prefer to find them all together in a dedicated NZ archived website collection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer to use a url search in the NLNZ catalogue to find archived websites	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

8. Please tell us if there are any other services you would like the NLNZ to offer in regards to searching and accessing archived websites

9. If you were to access content in archived websites at NLNZ, to what extent do you agree with the following statements?

	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
I would prefer to find archived websites as part of my search for any items in the NLNZ collections (current approach)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer to be given archived websites as a data set that I could search using my own research tools (e.g. data mining)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer a full text search across all archived websites (e.g. like Papers Past)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer subject collections of archived websites to be available for my specific research or teaching area	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

I would prefer to go directly from a live website to archived versions of that site	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
For items collected by NLNZ, I would prefer to find them all together in a dedicated NZ archived website collection	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer to use a url search in the NLNZ catalogue to find archived websites	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Please tell us if there are any other services you would like the NLNZ to offer in regards to searching and accessing archived websites

Using Archived Websites for Teaching

The Library’s analytics tell us that there are several tertiary courses that use archived websites as a learning resource. The purpose of this section is to explore the relationship between teaching and the use of archived websites in New Zealand.

11. Have you used archived websites or blogs as a resource in your teaching?

- Yes
- No

12. How do you use archived websites or blogs in your teaching?

(Tick all that apply)

- I provide a link to archived websites
- I use screen shots of archived websites in my presentations
- Archived websites are part of reading lists I provide to students
- Other (please specify)

13. Why have you used archived websites or blogs as a resource in your teaching?

(Tick all that apply)

- Archived websites contain the only content I can find on a topic (content no longer available in print or online)
- I want my students to use a wide range of resources
- Archives websites reflect how society engages and communicates online, at a particular point in time
- Archived websites reflect the design sensibility of society and culture, at a particular point in time

Please explain in your own words why you have used archived websites or blogs as a resource in your teaching

14. What are some of the barriers or frustrations to using archived websites in your teaching?

(Tick all that apply)

- I am not aware of archived websites as a resource
- It is difficult to provide a link to the archived websites
- I do not know how to find archived websites
- Archived websites relevant to my course are difficult or time consuming to find
- I am unsure of the credibility or authority of archived websites as a resource
- I am unsure about copyright and what I can do with archived websites
- Other (please specify)

Content of the New Zealand Web Collection

NLNZ collects websites and blogs in a wide range of subject areas but the collection is especially strong in key areas. The purpose of this section is to seek feedback on the Library’s priorities and identify any gaps.

15. How important are the existing NZ Web Collection subject priorities to you in your research?

	Very Important	Somewhat Important	Somewhat Unimportant	Not Important
Government (e.g. official websites or government agencies)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Politics (e.g. political parties and blogs and websites from across the political spectrum)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Music (e.g. bands, musicians)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Sport and leisure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Māori (e.g. iwi, Treaty of Waitangi, language)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
History	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Environment (e.g. climate change, conservation)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Literature	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Community groups (e.g. clubs, societies, advocacy groups, migrant groups)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Art and design	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Events (e.g. Canterbury earthquakes, elections)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16. Please tell us if there are any other subject areas you would like to ensure are represented in the web history of New Zealand, and why they are important to you in your current or future research

The Importance of Social Media

A significant number of social, political, and cultural interactions take place on social media. The purpose of this section is to determine a demand from the academic community for archiving publicly available social media. Please note, the Library HAS NO INTENTION to collect private data or private messages communicated via social media, without the owner's permission.

17. Would archives of social media be a useful resource for your current or future research?

- Yes
- No
- Don't Know

Please explain why or why not

18. How much do you agree or disagree with the following statements?

	Strongly Agree	Agree	Disagree	Strongly Disagree
Archived, publicly available, personal identity channels (e.g. Facebook, Myspace, LinkedIn) will be important to my current or future research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Archived, publicly available, microblogging channels (e.g. Twitter) will be important to my current or future research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Archived, publicly available, video channels (e.g. YouTube, Vimeo) will be important to my current or future research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Archived, publicly available, photo sharing channels (e.g. Instagram, Flickr) will be important to my current or future research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Archived, publicly available, audio channels (e.g. Bandcamp, Soundcloud) will be important to my current or future research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Archived, publicly available, discussion forums (e.g. 4chan, Reddit) will be important to my current or future research	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Other, please specify

Domain Harvests

NLNZ regularly collects copies of the New Zealand internet from the .nz domain as part of the Whole of Domain web harvest programme. The Library has sets collected from 2008, 2010, 2013 and the 2014 set is being collected this summer. Each set contains at least 150 million pages, and the whole collection comprises over 30TB of data, approximately 1 billion files, and over 600 million URL captures.

19. How would you prefer to access the domain harvest collections?

	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't Know
I would prefer access through an online full text search	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I would prefer to access the full datasets	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

The Research Value of Archived Websites

The purpose of this section is to determine whether archived websites, blogs, social media, domain harvests and other web content are a valuable research resource, now or in the future.

20. Is it important for New Zealand websites and blogs to be archived?

- Yes
- No
- Don't Know

21. Is the NLNZ collection of archived websites important for current, medium or long term future research in your discipline?

	Yes	No	Don't Know
Current research (next 5 years)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Medium term research (5-15 years)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Long term research (15+ years)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

22. Could the NLNZ domain harvests be important for current, medium or long term future research in your field?

	Yes	No	Don't Know
Current research (next 5 years)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Medium term research (5-15 years)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Long term research (15+ years)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

23. Do you have any comments to make as to why it is or is not important for New Zealand websites to be archived?

Final Page

Thank you for taking this survey. The information will be used to support the National Library's approach to collecting archived websites, as well as contributing to international discourse on the scholarly use of archived websites.

The results from this survey are anonymised. However, if you would like to be contacted at some stage in the future for focus groups on the NLNZ's collection of archived websites please email me at web.archive@dia.govt.nz with your name and position.

Please note that providing this information does not compromise the confidentiality and anonymity of the research. It is impossible to link an email sent to this address to a survey response.

24. Do you have any final comments, feedback or ideas about the current or future use of archived websites?