Report: 2017 New Zealand Government Web Standards Self Assessments

20 July 2018

Prepared for Government Information Services, Department of Internal Affairs,

by Access Advisors (a Blind Foundation initiative).

Contents

1. Introduction

In December 2003, Cabinet [CAB Min (03) 41/2B] noted that the New Zealand Government had set Web Guidelines to help government agencies make online services as accessible as possible to the widest range of New Zealanders. The websites developed with these Web Guidelines would give effect to core Public Service values, and meet obligations under the Official Information Act 1992, the Human Rights Act 1993, the Policy Framework for Government-held Information, the United Nations Convention on the Rights of Persons with Disabilities, and Māori Language strategies.

In 2010, international web standards (WCAG 2.0) were adopted to drive increased conformance with the New Zealand Government's accessibility goals. By 2013, a set of New Zealand specific Web Accessibility and Web Usability Standards were mandated. All Public Service departments and Non-Public Service departments are required to assess and report on their conformance with these Standards on request. DIA administers this process, and on 29 September 2017, the mandated agencies were invited to participate in the 2017 Web Standards Self-Assessments, with the following aims:

- identify common Web Standards issues across government websites
- support the development of new guidance to help agencies meet the Web Standards
- assess the effectiveness and value of the self-assessment methodology.

This report summarises the 2017 Self-Assessment process. It provides an overview of the results submitted by agencies, identifies top issues and trends, and proposes recommendations to address some of the challenges raised. At various points throughout the report, significant findings, as well as key issues calling for solutions, are noted. These have been collated in Appendix A and Appendix B.

2. Executive Summary

The 33 Public Service departments and Non-Public Service departments subject to the Cabinet mandate to self-assess their conformance against the Web Accessibility and Web Usability Standards were invited in September 2017 to participate in the 2017 Web Standards Self-Assessment programme. DIA developed a 2017 Web Standards Self-Assessment Methodology (SAM) to facilitate this process.

Mandated agencies submitted their self-assessment results to DIA in April 2018. A review and external audit of the results submitted by agencies was undertaken with the assistance of our external third-party private sector partners, Access Advisors (a Blind Foundation initiative). According to the external audit, the agency web pages tested had an average compliance rate of 65% against the SAM. Agencies had some difficulty assessing certain tests, and their own self-assessment results were, on average, only 75% accurate, as judged by the external audit. However, the tests that were most commonly failed in the agency results were the tests most commonly failed in the external audit (see Figure 1 below).

Key findings were that:

- The Self-Assessment Methodology results do not represent the full suite of Web Standards requirements to serve as a reliable broad indicator of Web Standards conformance.
- NZ Government Web Standards capability across government is low, and much of this work is currently outsourced.
- Agency conformance was variable, however the three persistent conformance issues for agency websites were:
    1. making information in images available to people who cannot see them

2. ensuring that web pages can be used by people who rely on a keyboard instead of a mouse

3. using headings properly to structure content to make it easy for people to understand and navigate, including people who use software to support their interpretation of and interaction with web pages.

The recommendations to treat the issues identified are:

- To review and improve the NZ Web Standards and supporting SAM conformance model, in line with the new international accessibility standard (WCAG 2.1), recently published in June 2018. Options for alternative or improved methods of assessment should be investigated.

- The level of maturity and capability across government practitioners is in general low, we need to increase the knowledge and understanding of the Web Standards, self-assessment tests, and how to perform them. To inform this process of upskilling, feedback from agencies about their experience of the 2017 Self-Assessments should be sought.

- We should target our remediation efforts on addressing the top three persistent conformance issues that were identified in the 2017 results, that broadly align with the 2014 results.

A set of detailed results and recommendations is included in the report, which includes a comparison with the previous Self-Assessment results from 2014.

Figure 1. Comparison of agency and external audit compliance rates per SAM test.

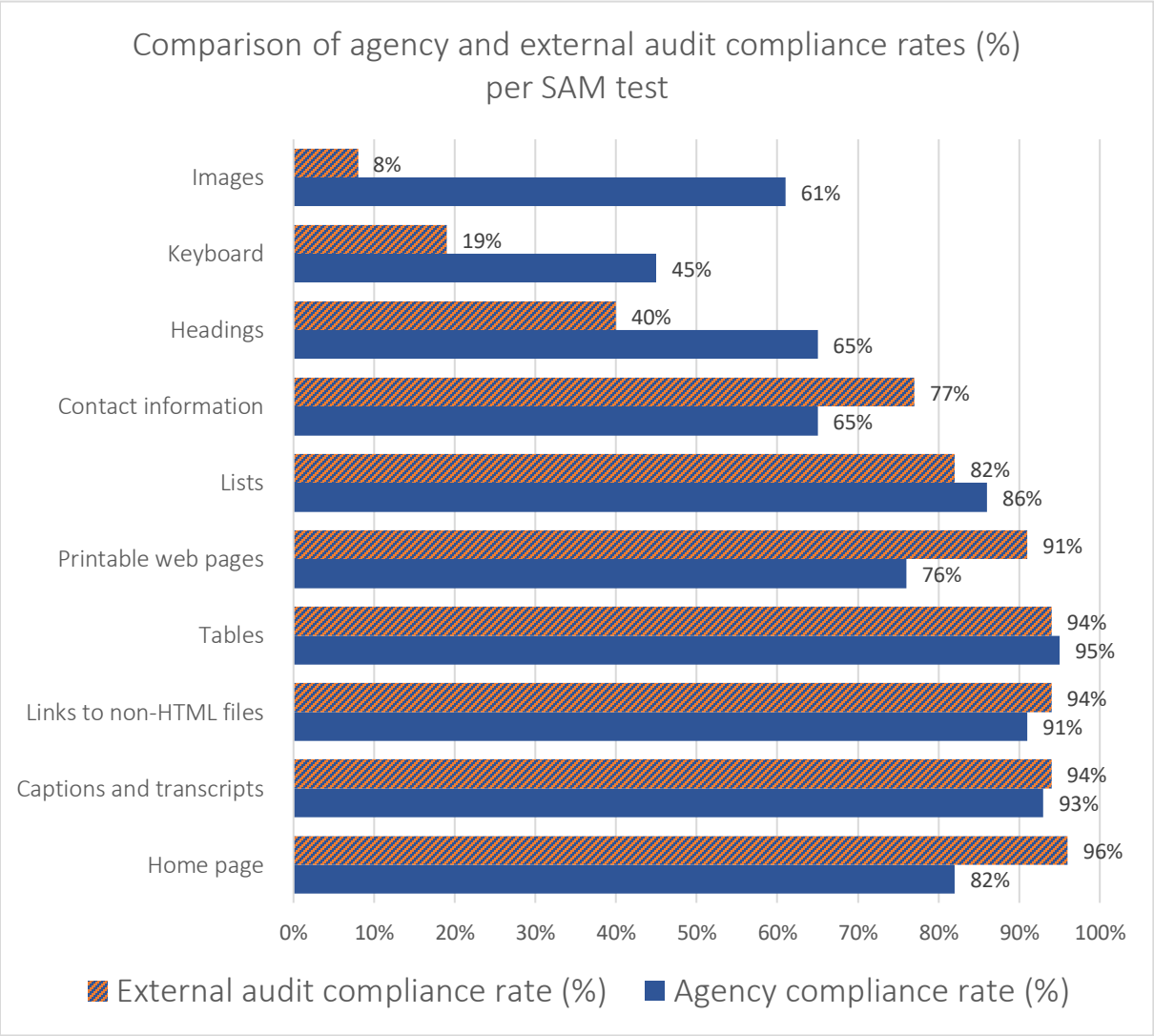


Figure 1 above compares the overall compliance rates per SAM test as assessed by agencies versus the external audit. (See Table 4 for the data represented in Figure 1), and highlights where the greatest differences in compliance scores were located.

3. Background

3.1 New Zealand Government Web Standards

The New Zealand Government Web Standards are made up of 2 separate standards, the Web Accessibility Standard and the Web Usability Standard. As established by Cabinet in 2003 [CAB Min (03) 41/2B], these Standards are mandatory for Public Service departments and Non-Public Service departments in the State Services.

The Standards set requirements for the design, development, and content of Government websites to help make them easier for the public to use. They also require that agencies, when asked, assess and report on their conformance with the Standards.

The New Zealand Government Web Standards can be found on the Web Toolkit website.

3.1.1 Web Accessibility Standard

The Web Accessibility Standard is a profile of the World Wide Web Consortium's (W3C) Web Content Accessibility Guidelines v2.0 (WCAG 2.0). With some exceptions (notably around complex images and audio description for video), it requires that each web page conform to all WCAG 2.0 Level A and Level AA requirements, or Success Criteria (SC) as they are called in the WCAG specification. WCAG 2.0 is the de facto international standard for web accessibility, and serves as the basis for the web accessibility requirements of several jurisdictions, including Australia, Canada, United Kingdom and the European Union.

3.1.2 Web Usability Standard

The Web Usability Standard sets a number of policy-related requirements to do with privacy, copyright and licensing. It also includes a small number of best practices for improving usability,

such as requiring that links to downloadable files include an indication of the file's size and format, and that each site's home page include a clear link to a page with contact information.

3.1.3 2017 Self-Assessment Methodology (SAM)

The 2017 Self-Assessment Methodology (SAM) is a collection of 10 manual tests, and 1 test using an automated tool. Developed to address the common issues identified by the 2014 Web Standards Self-Assessments, these tests were intended to highlight issues for repair and indicate how well a web page meets certain indicators of accessibility and usability (as defined by the Web Standards). The SAM was meant to be easy to use by almost anyone, without requiring advanced expertise, and reduce the cost and effort to agencies. The SAM results do not, however, sufficiently address or represent the full suite of Web Standards requirements to serve as a complete measure of Web Standards compliance; they do however set out the key issues.

3.2 Previous self-assessments

Over the years, agencies have self-assessed their websites against the NZ Government Web Standards in different ways.

3.2.1 2011 Self-Assessments

In 2011, agencies assessed only a handful of web pages from each of their websites against the Web Standards. The results from those self-assessments, along with subsequent, informal website reviews and feedback from the government web community indicated significant variability in how well agencies and their web vendors were able to implement and assess against the Web Standards. This variability in addressing the Web Standards was especially evident with regard to the accessibility-related requirements.

3.2.2 2014 Self-Assessments

In July 2013, the Web Standards were revised and split into the Web Accessibility and Web Usability Standards that are in force today. Between November 2014 and June 2015, mandated agencies participated in the 2014 Web Standards Self-Assessments, a much more comprehensive activity where each agency assessed upwards of 78 pages against close to 40 different requirements. The results from those self-assessments identified common and priority areas for improvement, and confirmed existing impressions of agencies and vendors' variable capability with respect to the Web Standards. However, compared to 2011, the 2014 Self-Assessments were a relatively costly endeavour for agencies.[1]

Taking into account what was learned from 2014, a new, simplified approach for the 2017 Self-Assessments was devised to help agencies meet their Web Standards obligations, while reducing cost and effort to them.

3.3 Comparing self-assessments over the years

The 2011, 2014, and 2017 Self-Assessments were each very different in the number of pages that were assessed, the tests that were performed, and how results were recorded. As such, it is not possible to directly compare the results from the various self-assessments for any robust measure of change or progress over time.

However, it is possible to compare the list of most common issues identified in 2014 with those found in the 2017 Self-Assessments, and note the relative prevalence of those issues. Insights from that comparison can be found in this report under Comparing 2017 results to 2014 results.

Table 1 summarises the main procedural differences between the 2014 and 2017 Self-Assessments.

[1] Report: 2014 Web Standards Self Assessments. Available at https://www.ict.govt.nz/guidance-and-resources/standards-compliance/web-standards/2014-web-standards-self-assessments/2014-web-standards-self-assessment-report/

Table 1. Summary of the main differences between the 2014 and 2017 Self-Assessments.

2017 Self-Assessments	2014 Self-Assessments
Each agency assessed a total of 20 pages, including up to 3 home pages, 3 "Contact us" pages, and pages with different content types (e.g. lists, tables, forms, images, video, date pickers, etc.).	Each agency assessed up to 5 home pages, 5 “Contact us” pages, and a maximum of 68 randomly selected pages from across all its websites.
Three views (desktop, tablet, and phone) of each page was assessed against 7 tests related to the Web Accessibility Standard, and 4 tests related to the Web Usability Standard.	Each page was assessed against each of the 37 WCAG 2.0 success criteria required by the Web Accessibility Standard, and each requirement from the Web Usability Standard.
Assessment results were recorded in a single spreadsheet. One pass/fail mark per requirement was assigned for each of the 3 views of each page, specified by URL.	Assessment results were recorded in a single spreadsheet, with one pass/fail mark per requirement for each web page, specified by URL.
Agencies were required to review their own assessment results, develop and submit an action plan report.	Agencies were required to review their own assessment results, develop and submit a risk management plan.

4. Self-Assessment process and timeline

Agencies participating in the Self-Assessments were required to deliver the following artefacts:

- results for all web pages assessed against the tests included in the 2017 Web Standards Self-Assessment Methodology (SAM)

- an action plan report outlining the agency's intentions and a timeline for addressing the issues raised by the self-assessment results, and improving the agency's overall position with regard to the Web Standards.

The SAM included 1 automated test and 10 manual tests. The automated test involved using the aXe extension for Chrome. The 10 manual tests required an assessor to review and interpret, with the aid of tools, different aspects of the web page. These two types of test complemented each other, as the automated test finds a range of WCAG 2.0 failures that the manual tests do not address at all. Of the manual tests, 6 were related to the Web Accessibility Standard, and 4 to the Web Usability Standard.

The self-assessment results (but not the action plan reports) were audited by accessibility consultants, Access Advisors, who have more than 15 years web accessibility expertise. From those external audits, the following artefacts were produced:

- assessment results for a 3 page subset of each agency's sample of web pages, with pass/fail marks for each requirement, along with automated test results, as required by the SAM.

- a list of the most common SAM failures across government websites as determined by both the agencies' self-assessments and the external audits

- comparison of agency results with those from the external audit

Details of the SAM and tests can be found in Appendix C.

4.1 Timeline

The call for the 2017 Web Standards Self-Assessments was issued 29 September 2017 to all agencies mandated to meet the Web Standards. Agencies had almost 5 months (albeit bridging the summer holiday period), until 23 February 2018, to submit their self-assessment results and action plan reports.

Not all agency submissions were received by the 23 February 2018 due date. To help agencies complete their self-assessments, three deadline extensions were eventually granted. The first extension to the end of March was formally issued on 21 March 2018.

Several agencies required more time, and on 5 April, a second extension to 30 April 2018 was communicated directly to the 8 agencies yet to submit.

Finally, an absolutely final deadline of 4 May was sent out to the remaining few agencies that missed the 30 April deadline. In response, 2 agencies confirmed that they would not be completing the 2017 Self-Assessments.

In the end, all expected self-assessment spreadsheets were delivered by 7 May 2018, with one exception, which was not submitted until 15 June 2018, too late to be included in the analysis.

4.2 Agency response

At the time of the Self-Assessments call in September 2017, there were 30 Public Service departments and 3 Non-Public Service departments in the State Services mandated by Cabinet to meet the Web Standards[2]. Each of these 33 agencies responded to the call.

[2] The Pike River Recovery Agency was established 31 January 2018, and was never included in the 2017 Self-Assessments programme.

While all mandated agencies responded to the call for self-assessment, only 30 of the mandated agencies submitted full self-assessment results and action plan reports. One agency, the New Zealand Customs Service, submitted self-assessment results, but did not submit an action plan, citing priority work on the Customs and Excise Act as the reason.

Two agencies submitted neither self-assessment results nor action plan reports. The Crown Law Office indicated that it had recently undergone operational IT restructuring and, taking the Self-Assessments as part of a broader package of ICT assurance work, would be looking to address the Web Standards following the agency's establishment of a new web platform. Ministry for the Environment did not submit any documents because of budgetary and resource constraints.

The Ministry for Pacific Peoples did submit self-assessment results only, but unfortunately those results were received too late to be included in the final analysis and review of agencies' submissions.

5. Goals of the 2017 Self-Assessments

Compared to the 2014 Web Standards Self-Assessments, the 2017 Self-Assessments were much simpler, with a new test methodology that reduced cost and effort to agencies, while delivering more practical results. The new methodology is meant to be easy to use by almost any web practitioner to test for common Web Standards issues, especially accessibility problems. It is a methodology that lends itself to formal self-assessments, as was the case here, but it can also be re-used on demand for any web page by any agency at any time.

The 2017 Web Standards Self-Assessment Methodology (SAM) represents a move away from more traditional compliance-based assessment methodologies. Such conformance-oriented approaches might clearly identify which specific requirements a web page fails to meet, but they do not necessarily translate as readily into practical or actionable results. And so the 2017 SAM does not serve to identify a web page's compliance with each of the technical requirements specified in the Web Standards. Instead, it is a relatively small collection of tests developed to address the most common issues identified by the 2014 Web Standards Self-Assessments. These tests highlight problems needing to be fixed, and indicate how well a web page meets certain indicators of accessibility and usability (as defined by the Web Standards).

The SAM itself has the following aims:

- raise staff knowledge and skill with regard to the Web Standards
- identify notable accessibility and other Web Standards issues for prioritisation and fixing by agencies
- report existing issues to management to get their support for training, remediation, resources, etc.

- test web content built by external companies for common accessibility issues
- reduce the effort involved in testing and identifying common Web Standards issues
- enable the testing process throughout a website's development lifecycle (as opposed to at the end) to ensure it is continually accessible

In addition to the above goals, the 2017 Self-Assessments aimed to:

- identify which accessibility issues covered by the SAM are most common across agency websites
- inform the development of future guidance and support for agencies and practitioners delivering government information and services on the web
- test the new Methodology, for its practicality and effectiveness in identifying common Web Standards issues with NZ Government websites, and for its viability as an indicator of or proxy for accessibility generally, and WCAG 2.0 conformance in particular.

6. Analysis and results

6.1 Summary

Once agencies submitted their results, those results were externally audited by accessibility consultants. A review of agencies' results alongside those from the external audit highlights which SAM tests were failed by NZ Government web pages most frequently, and which SAM tests agencies had the most difficulty assessing accurately. Based on this audit and review, agency web pages have an average compliance rate of 65% against the SAM.

Agencies' self-assessment results were, on average, only 75% accurate. Despite this, the SAM tests most commonly failed in the agency results were confirmed by the external audit results. Those tests were:

- the **Images** test, which looked for image content that did not have a proper text equivalent for people who cannot see the image for whatever reason
- the **Keyboard** test, which checked that all functionality worked via the keyboard, and that interactive elements had a visible indication when they were in focus
- the **Headings** test, which checked that content presented as a heading (e.g. bigger and bolder) had the proper HTML markup to programmatically identify it as a heading, and vice versa, that content marked up as a heading actually served as a heading to the content that followed it.

As was noted from the 2014 Self-Assessments, those tests or requirements with the least compliance also tended to be the least accurately assessed. This correlation indicates that the less one understands a requirement, the less likely they are to accurately assess for it, and the less likely they are to develop or design a web page that meets it.

According to the more accurate external audit results, the Images test was passed only 8% of the time. The Keyboard test had an average compliance rate of 19%, and the Headings test a rate of 40%. This suggests that NZ Government web pages will present accessibility issues for people who:

- cannot see image content (whether they are vision impaired or, for example, have not downloaded web page images in order to save bandwidth on a dial-up or mobile connection)
- cannot or prefer not to use a mouse or other pointer device, and instead rely on a keyboard to navigate and interact with web content
- rely on the HTML markup to expose, via special software (e.g. screen reader), the heading structure or hierarchy on the page in order to understand and navigate its content.

Each of the SAM test results were mapped to their relevant WCAG 2.0 Success Criteria (SC). In some cases, notably the Keyboard, as well as the Captions and Transcripts, a single test could reveal issues related to different WCAG SC. The assessor notes associated with these individual test results were reviewed, and where appropriate, a single Fail result was expanded into several issues, each representing a discrete WCAG SC failure.

This approach established a more granular picture of web page compliance with individual WCAG SC, providing more detailed information about specific causes underlying test failures. This enhanced detail also enabled a more complete comparison of the SAM with a full WCAG 2.0 audit, which was performed on a subset of 10 pages. In the end, however, it was concluded that the SAM simply does not check a sufficiently broad number of WCAG SC or causes of accessibility issues to serve as a reliable proxy for or indicator of accessibility, as defined by the Web Accessibility Standard.

6.2 Agency compliance scores

Individual agencies' compliance as measured by the SAM was not an explicitly requested measure. For this reason, agency compliance scores are not included in this report. However, such a measure is possible to extract from the dataset produced from agency's self-assessment results.

6.3 Action Plan reports

Action plans were not reviewed as part of this analysis. The action plan reports that agencies submitted remain with the Department of Internal Affairs for internal use.

However, the reports submitted ranged from formal plans approved by the agency's Chief Information Officer, to a few sentences via email regarding general plans to address the findings and renew efforts to meet the Web Standards.

6.4 Data quality

6.4.1 Inconsistent data entry

There were many empty results in agencies' submitted spreadsheets. In a good number of cases, these were the results for a page's Tablet and Phone viewports. Where there was no difference in the results from one viewport to the other, assessors were instructed to enter "No change". For example, if a page failed a certain test at the Desktop viewport, and the result was the same for the Tablet viewport, the assessor was expected to enter "Fail" in the Result column for Tablet, and "No change" in the Notes column. Similarly when moving from the Tablet to the Phone viewport.

The SAM instructions were not overly detailed around this procedure, and in some cases, the results for Tablet and Phone were left blank. In those instances, unless the blank Result value was accompanied by a "No change" (or similar, e.g. "Same as above") in the Notes, and also preceded by a "Fail" at the previous viewport, there was little choice but to interpret and record the blank as a Pass. However, the overall breakdown of Pass and Fail results across these three viewports for the subset of 3 pages included in the external audit showed no significant difference.

For many of the SAM manual tests, if the test was not applicable (because the particular element to be checked was not present on the page), assessors were to leave the Result column blank, and record "N/A" in the Notes column. For instance, if conducting the Tables test on a page with no tables, the result recorded in the spreadsheet was expected to be blank, with a note of "N/A". Unfortunately, assessors did not consistently record such results this way, and instead entered notes such as "No tables", "none on page", or in some cases, "Pass". This variability made it difficult to reliably infer any trends from the difference between a blank result associated with an "N/A" and an actual "Pass".

In the case of one agency, the assessors did not record repeated "Fail" results at every viewport on every page that had common template-level failures affecting every page on the site.

Instead, if the page had no other failures for the relevant SAM test, they left the Result field blank, so there was no explicit record of the page failing for those template-level issues. Without redoing the agency's self-assessment, such blanks were inferred to be "Pass" scores. Potentially, then, this interpretation of agency results may have resulted in a marginally higher level of Pass scores.

Key issue #1

There was significant variability in how agencies followed the Self-Assessment process and recorded their manual test results in the Self-Assessment results spreadsheet. This complicated the initial data, which required substantial effort to normalise, and forced some interpretations of the results.

Recommendation #1

To help reduce the variability in how agencies performed the SAM manual tests and recorded the results, it is recommended that assessors be asked simple yes/no or pass/fail questions, and that some mechanisms be established to ensure data consistency. For example, if self-assessment remains the approach taken for measuring Web Standards performance, provide a tool, ideally online (perhaps something like a survey), that restricts the answers that can be recorded. More detailed step-by-step instructions for tests with a more granular focus could also help, to avoid multiple failure issues being recorded under a single test result.

If agency web practitioners and others will be expected to continue performing this type of assessment, there will be a balance to strike between enough detailed instruction for clarity and ease, and not overwhelming assessors who are neither technical web nor accessibility experts.

6.4.2 Cumbersome procedure for recording aXe results

Based on agencies' submissions, it appears that agency assessors did not consistently follow the instructions for using aXe and for recording its automated test results in the spreadsheet. The

process for copying and pasting the aXe results did involve a somewhat awkward and repetitive text selection procedure that was prone to error. As such, many automated aXe results included not only clear violations as expected, but also issues that need review. Cleaning these automated results and separating out the violations from those issues that merely need review was not in scope for this project.

Key issue #2

The results from the aXe tool that were saved in the spreadsheet were prey to inconsistencies that could lead to incorrect data being recorded. This was due to the rather tricky copy and paste procedure required to select and save the aXe results.

Recommendation #2

A centralised approach to the automated testing, where one agency performs automated tests across the entire population of pages to be assessed would reduce the burden on agencies for performing what is essentially a machine-based test that should not and does not require manual activity by individual assessors at every agency.

6.5 Self-assessment Methodology (SAM) scores

6.5.1 Agency manual test results

The Accessibility tests, and 1 of the Usability tests (Links to non-HTML files) were applied to all 3 viewports (Desktop, Tablet, and Phone) for each page tested. This would equate to a total of 1,800 results (pass and fail) for each of these tests (30 agencies × 20 pages × 3 viewports = 1,800). However, one agency's submission was missing values for a single page (all 3 viewports), which means there were only 1,797 actual results for each of these SAM tests

Depending on the agency, anywhere from 1 to 7 home pages were included in an agency's self-assessment. A total of 70 home pages were assessed, which equates to a total of 210 results for that test across all 30 agencies (70 home pages × 3 viewports = 210).

Again, depending on the agency, anywhere from 1 to 5 "Contact us" pages were included in its self-assessment. A total of 62 "Contact us" pages were assessed, which makes a total of 186 results for that test (62 "Contact us" pages × 3 viewports = 186).

Table 2 below shows the number of fails recorded for each of the 10 SAM manual tests as assessed by the agencies. These include all the manual tests associated with both the Web Accessibility and the Web Usability Standards, but exclude the automated (aXe) test results. The failures are also represented as a percentage compliance rate, where 100% is full compliance.

Table 2. Average compliance rates as per agency self-assessment results.

Category	SAM test	Total results	Number of fails	Average compliance rate (%)
Accessibility	**Keyboard**	1,797	993	45%

Category	SAM test	Total results	Number of fails	Average compliance rate (%)
Accessibility	**Images**	1,797	701	61%
Accessibility	**Headings**	1,797	633	65%
Usability	**Contact information**	186	66	65%
Usability	**Printable web pages**	600	145	76%
Usability	**Home page**	210	38	82%
Accessibility	**Lists**	1,797	260	86%
Usability	**Links to non-HTML files**	1,797	162	91%
Accessibility	**Captions and transcripts**	1,797	123	93%
Accessibility	**Tables**	1,797	98	95%

Finding #1

Overall, according to agencies' self-assessment results, the NZ Government web pages assessed have an average compliance rate of 76% against the SAM. Correlatively, the web pages had an average failure rate of 24%. Compare with Finding #2 below regarding the average compliance rate as determined by the external audit results.

6.5.2 External manual test results

Agencies' internal self-assessment results were externally audited by Access Advisors. This external audit involved 3 pages (1 home page, 1 "Contact us" page, and 1 other page) from each agency's self-assessment being tested by accessibility experts against the SAM.

Consequently, the external audits involved testing a smaller number of pages than agencies tested. For each of the manual tests that apply to all 3 viewports, which is all the accessibility and 1 of the usability tests, there were 270 pass/fail results (30 agencies × 3 pages × 3 viewports = 270). For the 3 usability tests that apply only once to each page, there were only 90 total results (30 agencies × 3 pages = 90).

Table 3. Average compliance rates as per external audit results.

Category	SAM test	Total results	Number of fails	Average compliance rate (%)
Accessibility	**Images**	270	248	8%
Accessibility	**Keyboard**	270	219	19%
Accessibility	**Headings**	270	163	40%
Usability	**Contact information**	90	21	77%
Accessibility	**Lists**	270	48	82%
Usability	**Printable web pages**	90	8	91%
Accessibility	**Captions and transcripts**	270	15	94%
Accessibility	**Tables**	270	15	94%
Usability	**Links to non-HTML files**	270	15	94%
Usability	**Home page**	90	4	96%

Table 3 above shows the number of fails for each of the 10 SAM manual tests as applied to these pages by Access Advisors. Just as with agencies' own self-assessment results, these results include all the manual tests associated with both the Web Accessibility and the Web

Usability Standards, but exclude the automated (aXe) test results. The failures are also represented as a percentage compliance rate, where 100% is full compliance.

Finding #2

The external audit results established a 65% average compliance rate against the SAM. This represents an average failure rate of 35%, which is 11 percentage points greater than the failure rate measured by agencies. This has implications regarding the accuracy of agencies' own self-assessment results. See Key issue and Recommendation #3.

6.5.3 Comparing agency and external average compliance rates

While the population of pages externally audited was much smaller than that assessed by agencies, the external audit results are assumed to be more accurate, given the relative expertise of the external auditors. Accordingly, despite the smaller sample population of pages, the external audit results are considered to provide a closer representation of the current state of Web Standards implementation across NZ Government websites.

Except for some differences with respect to position, the 3 SAM tests with the lowest rates of compliance are the same for both agencies' self-assessment results and the external audits. Those are the Images, Keyboard, and Headings tests. See Table 4 below.

Table 4. Comparison of average compliance rates as recorded by agencies versus those from the external audit.

Category	SAM test	Agency compliance rate (%)	External audit compliance rate (%)
Accessibility	**Images**	61%	8%
Accessibility	**Keyboard**	45%	19%

Category	SAM test	Agency compliance rate (%)	External audit compliance rate (%)
Accessibility	**Headings**	65%	40%
Usability	**Contact information**	65%	77%
Accessibility	**Lists**	86%	82%
Usability	**Printable web pages**	76%	91%
Accessibility	**Captions and transcripts**	93%	94%
Usability	**Links to non-HTML files**	91%	94%
Accessibility	**Tables**	95%	94%
Usability	**Home page**	82%	96%

Despite the difference in number of fail results, the overall pattern of compliance derived from agency and external audit results is similar.

Finding #3

Despite the difference in expertise, agency results do conform in their general portrait of compliance to the external audit results, and that the 3 tests (Images, Keyboard, Headings) with the lowest compliance rates represent the tests that were least well-handled or most difficult to perform by agency assessors, and the areas where NZ Government websites commonly fail to meet Web Standards requirements. See related Key issue and Recommendation #3 below.

6.5.4 The difference between agency and external manual results

When the external audit scores for the SAM manual tests were compared with the agencies' results, it was noted where a pass was recorded by the agency, but a fail was recorded by the external audit, and *vice versa*.

By adding up these changes in the results we can get a score representing the average difference or variance between the agency self-assessments and the external audit scores for each manual test in the SAM. This difference or variance can be interpreted as an indication of how accurately web pages were assessed against a specific SAM test or requirement. In other words, the greater the variance, the less the SAM test was accurately assessed.

Table 5 below lists the average variance between the agencies' own scores and the external audit scores. The variance is expressed as a percentage, where 100% would indicate that the external audit recorded a different score for every result recorded by the agency.

Table 5. Average variance between agency and external audit SAM manual scores.

Category	SAM test	Number of Results	Changed to Fail	Changed to Pass	Total Changed Results (%)
Accessibility	Images	270	139	1	52%
Accessibility	Keyboard	270	93	13	39%
Accessibility	Headings	270	89	13	38%
Usability	Contact information	90	6	11	19%
Accessibility	Lists	270	30	51	30%

Category	SAM test	Number of Results	Changed to Fail	Changed to Pass	Total Changed Results (%)
Usability	Printable web pages	90	4	18	24%
Accessibility	Captions and transcripts	270	0	0	0%[3]
Usability	Links to non-HTML files	270	9	35	16%
Accessibility	Tables	270	12	11	9%
Usability	Home page	90	1	2	3%

Figure 2 below serves merely to emphasise, consistent with the different average compliance rates for agency vs. external audit results, that the SAM Images, Keyboard, and Headings tests were the most problematic for agencies.

Of particular note is the high number of changes from a Pass to a Fail for those 3 tests. This could signal that agencies had difficulty identifying actual failures for these tests.

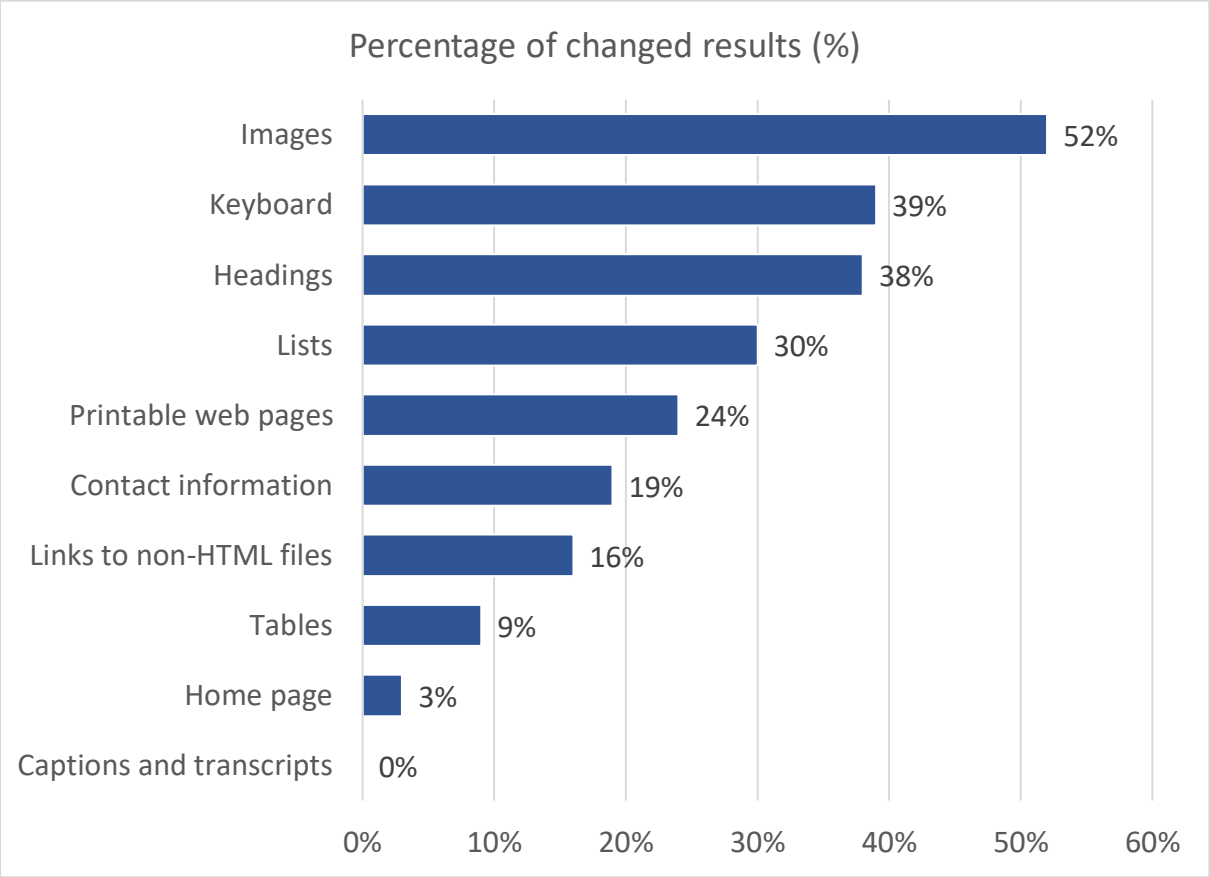
Meanwhile, there is a relatively high number of changes from a Fail result to a Pass for the Lists, Contact information, Printable web pages, and Links to non-HTML files tests. These 4 tests are arguably much simpler tests to perform than the others, which suggests that the instructions for performing these tests were somehow lacking sufficient detail or direction for agencies to reliably record an appropriate result.

Interestingly, there are relatively similar numbers for changes to Pass and changes to Fail for the Tables test, which would confirm that the test proved difficult to carry out with any

[3] Due to rounding, values below 0.5% are represented as 0%.

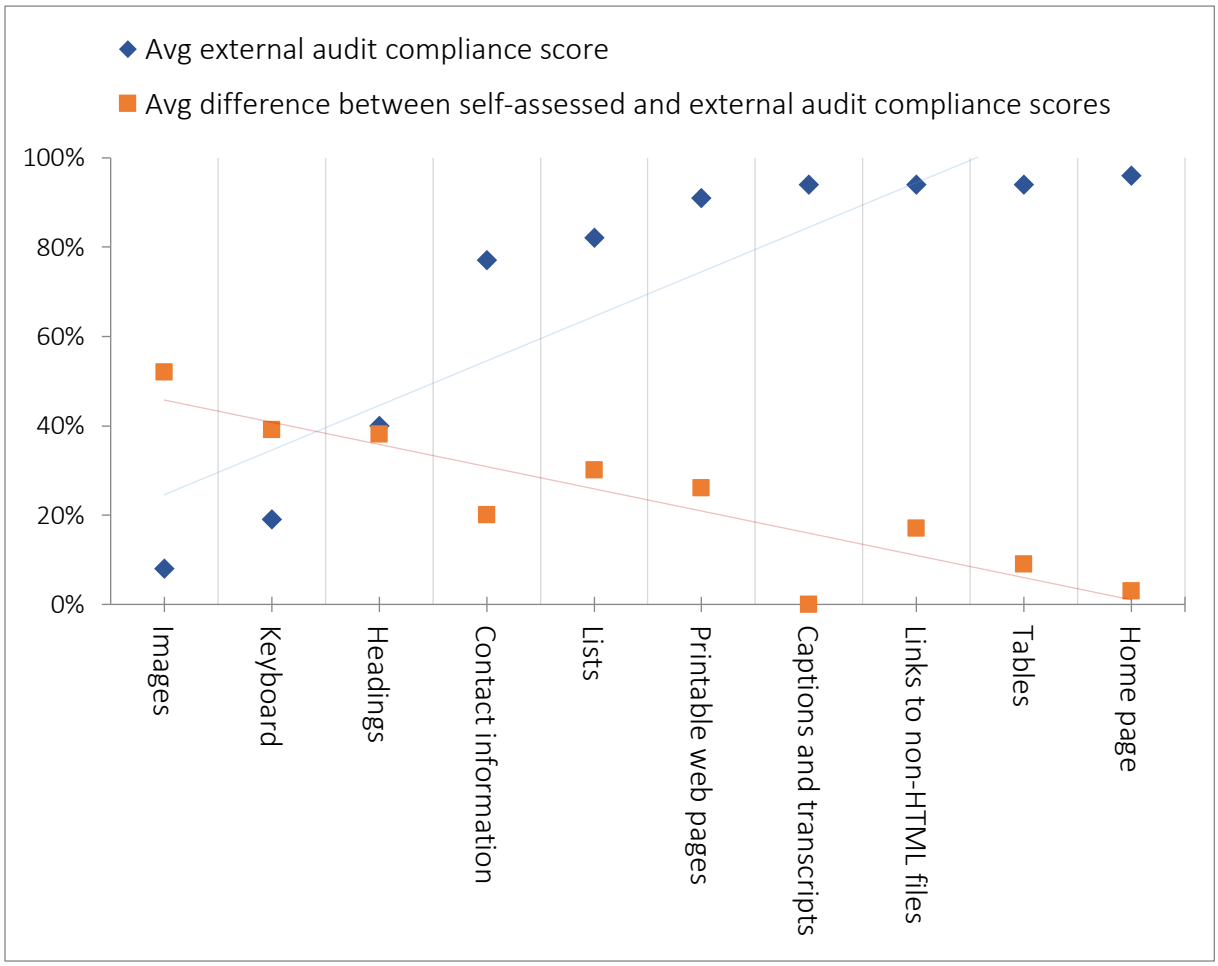
reliability, but there were very few actual tables compared with the number of other elements present in the pages assessed.

Figure 2. Percentage of agency pass/fail scores changed as part of the external audit.



If we assume that the variance between the agency and external audit score reflects the agencies' understanding of a requirement, we would expect to see an inverse relationship between that variance and the compliance score for that requirement. Such a relationship was identified from the 2014 Web Standards Self-Assessments, and is confirmed by the results from the 2017 Self-Assessments. As shown in Figure 3 below, as the compliance scores per requirement improve, the variance between agency and external audit scores go down.

Figure 3. SAM manual test scores from the external audit compared to the average difference between that score and agencies' self-assessed score. See Appendix E for raw data.



Key issue #3

Agencies' SAM manual test results were, on average, inaccurate by 25%. The 3 tests that were least accurately assessed by agencies (the Images, Keyboard, and Headings tests), were the 3 tests most commonly failed by the web pages that were audited. This suggests that agency websites fail these requirements, and agencies inaccurately assess their websites' conformance with those requirements, for the same reason: a lack in agencies' understanding of the requirements.

Recommendation #3

Workshops on how to conduct the self-assessments and follow the SAM were held early in the 2017 Self-Assessment programme. It is recommended that these continue. If the SAM is maintained as a practical collection of easy-to-use tests that can be run anytime and anywhere, then these workshops can be regular, ongoing occurrences that continually raise the visibility and practitioners' knowledge of the Web Standards.

6.5.5 Automated test results

The SAM automated tests were performed using the aXe extension for the Chrome browser. Since the aXe results fairly directly translate to WCAG 2.0 SC, they serve best as an expression of compliance to those particular WCAG SC. For this reason, the automated aXe results are addressed in the next section, SAM results as an expression of WCAG 2.0.

Additionally, given that they are automated results, there is little to no reason to review or compare agencies' aXe results with those from the external audit. In some ways, the agency aXe results are better: They're from the same tool, so should be just as robust as the external audit aXe results; but the agency aXe results are from a much larger sample (based on 20 pages per agency, as compared to the 3 pages per agency tested in the external audit), and therefore, taken on their own, will be more representative.

6.6 SAM results as an expression of WCAG 2.0

The Web Accessibility Standard is a slightly modified version of WCAG 2.0 Level AA. To understand what the SAM test results mean in terms of WCAG conformance, it is important to know which WCAG SC corresponds to which individual SAM test, both manual and automated.

6.6.1 Mapping SAM manual tests to WCAG

There is no one-to-one relationship between SAM manual tests and WCAG SC. While the Images, Headings, Lists, and Tables tests from the SAM each correspond to a single WCAG SC, the Keyboard and Captions and Transcripts tests can each relate to up to 3 WCAG SC.

Table 6. Mapping SAM manual tests to their relevant WCAG 2.0 Success Criteria.

SAM Test	Matching WCAG 2.0 Success Criteria
Images	1.1.1 Non-text Content
Headings	1.3.1 Info and Relationships
Lists	1.3.1 Info and Relationships
Tables	1.3.1 Info and Relationships
Keyboard	2.1.1 Keyboard 2.1.2 No Keyboard Trap 2.4.7 Focus Visible
Captions and Transcripts	1.2.2 Captions (Prerecorded) 1.2.3 Audio Description or Media Alternative

SAM Test	Matching WCAG 2.0 Success Criteria
	(Prerecorded) 4.1.2 Name, Role, Value

For instance, a single page might register a fail for the Keyboard test because it is not always visually indicated which link currently has keyboard focus (a violation of WCAG SC 2.4.7), and also because some widget on the page just does not work with a keyboard (a violation of SC 2.1.1). In such a case, a single SAM test result comprises violations of 2 distinct WCAG 2.0 SC.

Furthermore, three of the SAM tests correspond to WCAG SC 1.3.1. So, for example, a fail of the Headings test entails a fail for all the SC 1.3.1 related SAM tests, even if Lists and Tables record only passes. In such a scenario, WCAG SC 1.3.1 would receive both a pass and a fail result.

Even the manual Images test captured (in the auditors' notes) a range of different failure causes related to SC 1.1.1. Yet, in terms of the actual Pass/Fail results, the details of these different failure causes were not clearly captured by the SAM.

As a result of this relationship between the SAM manual tests and WCAG 2.0 SC, a SAM compliance score (based on failures of the SAM manual tests) cannot be translated directly to a meaningful WCAG 2.0 compliance score.

However, by expanding the SAM test results for Keyboard and Captions and Transcripts into their individual issues, and assigning those to relevant WCAG SC, the SAM results offer a more detailed view of WCAG conformance, albeit still for just a subset of the WCAG 2.0 SC required by the Web Accessibility Standard. We still do not get a full WCAG 2.0 compliance score, but we do know more about some of the specific WCAG SC failures for a particular page. When considering the whole sample of pages audited, we can also rank the incidence of WCAG SC failures, which helps to identify which are the most common.

Key issue #4

Details about specific causes of failure could not always be derived from the agency results without additional interpretation and refinement because of the way that multiple failures against several WCAG 2.0 SC and other important details could be encapsulated within a single SAM test, i.e. the manual Keyboard and Captions and Transcripts tests.

Recommendation #4

The SAM Keyboard and Captions and Transcripts tests should be revised so that the more detailed individual errors associated with discrete WCAG 2.0 SC can be recorded and planned for remediation in a programme of work. It is recommended that the SAM manual tests related to SC 1.1.1, 1.3.1, 2.1.1 and 2.4.7 be refined into a number of more discrete tests to elicit more detailed, actionable results.

6.6.2 WCAG indications from SAM manual test results

The SAM manual test scores from both agency and external audit results were converted to their representative WCAG 2.0 SC, based on the assessor-provided notes accompanying each result. This included expanding the Keyboard test failures into their individual WCAG-related issues under WCAG SC 2.1.1, 2.1.2, 2.4.7. Similarly, the Captions and Transcripts test failures were expanded as appropriate into their discrete issues under WCAG SC 1.2.2, 1.2.3, and 4.1.2.

As noted above, because there is no direct one-to-one relationship between the SAM manual tests and WCAG 2.0, the SAM manual tests cannot be converted into a WCAG compliance score.

Note the similar distribution of failures in the agency and external audit results, as shown in Table 7 below. This again suggests (see Finding #3 above) that the overall pattern of issues identified by agencies working through the SAM is representative, in spite of a 25% inaccuracy rate.

Table 7. Percentage distribution of WCAG failures from SAM manual results as recorded by agencies compared with those recorded in the external audit.

WCAG	Failures as per agency results (%)	Failures as per external audit (%)
1.1.1 Non-text Content	28%	31%
1.3.1 Info and Relationships	32%	28%
2.4.7 Focus Visible	27%	25%
2.1.1 Keyboard	7%	13%
1.2.2 Captions (Prerecorded)	2%	1%
1.2.3 Audio Description or Media Alternative (Prerecorded)	2%	1%
2.1.2 No Keyboard Trap	1%	0%
4.1.2 Name Role Value	1%	0%
Total	100%	100%

Finding #4

The SAM manual tests, by their makeup, can be mapped to only 8 of the 37 WCAG SC required by the Web Accessibility Standard. Once mapped to their associated WCAG SC, the SAM manual tests that were most commonly failed (Images, Keyboard, Headings) relate to just 3 WCAG SC (1.1.1, 1.3.1, 2.4.7). See related Key issue and Recommendation #7 below.

6.6.3 Mapping SAM automated tests to WCAG

The SAM automated tests were performed using the aXe extension for the Chrome browser. The aXe extension runs a series of tests on the page currently loaded in the browser, and returns a list of violations found. Based on the documentation for the aXe tests, we associated each of these violations with a specific WCAG 2.0 SC. See Appendix D for a list of aXe errors and their associated WCAG 2.0 SC. Note that a number of violations identified by aXe are considered best practice or otherwise classified, but are not explicit WCAG SC errors. For the purposes of this exercise, which is about how the SAM identified Web Accessibility Standard or WCAG issues, those non-WCAG violations were ignored.

6.6.4 WCAG indications from SAM automated test results

Table 8. Percentage distribution of WCAG failures from SAM automated test results as recorded by agencies.

WCAG Success Criteria	Number of fails	Percentage of total results
1.4.3 Contrast (Minimum)	1253	26%
1.3.1 Info and Relationships	1002	21%
4.1.2 Name Role Value	821	17%
1.1.1 Non-text Content	490	10%
4.1.1 Parsing	406	8%
1.4.4 Resize text	372	8%
3.1.1 Language of Page	275	6%
2.4.1 Bypass Blocks	93	2%

WCAG Success Criteria	Number of fails	Percentage of total results
1.2.1 Audio-only and Video-only (Prerecorded)	57	1%
1.2.2 Captions (Prerecorded)	54	1%
2.4.2 Page Titled	9	0%
3.1.2 Language of Parts	3	0%
Total	4835	100%

According to the aXe results (as seen in Table 8), the greatest number of issues belong to WCAG SC 1.4.3, 1.3.1, and 4.1.2. The SC 1.4.3 errors are colour contrast issues where text and background colours are not sufficiently distinct to enable easy reading by sighted users. Colour contrast issues are well-captured by aXe, which is why they were not included as part of the SAM manual tests.

Looking in more detail at the error messages associated with the aXe findings, the most common issues under SC 1.3.1 were form elements lacking properly associated labels, and heading elements with no content. Neither of these issues were explicitly checked for by the SAM tests, which makes their identification by aXe useful.

The SC 4.1.2 failures found by aXe had mostly to do with links not having discernible text, which results in links with no programmatic name that software, such as screen reader or speech recognition software, can use to identify or refer to it.

Key issue #5

The aXe tool identified common issues with colour contrast, form input labels and empty headings, and links with no accessible name or identifier. These relate to WCAG SC 1.4.3, 1.3.1,

and 4.1.2, respectively. These issues would not have been found through the SAM manual tests alone, making aXe a useful addition to the self-assessment methodology.

Recommendation #5

Continue to advise agencies and web development firms to integrate automated testing tools like aXe (Tenon.io is another example) into their regular work practices. While such tools cannot address all accessibility issues and failures, they can be used to provide reliable, consistent, and accurate results, as opposed to manual tests that require time, effort, and interpretation.

6.7 SAM as indicator of WCAG compliance

For assessing a web page's accessibility, defined by the Web Accessibility Standard as conformance to WCAG 2.0 AA, the 2017 Self-Assessment Methodology (SAM) combined a small number of manual tests with one automated test. The intent of the SAM was to provide practical, easy-to-run manual tests that addressed known common accessibility issues (as identified by the 2014 Self-Assessments). By design, those manual tests did not cover issues that the automated test tool, aXe, was known to address. While the SAM certainly identified accessibility errors across those pages that were tested, to what degree do the SAM results for a web page serve as an indicator of, or proxy for, WCAG 2.0 AA compliance?

To answer this question, a sample of 10 pages from the total population of pages assessed in the 2017 Self-Assessments, was additionally assessed against the full WCAG 2.0 specification, at Level AA (minus the one exemption for SC 1.2.5 under the Web Accessibility Standard). The SAM results for those same 10 pages were compared with the full WCAG 2.0 audit results.

6.7.1 Incomplete WCAG coverage

From the outset, is was clear that the SAM could not fully represent WCAG 2.0, given that the SAM tests do not address every issue covered by the 37 WCAG 2.0 SC required by the Web Accessibility Standard. In total, the SAM manual and automated accessibility tests cover only 16 of the 37 relevant SC:

- 1.1.1 Non-text Content
- 1.2.1 Audio-only and Video-only (Prerecorded)
- 1.2.2 Captions (Prerecorded)
- 1.2.3 Audio Description or Media Alternative (Prerecorded)
- 1.3.1 Info and Relationships

- 1.4.3 Contrast (Minimum)
- 1.4.4 Resize text
- 2.1.1 Keyboard
- 2.1.2 No keyboard trap
- 2.4.1 Bypass Blocks
- 2.4.2 Page Titled
- 2.4.7 Focus visible
- 3.1.1 Language of Page
- 3.1.2 Language of Parts
- 4.1.1 Parsing
- 4.1.2 Name Role Value

Further, certain of these SC, in particular 1.3.1 and 4.1.2, involve a broad range of possible failure conditions that are not explicitly tested for, neither by the SAM manual tests nor the rules applied by the aXe tests. In comparing the full WCAG audit results with those from the SAM for the same pages, there were a significant number of 1.3.1 and 4.1.2 errors representing critical accessibility issues that were identified by the former, but not the latter.

For instance, the SAM manual tests are such that they simply could not possibly have identified the following type of important 1.3.1 errors:

- form labels not programmatically associated with their fields
- interactive states visually indicated, but not programmatically provided
- links with no discernible text

- interactive controls not marked up as interactive controls
- major, discrete page regions (e.g. footer) not programmatically demarcated.

It is worth repeating that the above are critical accessibility issues that can present serious barriers to some users, especially those that rely on assistive technologies. However, the aXe tool is able to find some of the above failures, so the SAM as a whole does provide more coverage of accessibility issues than the manual tests alone.

Still, the aXe tool was only able to find approximately 12% of the 1.3.1 errors, and 30% of the 4.1.2 errors identified in the full WCAG audit. Among the SC 4.1.2 errors identified in the full WCAG audit, but missed by aXe, were the following common accessibility issues:

- interactive elements assigned an incorrect role or state (e.g. an element that acts like a button, but is exposed in the HTML markup as some other kind of element; a push button whose pressed state is marked up as not pressed)
- User interface components with no accessible name (e.g. a button with no discernible content by which it can be named or referred to; form inputs without labels associated with them in the HTML markup).

This is not a criticism of the aXe tool, which purposely limits its tests to those that will not raise false positives. But it does emphasise the limitations of the aXe tool. Unfortunately, these types of accessibility issues are more technically complicated to assess, and typically require more advanced understanding of web technologies like HTML, ARIA, and the way that browsers work with assistive technologies such as screen readers.

Key Issue #6

The aXe tool is useful for testing certain characteristics of web accessibility, but is limited in what it tests for, and cannot reveal all critical accessibility errors or WCAG failures, which requires manual testing.

Recommendation #6

When comprehensively testing for accessibility, whether to a specific standard, e.g. WCAG 2.0 AA, or to inclusive design principles and best practice, the use of automated tools must be supplemented with manual testing, ideally by someone with expertise in how web technologies work to deliver accessible user experiences.

6.7.2 SAM and WCAG 2.0 compliance rates do not compare

The SAM manual accessibility tests did find many other, often significant, accessibility issues that the aXe tool did not, particularly to do with WCAG SC 1.1.1, 1.3.1, 2.1.1 and 2.4.7. However, even if the SAM manual and automated accessibility test failures for a web page are combined to produce a SAM compliance rate for that page, that number just does not compare in any meaningful or consistent fashion with the full WCAG audit scores for the same page.

Certainly, SAM results may provide an indicator of comparative accessibility when applied to different web pages, and therefore can serve to rate those pages' relative accessibility, as defined by the SAM tests. However, because of the differences between the SAM and a full WCAG 2.0 assessment, the SAM results do not provide a robust indicator of accessibility as defined by the Web Accessibility Standard.

Key issue #7

The SAM does not deliver a representative Web Standards or WCAG compliance measure. One could develop a SAM with tests that translate to and represent a much greater number of WCAG SC. However, that collection of tests could never reasonably represent all WCAG failure conditions, and so there still would be no one-to-one correlation between the SAM and a full WCAG compliance score.

Recommendation #7

As opposed to preparing and running a collection of tests to address all the potential WCAG failure conditions, an expert WCAG audit of a representative sample of pages will be the more cost-effective approach for establishing an average WCAG compliance score for NZ Government websites overall.

For instance, one option might be a centralised full WCAG audit of approximately 70-80 pages (for a reasonably representative sample) that combines manual and automated tests; or a centralised manual WCAG audit of 70-80 pages, plus a much broader automated assessment of 100s or even 1000s of pages from across the NZ Government's web presence.

6.8 Comparing 2017 results to 2014 results

The 2014 Self-Assessments involved complete WCAG 2.0 audits of web pages. As such, comparing the 2017 results with those from 2014 suffers the same limitations as comparing the SAM results to the results of the full WCAG audits. However, if we consider the SAM manual and automated test results represented as WCAG SC failures, there is some clear alignment between them and the 2014 results.

In 2014, the most commonly failed WCAG SC were, in ascending order of compliance:

1. 1.3.1 Info and Relationships
2. 1.1.1 Non-text Content
3. 1.4.3 Contrast (Minimum)
4. 4.1.2 Name, Role, Value
5. 2.4.7 Focus Visible

In 2017, the SAM manual results have WCAG SC 1.1.1, 1.3.1, and 2.4.7 in their top 4, while the top 4 WCAG SC failures as per the aXe test results are the same as those from the 2014 Self-Assessments.

Finding #5

Despite the different approaches between the two Self-Assessment programmes, many of the common issues identified in 2017 remain common in today's NZ Government web pages.

6.9 Other trends

6.9.1 SAM Keyboard results

Each SAM Keyboard test result could represent several issues related to different WCAG SC. For example, a page with links lacking visible focus indicators, and a widget in which keyboard focus gets trapped, preventing keyboard access to the rest of the page, would be recorded as a single failure of the SAM Keyboard test, but represent distinct errors under SC 2.4.7 Focus visible and SC 2.1.2 No keyboard trap.

When these SAM Keyboard results are expanded from a single Fail result into their discrete WCAG SC failures, we get an impression of the relative frequency of those different WCAG SC failures. We can also compare the distribution of these keyboard-related WCAG failures from agencies' own self-assessment results to those of the external audit (Table 9).

Table 9. Percentage distribution of keyboard-related WCAG failures from SAM manual results as recorded by agencies compared with those as recorded in the external audit.

WCAG Success Criteria	Failures as per agency results (%)	Failures as per external audits (%)
2.4.7 Focus Visible	56%	66%
2.1.1 Keyboard	41%	33%
2.1.2 No Keyboard Trap	3%	1%
Total	100%	100%

Comparing agency results to external audit results for keyboard-related WCAG failures, the distribution of those failures is similar across the two result sets. However, the external audit results recorded twice as many SC 2.4.7 failures as SC 2.1.1 failures, whereas the agency results

show only 1.4 times as many SC 2.4.7 failures over SC 2.1.1. In either case, there is clearly a higher incidence of visible focus issues than there are issues with basic keyboard functionality.

While ensuring that all page content is operable via keyboard is a critical requirement for accessibility, if it is not visibly clear which interactive control is currently in focus, sighted keyboard users will have an extremely difficult, if not impossible time taking advantage of any otherwise accessible keyboard functionality.

Finding #6

While agency websites have difficulty making their interactive controls usable by keyboard, the more common keyboard accessibility issue is interactive components lacking a visible indication of when they have keyboard focus and are ready to be activated by the user. Accordingly, educating designers and developers on the importance of visible focus indicators is a relatively clear priority for improving the accessibility of government websites. See Key issue and Recommendation #8 below.

6.9.2 SAM Captions and Transcripts

Just as with the SAM Keyboard test results, each Captions and Transcripts test result could represent different WCAG issues. A video might lack both captions and a descriptive text transcript, representing failures of both WCAG SC 1.2.2 and 1.2.3 within a single SAM result for Captions and Transcripts.

The Captions and Transcripts test was also intended to identify embedded videos, e.g. YouTube or Vimeo, lacking an HTML title attribute on the video's <iframe> container element, which is a failure of WCAG SC 4.1.2.

Table 10. Distribution of WCAG failures related to the SAM Captions and Transcripts test as recorded in the external audit.

WCAG Success Criteria	Number of Fails	Percentage of total results
1.2.3 Audio Description or Media Alternative (Prerecorded) [represents missing transcript]	15	62.5%
4.1.2 Name Role Value	6	25%
1.2.2 Captions (Prerecorded)	3	12.5%
Total	24	100%

Table 10 shows the distribution of WCAG SC failures comprising the Captions and Transcripts results from the external audit. Note that the total number of fails is relatively low compared to some other measures because there were few videos across the 90 web pages that were externally audited.

Despite the smaller sample of web pages with video content, these expanded results for Captions and Transcripts indicates that captions tend to be provided most of the time, but descriptive text alternatives or transcripts are not.

Finding #7

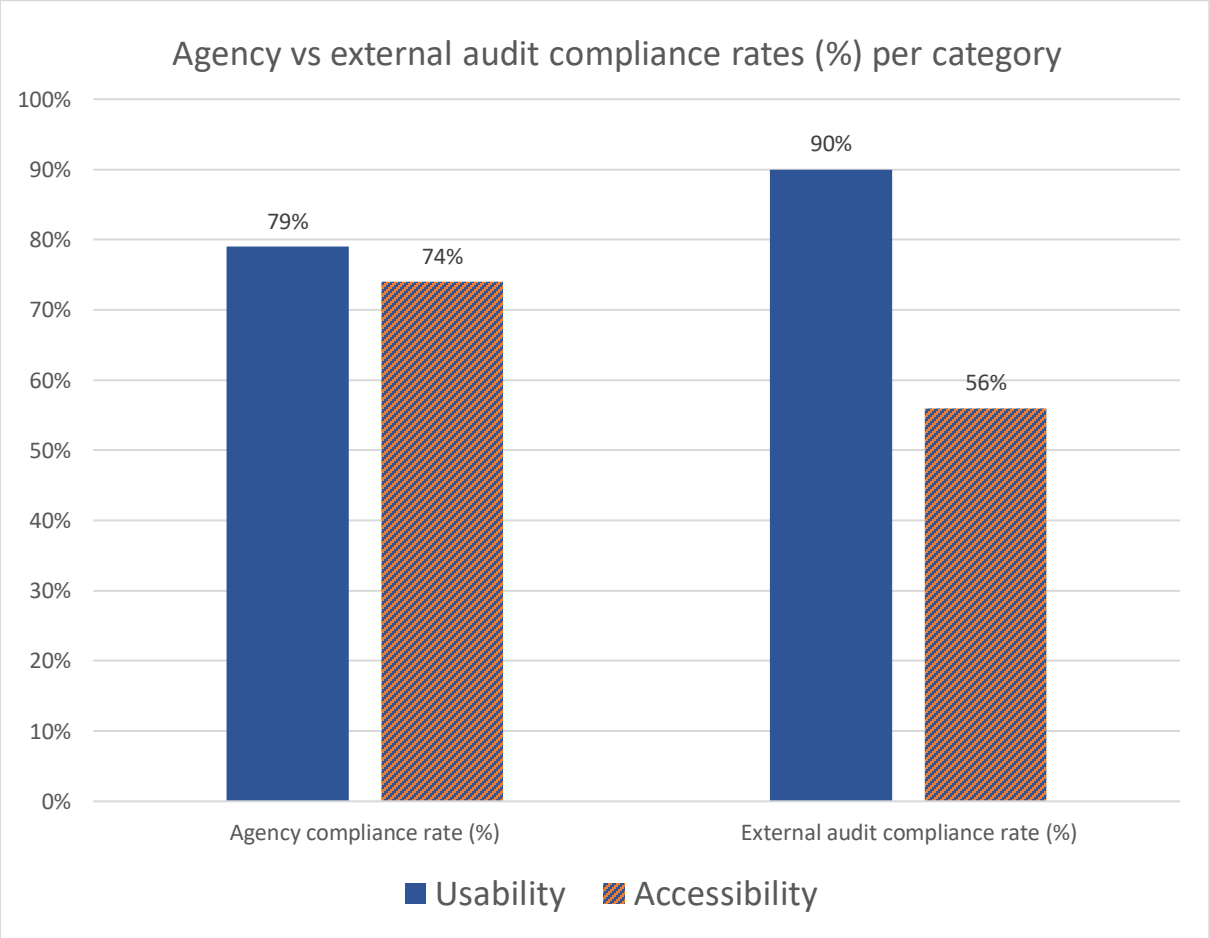
Agencies need cost-effective approaches to providing text alternatives to their online videos. Depending on the video content, the text alternative does not have to include significantly more content than is available already from the captions (should they exist). So, improving the accessibility of government online video may be more a matter of education about what is required as opposed to the technical how-to. See Key issue and Recommendation #8 below.

6.9.3 Comparing SAM manual accessibility and usability results

By filtering the SAM test results by category, i.e. Accessibility vs. Usability, we can compare their average compliance scores as determined by agencies and the external audit.

While the agency self-assessments recorded average compliance rates of 79% and 74% for the accessibility- and usability-related SAM tests, respectively, there was a much greater difference between the two rates as recorded in the external audit. Where the external audits revealed an average compliance rate of 90% for the usability-related tests, that rate dropped to 56% for the accessibility-related tests. The visual comparison in Figure 4 below makes this especially clear.

Figure 4. Comparison of agency vs external audit compliance rates per compliance category.



Finding #8

According to the external audit results, NZ Government web pages do much better meeting the Web Usability Standard (90% compliance) than they do the Web Accessibility Standard (56% compliance). If effort can be spent on improving performance against only one of these Standards, the greatest room for improvement is with the Web Accessibility Standard. See Key issue and Recommendation #8 below.

6.9.4 Desktop vs Tablet vs Phone

As evident from the table that follows, there was effectively little to no difference in the incidence of failures (including all SAM manual and automated test results) between the Desktop viewport to the Tablet viewport to the Phone viewport. This trend was effectively the same for both agencies' own results and the external audit. This suggests that, where there were differences in the content, functionality or layout of a page at these 3 viewports, those differences did not result in any significant change in the general rate of accessibility failures.

Table 11.Percentage distribution of failures across viewports from all SAM results as recorded by agencies compared with those as recorded in the external audit.

Viewport	Failures as per agency results (%)	Failures as per external audits (%)
Desktop	35%	33%
Tablet	33%	33.5%
Phone	32%	33.5%
Total	100%	100%

It could be that in the move, for example, from the Desktop to the Tablet viewport, certain changes in content removed some failures while adding others, effectively balancing out the

overall number of failures on that page. However, more analysis would be required to determine just what kinds of changes to the page are occurring at different viewports.

Finding #9

There is little difference in the number of failures between the Desktop, Tablet, and Phone views of the individual web pages tested. Determining the cause of this consistency across the viewports would require additional analysis.

6.9.5 Home vs "Contact us" vs Other

The 20 pages that each agency had to assess included at least 1 home page, 1 "Contact us" page, and then a collection of other pages to complete the sample. Reviewing the average compliance rate against all SAM manual tests, there is only a slight difference based on whether it was a home, a "Contact us", or other page being assessed (Table 12).

Table 12. Average compliance rates against SAM manual tests for Home, "Contact us", and Other pages as recorded by agencies and the external audit.

Page type	Average compliance rate as per agency results (%)	Average compliance rate as per external audits (%)
Home	72%	54%
"Contact us"	75%	54%
Other	66%	53%

Finding #10

In both agency self-assessments and the external audits, the "Contact us" pages had the lowest average compliance rate, though not by much. Given that many "Contact us" pages include a

contact form, the lower compliance rate may simply reflect the fact that accessibility issues with forms are typical on the web.

6.10 Common accessibility issues

Combining all SAM test results from agencies and the external audits identified some accessibility errors common to NZ Government websites. These accessibility errors are not unusual, and represent typical design, development, and content authoring shortfalls:

- images missing appropriate text alternatives
- interactive elements without visible focus indicators
- videos lacking fully descriptive text transcripts (in addition to captions)
- text and background colours with insufficient contrast
- form elements not properly labelled
- empty heading elements, or headings at the wrong level
- links with no discernible content to provide an accessible name.

Fortunately, being common accessibility issues, there are known solutions that designers, developers, and content authors can apply to rectify those issues.

Key issue #8

As expressed by the SAM manual and auto results, those responsible for delivering NZ Government web content still lack the knowledge and skill to ensure that some common accessibility issues are avoided.

Recommendation #8

Consider the following options for enabling those who contribute to NZ Government websites to deliver accessible experiences for the people visiting and using those websites.

Option A

Develop a collection of guidelines or notes on how to fix each error that the SAM is intended to find. If the SAM tests are refined to address more discrete issues, the results will likely translate more easily into solutions, as opposed to the less practicable outcome of simply aligning the results with specific WCAG SC, for instance. For each test in the next iteration of the SAM, provide relevant practical notes about how to fix the issues addressed by that test.

Option B

Develop practical guidance specific to each of the top issues revealed by the SAM manual and auto results. Issues such as colour contrast, visible focus indicators, and properly structured headings are in fact rather simple to fix. One approach might be a guerrilla campaign of awareness raising and sharing techniques focussed on these specific issues to get them fixed in short order and reduce the overall number of accessibility errors across NZ Government websites.

Option C

Much NZ Government web work is outsourced, so vendors are responsible for many of the issues identified by the self-assessments. To improve vendors' performance around Web Standards compliance, it is important to include and engage with those doing most of the design and development work. Familiarise vendors with the SAM, and solicit their input on potential improvements. This will help get vendors more familiar with the Web Standards, as well as the responsibility for meeting them that they share with their government clients.

Option D

With most NZ Government websites adopting their own design and layout, the variability in Web Standards compliance is only multiplied and more difficult to address. Each unique website needs a separate evaluation and set of solutions. While it would represent a significant change from the status quo, requiring that NZ Government websites partake of a single (but

still customisable) design, layout, and structure could ensure near 100% compliance with the Web Standards for those common elements.

This approach is not without precedent internationally. Government of Canada websites, for instance, all use a standard template known to meet their web standards, but that still allows each site to customise its presentation to some degree. An additional benefit of a common look and feel is that once someone learns how to use one government website, they are well-positioned to using all the others.

7. Conclusion

Were the 2017 Self-Assessments worth the cost and effort to agencies and the NZ Government? The answer to this question will be partly decided by what follows from the 2017 process in terms of new or modified approaches to delivering web content that meets the Web Standards. The 2017 Web Standards Self-Assessments has identified some high priority targets for guidance and training to help agencies improve their Web Standards performance. However, it is recommended that feedback be sought from agencies that participated in the Self-Assessments. This could take the shape of open forums or online surveys regarding how they found the process, what they struggled with, what they found useful about it and why, etc?

Throughout this report on the 2017 Self-Assessments, a number of interesting findings are noted, and a range of key issues raised with associated solutions recommended. See Appendix A for a list of noteworthy findings, and Appendix B for all the key issues and recommendations. In addition to those recommendations, there is likely to be a range of other strategic activities and interventions that could be planned, but that require greater familiarity with the current plans, priorities, and structures currently in place within the department responsible for the Web Standards, and across NZ Government agencies.

Appendix A: Findings

Finding #1

Overall, according to agencies' self-assessment results, the NZ Government web pages assessed have an average compliance rate of 76% against the SAM. Correlatively, the web pages had an average failure rate of 24%. Compare with Finding #2 regarding the average compliance rate as determined by the external audit results.

Finding #2

The external audit results established a 65% average compliance rate against the SAM. This represents an average failure rate of 35%, which is 11 percentage points greater than the failure rate measured by agencies. This has implications regarding the accuracy of agencies' own self-assessment results. See Key issue and Recommendation #3.

Finding #3

Despite the difference in expertise, agency results do conform in their general portrait of compliance to the external audit results, and that the 3 tests (Images, Keyboard, Headings) with the lowest compliance rates represent the tests that were least well-handled or most difficult to perform by agency assessors, and the areas where NZ Government websites commonly fail to meet Web Standards requirements. See related Key issue and Recommendation #3.

Finding #4

The SAM manual tests, by their makeup, can be mapped to only 8 of the 37 WCAG SC required by the Web Accessibility Standard. Once mapped to their associated WCAG SC, the SAM manual tests that were most commonly failed (Images, Keyboard, Headings) relate to just 3 WCAG SC (1.1.1, 1.3.1, 2.4.7). See related Key issue and Recommendation #7.

Finding #5

Despite the different approaches between the two Self-Assessment programmes, many of the common issues identified in 2017 remain common in today's NZ Government web pages.

Finding #6

While agency websites have difficulty making their interactive controls usable by keyboard, the more common keyboard accessibility issue is interactive components lacking a visible indication of when they have keyboard focus and are ready to be activated by the user. Accordingly, educating designers and developers on the importance of visible focus indicators is a relatively clear priority for improving the accessibility of government websites. See Key issue and Recommendation #8.

Finding #7

Agencies need cost-effective approaches to providing text alternatives to their online videos. Depending on the video content, the text alternative does not have to include significantly more content than is available already from the captions (should they exist). So, improving the accessibility of government online video may be more a matter of education about what is required as opposed to the technical how-to. See Key issue and Recommendation #8.

Finding #8

According to the external audit results, NZ Government web pages do much better meeting the Web Usability Standard (90% compliance) than they do the Web Accessibility Standard (56% compliance). If effort can be spent on improving performance against only one of these Standards, the greatest room for improvement is with the Web Accessibility Standard. See Key issue and Recommendation #8.

Finding #9

There is little difference in the number of failures between the Desktop, Tablet, and Phone views of the individual web pages tested. Determining the cause of this consistency across the viewports would require additional analysis.

Finding #10

In both agency self-assessments and the external audits, the "Contact us" pages had the lowest average compliance rate, though not by much. Given that many "Contact us" pages include a contact form, the lower compliance rate may simply reflect the fact that accessibility issues with forms are typical on the web.

Appendix B: Key issues and recommendations

Key issue #1

There was significant variability in how agencies followed the Self-Assessment process and recorded their manual test results in the Self-Assessment results spreadsheet. This complicated the initial data, which required substantial effort to normalise, and forced some interpretations of the results.

Recommendation #1

To help reduce the variability in how agencies performed the SAM manual tests and recorded the results, it is recommended that assessors be asked simple yes/no or pass/fail questions, and that some mechanisms be established to ensure data consistency. For example, if self-assessment remains the approach taken for measuring Web Standards performance, provide a tool, ideally online (perhaps something like a survey), that restricts the answers that can be recorded. More detailed step-by-step instructions for tests with a more granular focus could also help, to avoid multiple failure issues being recorded under a single test result.

If agency web practitioners and others will be expected to continue performing this type of assessment, there will be a balance to strike between enough detailed instruction for clarity and ease, and not overwhelming assessors who are neither technical web nor accessibility experts.

Key issue #2

The results from the aXe tool that were saved in the spreadsheet were prey to inconsistencies that could lead to incorrect data being recorded. This was due to the rather tricky copy and paste procedure required to select and save the aXe results.

Recommendation #2

A centralised approach to the automated testing, where one agency performs automated tests across the entire population of pages to be assessed would reduce the burden on agencies for performing what is essentially a machine-based test that should not and does not require manual activity by individual assessors at every agency.

Key issue #3

Agencies' SAM manual test results were, on average, inaccurate by 25%. The 3 tests that were least accurately assessed by agencies (the Images, Keyboard, and Headings tests), were the 3 tests most commonly failed by the web pages that were audited. This suggests that agency websites fail these requirements, and agencies inaccurately assess their websites' conformance with those requirements, for the same reason: a lack in agencies' understanding of the requirements.

Recommendation #3

Workshops on how to conduct the self-assessments and follow the SAM were held early in the 2017 Self-Assessment programme. It is recommended that these continue. If the SAM is maintained as a practical collection of easy-to-use tests that can be run anytime and anywhere, then these workshops can be regular, ongoing occurrences that continually raise the visibility and practitioners' knowledge of the Web Standards.

Key issue #4

Details about specific causes of failure could not always be derived from the agency results without additional interpretation and refinement because of the way that multiple failures against several WCAG 2.0 SC and other important details could be encapsulated within a single SAM test, i.e. the manual Keyboard and Captions and Transcripts tests.

Recommendation #4

The SAM Keyboard and Captions and Transcripts tests should be revised so that the more detailed individual errors associated with discrete WCAG 2.0 SC can be recorded and planned for remediation in a programme of work. It is recommended that the SAM manual tests related to SC 1.1.1, 1.3.1, 2.1.1 and 2.4.7 be refined into a number of more discrete tests to elicit more detailed, actionable results.

Key issue #5

The aXe tool identified common issues with colour contrast, form input labels and empty headings, and links with no accessible name or identifier. These relate to WCAG SC 1.4.3, 1.3.1, and 4.1.2, respectively. These issues would not have been found through the SAM manual tests alone, making aXe a useful addition to the self-assessment methodology.

Recommendation #5

Continue to advise agencies and web development firms to integrate automated testing tools like aXe (Tenon.io is another example) into their regular work practices. While such tools cannot address all accessibility issues and failures, they can be used to provide reliable, consistent, and accurate results, as opposed to manual tests that require time, effort, and interpretation.

Key Issue #6

The aXe tool is useful for testing certain characteristics of web accessibility, but is limited in what it tests for, and cannot reveal all critical accessibility errors or WCAG failures, which requires manual testing.

Recommendation #6

When comprehensively testing for accessibility, whether to a specific standard, e.g. WCAG 2.0 AA, or to inclusive design principles and best practice, the use of automated tools must be supplemented with manual testing, ideally by someone with expertise in how web technologies work to deliver accessible user experiences.

Key issue #7

The SAM does not deliver a representative Web Standards or WCAG compliance measure. One could develop a SAM with tests that translate to and represent a much greater number of WCAG SC. However, that collection of tests could never reasonably represent all WCAG failure conditions, and so there still would be no one-to-one correlation between the SAM and a full WCAG compliance score.

Recommendation #7

As opposed to preparing and running a collection of tests to address all the potential WCAG failure conditions, an expert WCAG audit of a representative sample of pages will be the more cost-effective approach for establishing an average WCAG compliance score for NZ Government websites overall.

For instance, one option might be a centralised full WCAG audit of approximately 70-80 pages (for a reasonably representative sample) that combines manual and automated tests; or a

centralised manual WCAG audit of 70-80 pages, plus a much broader automated assessment of 100s or even 1000s of pages from across the NZ Government's web presence.

Key issue #8

As expressed by the SAM manual and auto results, those responsible for delivering NZ Government web content still lack the knowledge and skill to ensure that some common accessibility issues are avoided.

Recommendation #8

Consider the following options for enabling those who contribute to NZ Government websites to deliver accessible experiences for the people visiting and using those websites.

Option A

Develop a collection of guidelines or notes on how to fix each error that the SAM is intended to find. If the SAM tests are refined to address more discrete issues, the results will likely translate more easily into solutions, as opposed to the less practicable outcome of simply aligning the results with specific WCAG SC, for instance. For each test in the next iteration of the SAM, provide relevant practical notes about how to fix the issues addressed by that test.

Option B

Develop practical guidance specific to each of the top issues revealed by the SAM manual and auto results. Issues such as colour contrast, visible focus indicators, and properly structured headings are in fact rather simple to fix. One approach might be a guerrilla campaign of awareness raising and sharing techniques focussed on these specific issues to get them fixed in short order and reduce the overall number of accessibility errors across NZ Government websites.

Option C

Much NZ Government web work is outsourced, so vendors are responsible for many of the issues identified by the self-assessments. To improve vendors' performance around Web Standards compliance, it is important to include and engage with those doing most of the design and development work. Familiarise vendors with the SAM, and solicit their input on potential improvements. This will help get vendors more familiar with the Web Standards, as well as the responsibility for meeting them that they share with their government clients.

Option D

With most NZ Government websites adopting their own design and layout, the variability in Web Standards compliance is only multiplied and more difficult to address. Each unique website needs a separate evaluation and set of solutions. While it would represent a significant change from the status quo, requiring that NZ Government websites partake of a single (but still customisable) design, layout, and structure could ensure near 100% compliance with the Web Standards for those common elements.

This approach is not without precedent internationally. Government of Canada websites, for instance, all use a standard template known to meet their web standards, but that still allows each site to customise its presentation to some degree. An additional benefit of a common look and feel is that once someone learns how to use one government website, they are well-positioned to using all the others.

Appendix C: Self-Assessment Methodology

The complete 2017 Web Standards Self-Assessment Methodology (SAM) was published online[4] for easy access by all participating agencies.

At a high level, the SAM had three main steps:

- Identify the pages to assess.
- Assess those pages by completing the tests specified in the SAM.
- Complete and submit the self-assessment spreadsheet and action plan report.

Prerequisites

To be able to follow the SAM, assessors were expected to have:

- enough basic knowledge of HTML to be able to inspect code and understand how elements are structured
- enough familiarity with their agency's websites to pick pages based on their content
- access to the Chrome browser with the Web Developer and aXe extensions installed
- access to either Internet Explorer or Firefox for testing keyboard accessibility.

The Methodology included a link to a downloadable copy of the Self-Assessment Spreadsheet.

[4] The 2017 Web Standards Self-Assessment is available online at https://govtnz.github.io/web-standards/guidance/assessment/2017-web-standards-self-assessment-methodology.html.

Selecting pages to assess

Each agency was to assess 20 pages from across its top 3 most visited sites. These 20 pages were to include the home page and the "Contact us" page from each of those sites. The remaining pages were to be selected based on their content to ensure that the total sample included representative pages with:

- lists (ordered and/or unordered)
- tables or information presented in HTML tables
- forms and/or widgets (such as date pickers, calculators etc)
- images
- video.

These remaining pages were to be selected from the most visited pages on the site(s), favouring the most visited site. So for example, if an agency was assessing 3 sites, it would have its 3 home page and 3 "Contact us" pages, leaving 14 more pages to be selected from across the 3 sites. Of these 14 pages, 6 might come from the busiest site, and 4 each from the second and third busiest. Agencies were expected to use their own judgement.

If, for example, a smaller agency only had 1 website, then it would need to include its home and "Contact us" pages, and then 18 other representative pages selected from the website's most visited pages.

Assessing the web pages

With its sample of pages selected, agencies were to assess those pages following the process outlined in the SAM. Some agencies performed this work in-house, and some hired an external vendor to support them.

Each of the tests in the Methodology was related either to the Web Accessibility or the Web Usability Standard.

Under the Web Accessibility Standard there were 7 tests. The first involved using an automated accessibility testing tool, the aXe extension for the Chrome browser. The remaining 6 were manual tests that addressed the following common types of web content or interactivity:

- Headings
- Lists
- Tables
- Keyboard access
- Images
- Video captions and transcripts

Detailed instructions were provided on how to use the Chrome browser's built-in developer tools and Web Developer extension in different ways to complete the tests.

Under the Web Usability Standard there were 4 tests. The first 2 involved checking the home and "Contact us" pages for specific information that every website must have as per the Standard. The last 2 checked that links to downloadable files (e.g. PDF, Word) included the file's format and size, and that the core content on each web page was printable on standard sheets of paper.

Additionally, with the exception of those last 2 Web Usability Standard tests, each page was to be assessed at 3 different responsive layouts or viewports, one representing a desktop or large laptop, one a tablet-sized device, and one a smartphone. This approach was taken to acknowledge the often quite different content, layout or interactivity that a web page presents to the user depending on the size of the device that is used. In effect, this almost tripled the number of test results recorded by each agency.

Action plan reports

Each agency was also to submit, along with its self-assessment results spreadsheet, an action plan report. This report was to include any plans the agency's Chief Information Officer or similarly placed executive manager had for addressing risks and opportunities raised by the self-assessment results, and for improving the agency’s web presence with regard to the Web Standards.

No template for the report was provided. Agencies were simply instructed to describe areas of non-conformance revealed in the self-assessments, and describe their plans and timeframes for fixing the issues identified.

Appendix D: Mapping aXe errors to WCAG 2.0

aXe errrors	WCAG 2.0 Success Criteria
<dl> elements must only directly contain properly-ordered <dt> and <dd> groups, <script> or <template> elements	1.3.1 Info and Relationships
<dt> and <dd> elements must be contained by a <dl>	1.3.1 Info and Relationships
<html> element must have a lang attribute	3.1.1 Language of Page
<html> element must have a valid value for the lang attribute	3.1.1 Language of Page
elements must be contained in a or	1.3.1 Info and Relationships
<object> elements must have alternate text	1.1.1 Non-text Content
and must only directly contain , <script> or <template> elements	1.3.1 Info and Relationships
<video> elements must have an audio description track	1.2.1 Audio-only and Video-only (Prerecorded)
<video> elements must have captions	1.2.2 Captions (Prerecorded)
Active <area> elements must have alternate text	1.1.1 Non-text Content
All th elements and elements with role=columnheader/rowheader must have data cells they	1.3.1 Info and Relationships

aXe errrors	WCAG 2.0 Success Criteria
describe	
Buttons must have discernible text	4.1.2 Name Role Value
Certain ARIA roles must be contained by particular parents	1.3.1 Info and Relationships
Certain ARIA roles must contain particular children	1.3.1 Info and Relationships
Documents must have <title> element to aid in navigation	2.4.2 Page Titled
Elements must have sufficient color contrast	1.4.3 Contrast (Minimum)
Form elements must have labels	1.3.1 Info and Relationships
Frames must have title attribute	4.1.2 Name Role Value
Headings must not be empty	1.3.1 Info and Relationships
id attribute value must be unique	4.1.1 Parsing
Image buttons must have alternate text	1.1.1 Non-text Content
Images must have alternate text	1.1.1 Non-text Content
lang attribute must have a valid value	3.1.2 Language of Parts
Layout tables must not use data table elements	1.3.1 Info and Relationships
Links must have discernible text	4.1.2 Name Role Value

aXe errrors	WCAG 2.0 Success Criteria
Page must have means to bypass repeated blocks	2.4.1 Bypass Blocks
The skip-link target should exist and be focusable	2.4.1 Bypass Blocks
Zooming and scaling must not be disabled	1.4.4 Resize text

Appendix E: Raw data for Figure 3

Criteria	Average external audit compliance score	Average difference between agency and external audit compliance scores
Images	8%	52%
Keyboard	19%	39%
Headings	40%	38%
Contact info	77%	20%
Lists	82%	30%
Printable web pages	91%	26%
Captions and transcripts	94%	0%
Links to non-HTML files	94%	17%
Tables	94%	9%
Home page	96%	3%