Executive Summary: Gutenberg Accessibility Audit

Prepared by: Tenon, LLC.

Authors: Karl Groves, James Edwards, Mallory van Achterberg, Helen Burge, and Michael Beck

April 2, 2019

Table of Contents

Index of Figures

Index of Tables

Introduction and Overview of Findings

WordPress is a free and open source content management system in use by more than 60 million websites, including 30% of the top 10 million websites as of April 2018. Thanks to its flexible templating & plugin capabilities and robust community of users and developers, the product's users range from small individual bloggers to large companies, educational institutions, and government agencies across the world.

On December 6, 2018, version 5.0 of WordPress was released which includes the new Gutenberg editor. However, work on Gutenberg began in Spring of 2017 with accessibility-related issues being logged in the Gutenberg project's GitHub repo shortly thereafter. In March 2018, the WordPress accessibility team began testing Gutenberg in earnest and logging issues they discovered during testing.

In November 2018, WPCampus issued an RFP to find a provider to do an accessibility audit on Gutenberg. Specifically, the WPCampus community sought to identify:

- Potential legal risk (if any) posed by upgrading to Gutenberg
- Potential challenges when supporting the new editor, particularly for assistive technology users
- Potential impact of Gutenberg on the accessibility of public-facing web content

On December 20, 2018 Tenon's proposal was accepted. Our proposal included a technical audit of the Gutenberg experience as well as user-based testing with persons with disabilities. Testing began on January 15, 2019 and concluded on March 1, 2019. This Executive Summary document provides a description of Tenon's audit of Gutenberg, our methodology, key findings, and a summary of the data uncovered during testing.

Key Findings

During usability testing of Gutenberg, participants only had an average completion rate of 63% across three tasks, with an average System Usability Scale (SUS) score of 46. The scores within the UX Report indicate that Gutenberg can be described, at best, as "Poor" to "OK". The usability test results are backed by the data uncovered during technical review.

During the technical review, Tenon staff logged a total of 90 issues across 16 components tested. As the data in this report will show, there is an average of 18% failed check items per component, with 63% of issues impacting at least 3 user populations. 69% of issues are Medium or High severity. Finally, 53% of the issues logged in the Technical Report are for Level A Success Criteria.

On the ability to generate accessible markup with Gutenberg

For the most part, the markup generated from the Gutenberg editor is clean, semantically correct and accessible. It's clear that a lot of attention has been paid to things like using proper headings and lists in the output, which makes it easy for users to create accessible content without having to know HTML.

However, it's still relatively easy for users to unknowingly create accessibility problems, for example, by using the wrong heading levels within a post, creating multiple posts with the same title, or adding videos which autoplay. The editor could do more to warn users when they do this kind of thing.

Some specific limitations in the visual interface also make it difficult for certain kinds of content to be fully accessible, because it doesn't provide the tools to define all necessary semantics and metadata. The Video Block interface, for example, doesn't provide a way to associate a captions file with the video; the Table Block interface doesn't allow users to define a caption, header rows, header cells, or scopes. Gutenberg users can only add this information by using the HTML editor, but this of course requires them to have sufficient HTML knowledge.

Deliverables

The deliverables provided to WPCampus are in the form of a zip file containing the following:

- **allIssuesExport.csv:** A full export of the issues logged during the technical audit
- Executive Summary (This document)
- **GutenbergIssueData.xlsx:** The data used to create the charts and tables within this document. The data was pulled from the issues export.
- **TechnicalReportGutenberg.pdf:** The long form technical report describing each issue logged.
- **Tenon Audit Report Issue Import Guide:** A document describing how the results from the technical audit can be imported into issue tracking systems
- **Test Plan – Gutenberg:** Our Test Plan document which contains a full list, with screenshots, of each component tested.
- **UX Testing Results:** A sub folder containing the results of user-based testing, including a full report as well as data to substantiate the findings.

UX Report

Key Findings of the UX Report

Tenon performed a usability study using 9 participants:

- 3 participants were blind.
- 2 participants were blind with motor impairments
- 1 participant had cognitive impairments
- 2 participants had mobility impairments
- 1 participant was visually impaired

All participants had a higher-than-average level of technical experience, some of whom having professional experience in web development and some having direct and current experience with WordPress.

Each participant was asked to perform 3 tasks:

- Create new content
- Edit content
- Amend post options

Testing was conducted using a think aloud protocol whereby participants verbalize what comes to mind as they complete a task. For this project, the test facilitator also asked the participant to complete a System Usability Scale (SUS) questionnaire to measure effectiveness, efficiency, and satisfaction with the task. The below table shows the SUS Average scores and completion rates for each task.

Table 1: Data, SUS and Task Completion Rate

Task and Participant Type	SUS Average	Completion Rate
Task 1: All Users	**50**	**66.67%**
Task 1: Blind	39	66.67%
Task 1: Visually Impaired	90	0%
Task 1: Cognitively Challenged	75	100%
Task 1: Dexterity Impaired	38	66.67%
Task 2: All Users	**48**	**77.78%**
Task 2: Blind	33	66.67%
Task 2: Visually Impaired	95	100%
Task 2: Cognitively Challenged	65	100%
Task 2: Dexterity Impaired	45	66.67%
Task 3: All Users	**39**	**44.44%**
Task 3: Blind	8	16.67%
Task 3: Visually Impaired	95	100%
Task 3: Cognitively Challenged	75	100%
Task 3: Dexterity Impaired	27	66.67%

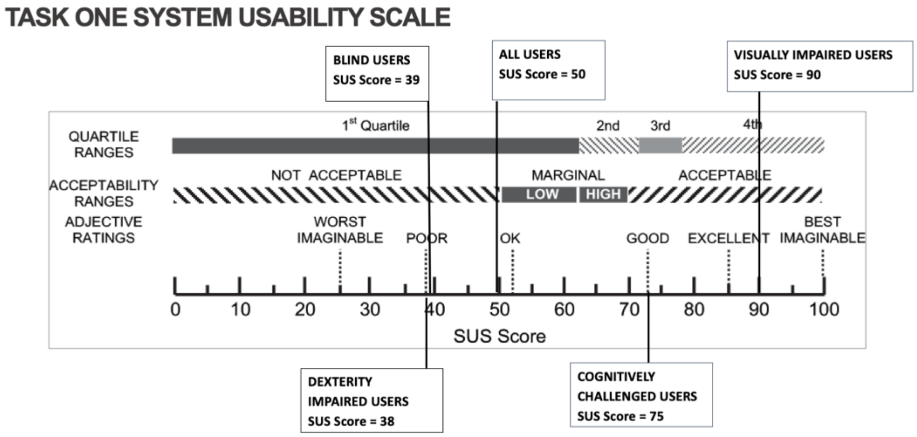


Figure 1: Task One SUS Scale

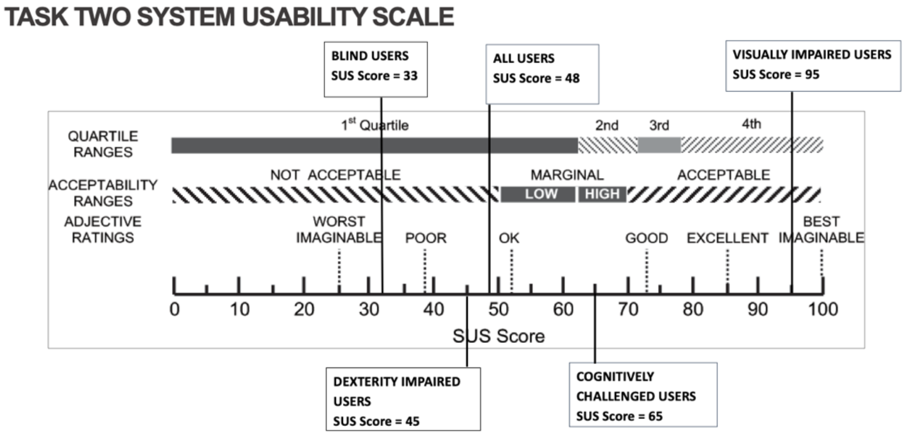


Figure 2: Task Two SUS Scale

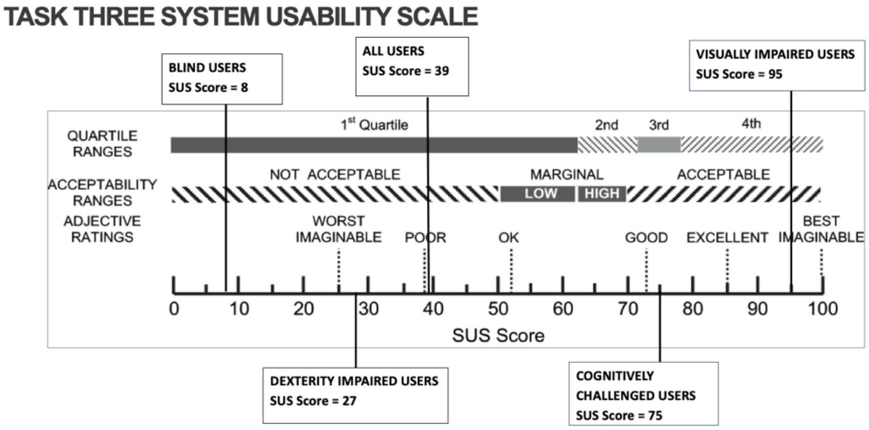


Figure 3: Task Three SUS Scale

As the above demonstrates, the SUS score, when measured across all test participants places Gutenberg squarely in the 1st Quartile, described as “Not Acceptable”. Cognitively Impaired and Visually Impaired users were the only participants who scored higher than the 1st Quartile. Overall, Gutenberg’s user experience is consistently poor. The data from our Technical Report provides insight into the development shortcomings that caused such low ratings.

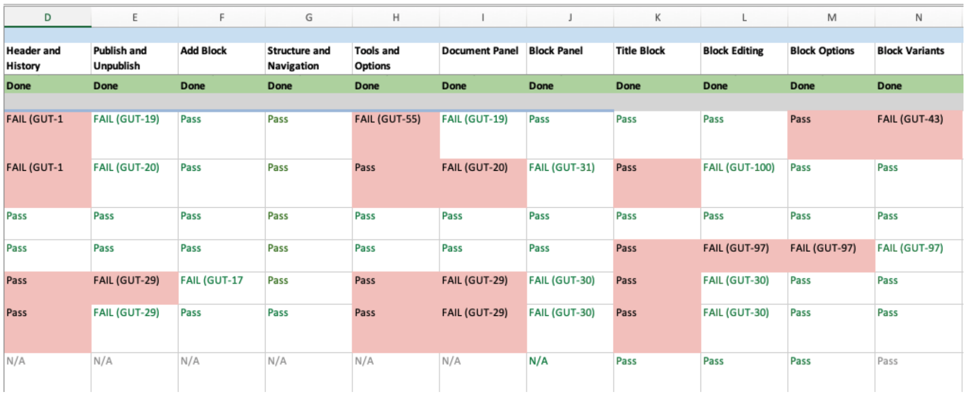
Technical Report

Methodology

Tenon’s technical audit methodology is derived from the collective experiences of our consulting staff who, all together, have performed nearly 300 accessibility audits of websites, software, and hardware. The goal of our methodology is to efficiently and accurately gather comprehensive data on the state of the tested system with respect to its accessibility. Our methodology ensures accuracy, reliability, repeatability, and defensibility.

The testing process begins by collecting the set of user interface (UI) components to be tested. These components are typically unique examples of common UI elements, such as forms, frames, tables, global navigation, and so on. The list of specific components tested during this audit can be found in the Test Plan document.

Once the list of components has been determined & the test plan approved by our customer, Tenon's testing staff evaluate each component using a list of approximately 200 check points. Tenon's checklist is a proprietary list of common failure conditions encountered during our cumulative history as consultants. As each item is tested, they're marked as N/A, Pass, or Fail.



D	E	F	G	H	I	J	K	L	M	N
Header and History	Publish and Unpublish	Add Block	Structure and Navigation	Tools and Options	Document Panel	Block Panel	Title Block	Block Editing	Block Options	Block Variants
Done	Done	Done	Done	Done	Done	Done	Done	Done	Done	Done
FAIL (GUT-1	FAIL (GUT-19)	Pass	Pass	FAIL (GUT-55)	FAIL (GUT-19)	Pass	Pass	Pass	Pass	FAIL (GUT-43)
FAIL (GUT-1	FAIL (GUT-20)	Pass	Pass	Pass	FAIL (GUT-20)	FAIL (GUT-31)	Pass	FAIL (GUT-100)	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	FAIL (GUT-97)	FAIL (GUT-97)	FAIL (GUT-97)
Pass	FAIL (GUT-29)	FAIL (GUT-17	Pass	Pass	FAIL (GUT-29)	FAIL (GUT-30)	Pass	FAIL (GUT-30)	Pass	Pass
Pass	FAIL (GUT-29)	Pass	Pass	Pass	FAIL (GUT-29)	FAIL (GUT-30)	Pass	FAIL (GUT-30)	Pass	Pass
N/A	N/A	N/A	N/A	N/A	N/A	N/A	Pass	Pass	Pass	Pass

Figure 4: Screenshot of Gutenberg checklist

When a check point fails, Tenon's tester logs the details relating to the failure into our issue tracking & reporting system.

In any case where testing uncovers an accessibility problem unaccounted for in the checklist, it is logged, and an internal discussion takes place as to whether the new issue type should be included in the checklist.

Understanding the Technical Report

The Technical Report provides an itemized list of each issue logged during the Gutenberg audit, organized by Content Type. For each Content Type, all relevant issues are listed in detail. Each issue provides a long form description of the problem, code samples & screenshots of the problem, and detailed remediation guidance with recommended code. Each issue also contains references to the applicable WCAG Success Criterion and metadata for understanding the severity, populations affected, and platform information. The level of detail provided within the Technical Report should support the complete remediation of each issue by Gutenberg's developers.

Overview of the data

The rest of this document provides a breakdown of the data found in the Technical Report. The following sections summarize the audit's findings in various contexts, helpful for understanding the scope, nature, and severity of the issues listed in the Technical Report:

- Test Outcomes & Rate of Failure
- Issues by Component
- Issues by Issue Type
- Issues by Content Type
- Issues by Platform
- Issues by Population
- Issues by Severity
- Conformance with the Web Content Accessibility Guidelines, version 2.1

Test Outcomes & Rate of Failure

As described in the discussion on Methodology, Tenon's testers evaluated 16 components of the Gutenberg product against approximately 200 check items. Each item in the checklist is given one of the following outcomes:

- **N/A:** This check item was not applicable for the component.
- **Pass:** This check item was applicable for the component and the component met the criteria for the check item.
- **Fail:** This check item was applicable for the component and the component *did not meet* the criteria for the check item.

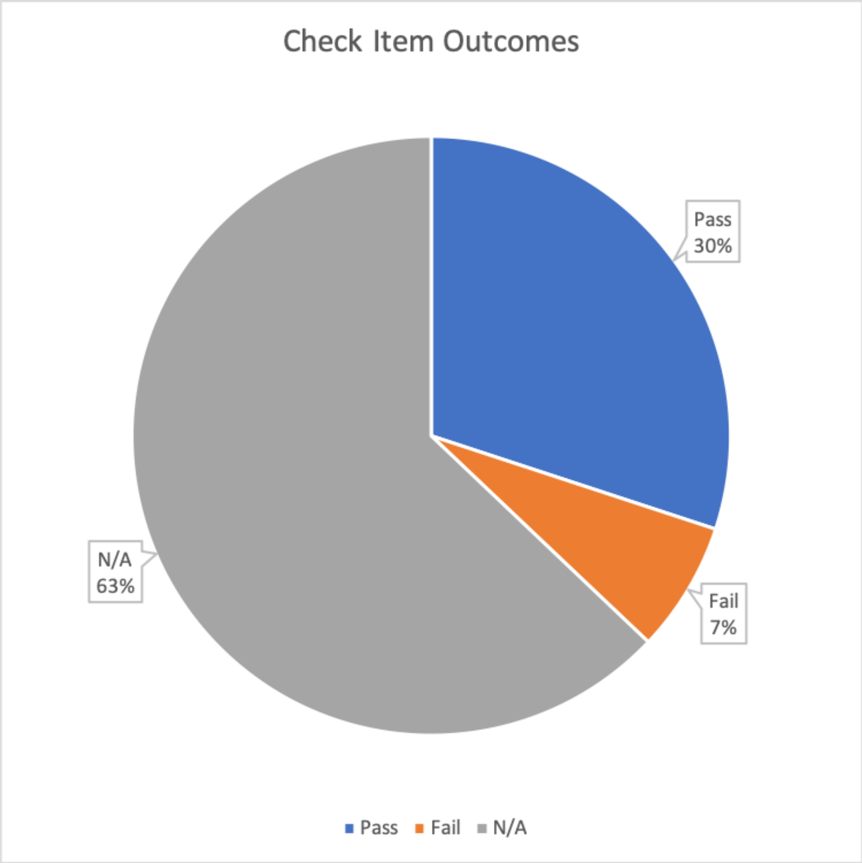


Figure 5 Chart, Check Item Outcomes

Table 2: Data, Check Item Outcomes

Result	Issue Count
Pass	892
Fail	209
N/A	1868

As the data shows, more than half of the check items were considered not applicable within Gutenberg. This includes many check items relating to multimedia content and non-text content. Depending on the component being tested, many of them will not have relevant check items for tables or forms, either. Of the relevant check items, Gutenberg had 209 items (19%) marked as failing.

Table 3: Rate of Failure Across Component

Statistic	**Percent**
Min	6%
Max	31%
Average	18%
Median	17%
Standard Deviation	0.066
Coefficient of Variation	0.367

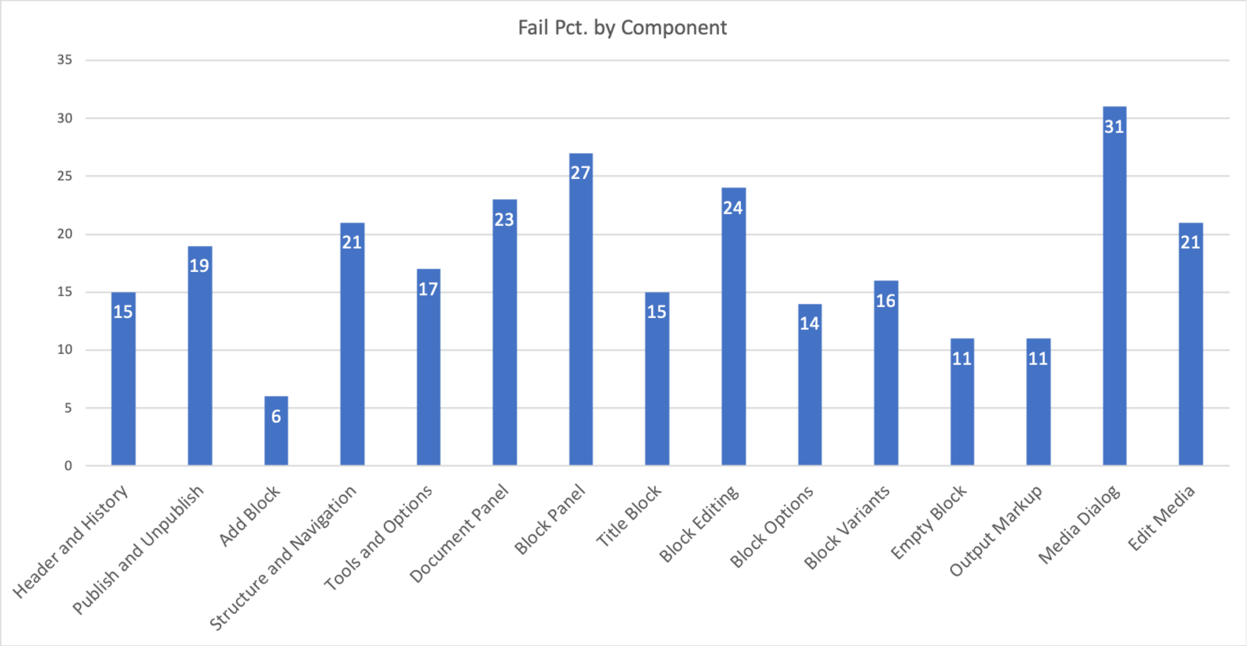


Figure 6: Chart, Rate of Failure by Component

Table 4: Data, Rate of Failure by Component

Component	Fail Pct.
Header and History	15%

Component	Fail Pct.
Publish and Unpublish	19%
Add Block	6%
Structure and Navigation	21%
Tools and Options	17%
Document Panel	23%
Block Panel	27%
Title Block	15%
Block Editing	24%
Block Options	14%
Block Variants	16%
Empty Block	11%
Output Markup	11%
Media Dialog	31%
Edit Media	21%

Issues by Component

This audit organizes the issues uncovered into 16 components: 15 components within Gutenberg and a 16th component we refer to as “Global”. The data below shows the raw issue count logged against each of these 16 components.

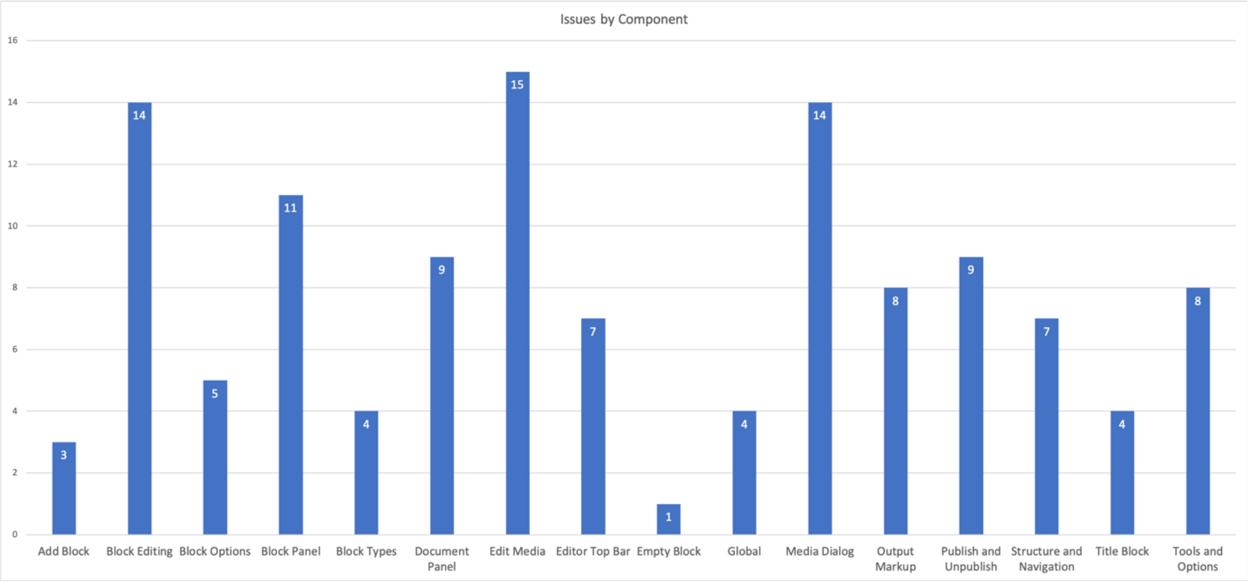


Figure 7: Chart, Issues by Component

Table 5: Data, Issues by Component

Component	Issue Count
Add Block	3
Block Editing	14
Block Options	5
Block Panel	11
Block Types	4
Document Panel	9
Edit Media	15
Editor Top Bar	7
Empty Block	1
Global	4
Media Dialog	14
Output Markup	8
Publish and Unpublish	9
Structure and Navigation	7

Component	Issue Count
Title Block	4
Tools and Options	8

Based on the data shown in the table above, Gutenberg has an average of 5.6 issues per component. This places Gutenberg in the top 30% ***worst performing*** products tested by this same team of testers.

Issues by Issue Type

Also logged within the Technical Audit are whether the issue being logged is regarded as a “Bug”, or a “Warning”. A “Warning” is a low-priority item, not explicitly mentioned in WCAG but nevertheless important enough for user experience to be mentioned in the report.   As the pie chart below indicates, 86% of the items logged in the Technical Report are considering bugs.

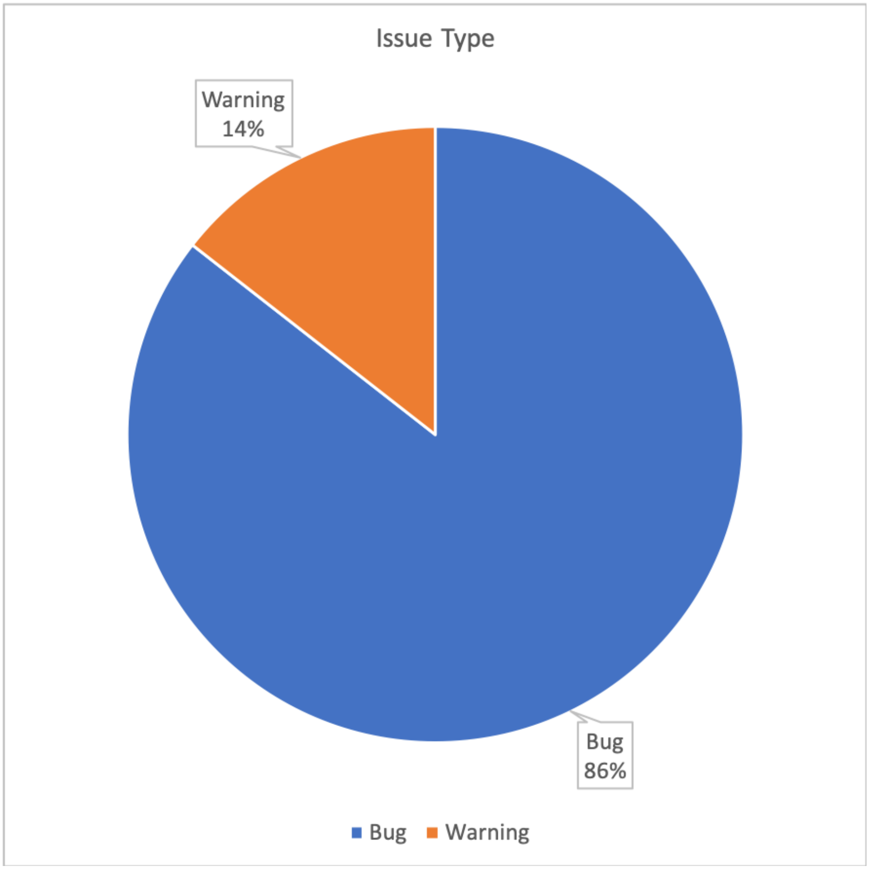


Figure 8: Chart, Issues by Issue Type

Table 6: Data, Issues by Issue Type

Issue Type	Issue Count
Bug	77
Warning	13

Issues by Content Type

Tenon's checklist separates out each item into Content Types. This provides for easier understanding of the issues and also gives additional context to the nature and cause of issues logged.

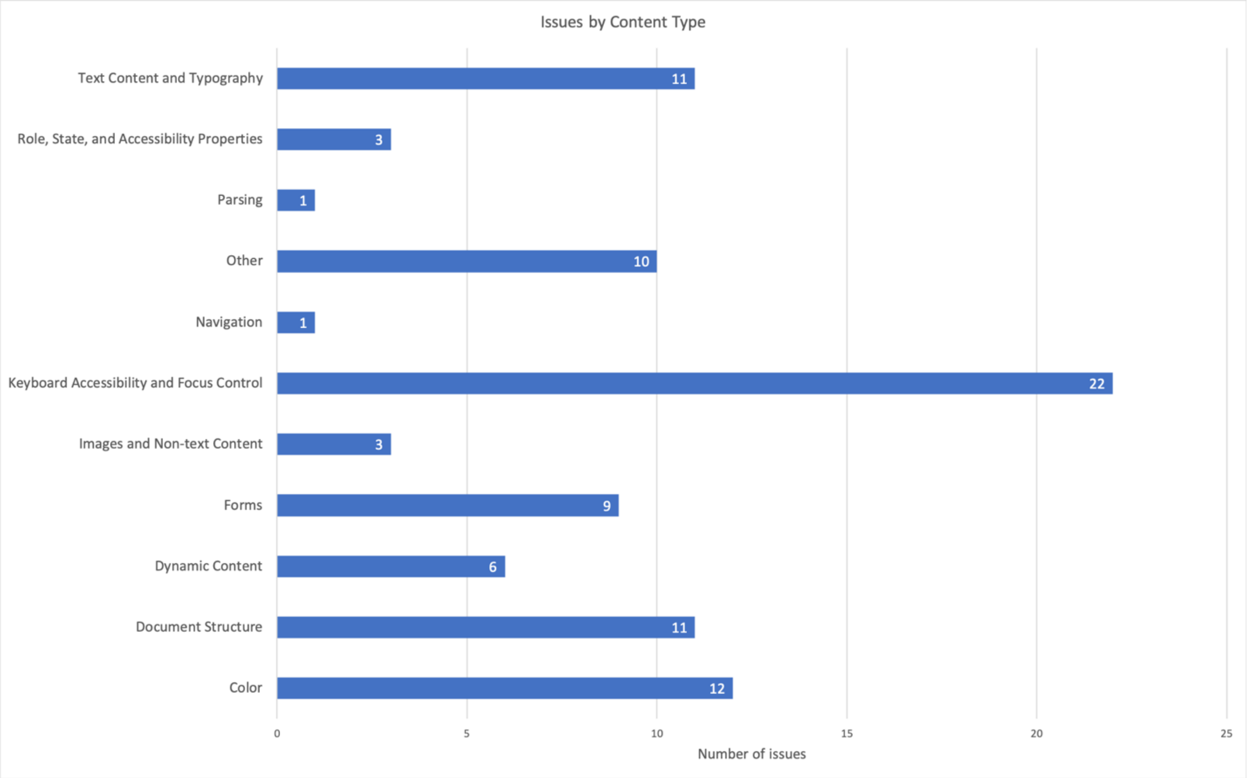


Figure 9: Chart, Issues by Content Type

Table 7: Data, Issues by Content Type

Content Type	Issue Count
Color	12
Document Structure	11
Dynamic Content	6
Forms	9
Images and Non-text Content	3
Keyboard Accessibility and Focus Control	22
Navigation	1
Other	10
Parsing	1
Name, State, Role, and Value	3
Text Content and Typography	11

Issues by Platform

Generally, Tenon does not perform testing for platform compatibility, but rather focuses on technical conformance against best practices. However, as issues are logged, Tenon does indicate which platform combination will be impacted by the issue being logged. In the vast majority of cases, the issues logged will align to “All/ Universal” because the issue’s impact is so broad that it is very likely to impact all users of all platforms. The following platforms are tracked in our reporting:

- All/ Universal
- Android w/ TalkBack
- iOS w/ VoiceOver
- macOS w/ VoiceOver
- Windows w/ Dragon Naturally Speaking
- Windows w/ native High Contrast Mode
- Windows w/ a screenreader (NVDA or JAWS)
- Windows w/ ZoomText

As noted elsewhere, issues may exist across multiple platforms, so the sum of issues-by-platform will be greater than the sum of all issues.

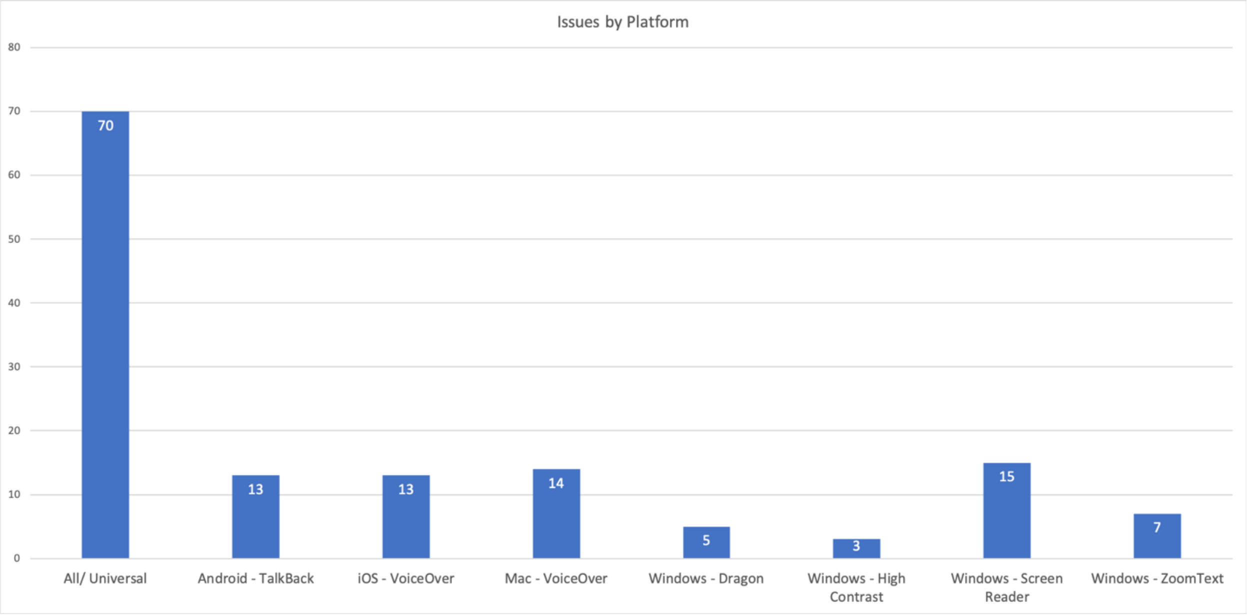


Figure 10: Chart, Issues by Platform

Table 8: Data, Issues by Platform

Platform	Issue Count
All/ Universal	70
Android - TalkBack	13
iOS - VoiceOver	13
Mac - VoiceOver	14
Windows - Dragon	5
Windows - High Contrast	3
Windows - Screen Reader	15
Windows - ZoomText	7

Issues by Population

As issues are logged by Tenon, we also take note of which user populations may be impacted by the issue. We log according to the following 6 populations:

- Blind
- Low Vision
- Cognitively Impaired
- Motor Impaired
- Hearing Impaired
- Speech Impaired

The chart & table below provides a breakdown of the issues by population. No issues were logged that affect hearing or speech impaired users, because no features under test required hearing or speech.

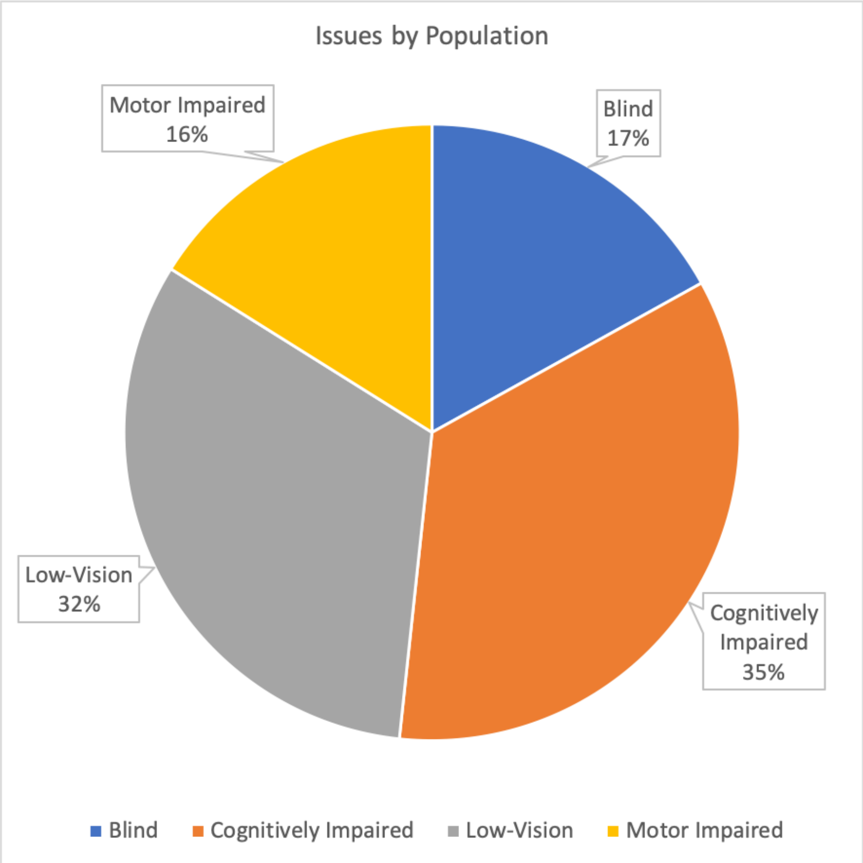


Figure 11: Chart, Issues by Affected Population

Table 9: Data, Issues by Affected Population

Populations	Issue Count
Blind	40
Cognitively Impaired	82
Low-Vision	76
Motor Impaired	38

Of particular importance when tracking issues by affected population is that many issues uncovered will impact more than one user population. For instance, issues relating to keyboard accessibility will cause problems for motor impaired users but may also impact users who are blind or low-vision.

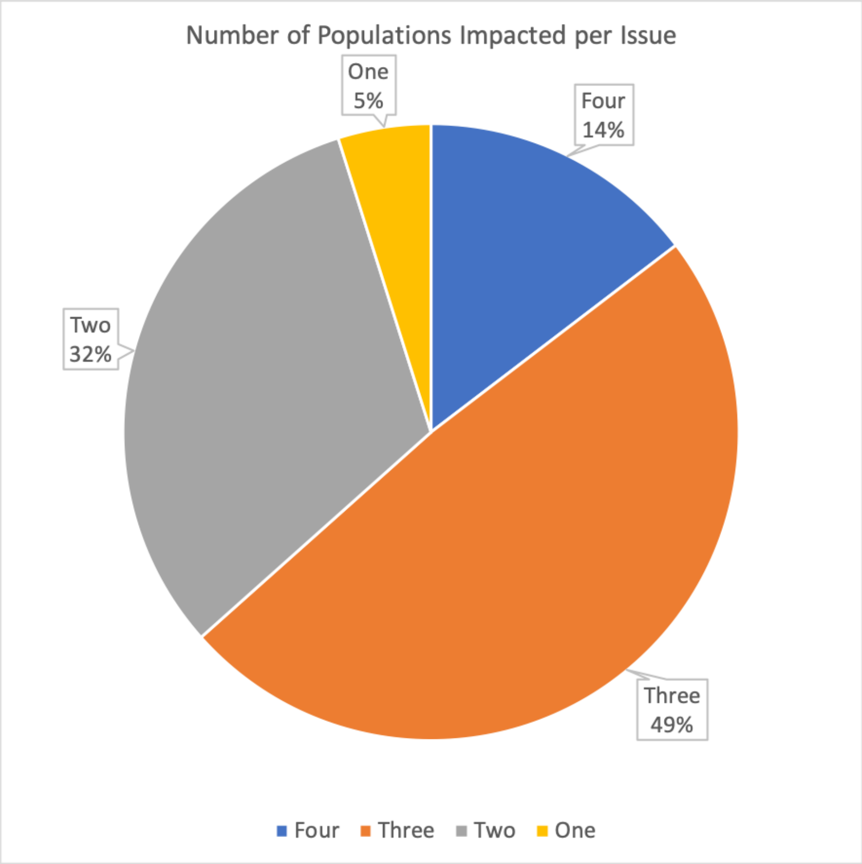


Figure 12: Chart, Number of Populations Affected per Issue

Table 10: Data, Number of Populations Affected per Issue

Number of Populations	Issue Count
Four	12
Three	40
Two	26
One	4

As shown above, only 5% of issues logged during this audit will impact only one user population, whereas nearly 50% of issues logged will impact three populations.

Issues by Severity

As issues are logged into our system, we give each item a “Severity” score as well. The severity score indicates the degree to which a user with a disability will be impacted by the issue:

- **Low:** A user with a disability will be able to complete the associated task with minor difficulties.
- **Medium:** A user with a disability will be able to complete the associated task with significant difficulty.
- **High:** A user with a disability will be prevented from completing the associated task.

As shown below, only 31% of issues are “Low” severity while 69% are “Medium” or “High”. 18% of the issues logged during this audit are likely to prevent a user from completing the associated task.

The reader should note that issues with Low severity can still create major problems their cumulative impact on usability is considered. As a consequence, we strongly recommend against disregarding issues listed as having "Low" severity, as this judgment is only in the context of the issue at the micro level and not in context of the cumulative impact across the system or its contribution to other issues on the same component.

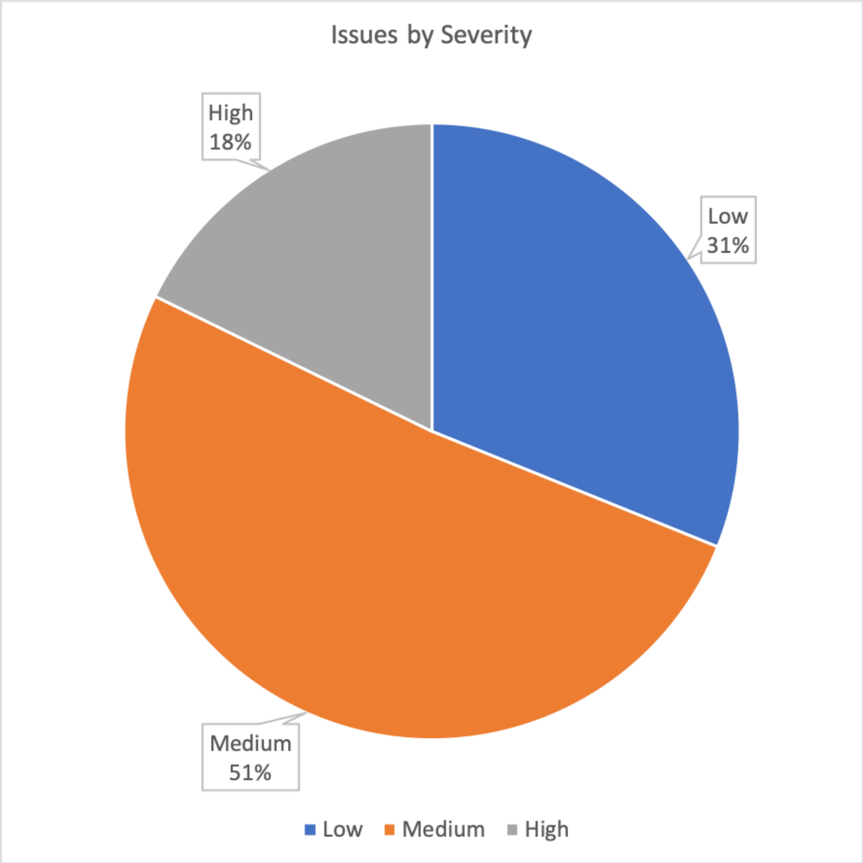


Figure 13: Chart, Issues by Severity

Table 11: Data, Issues by Severity

Severity	Issue Count
Low	28
Medium	46
High	16

Conformance with the Web Content Accessibility Guidelines, version 2.1

The Web Content Accessibility Guidelines (WCAG) is developed by the Accessibility Guidelines Working Group (AG WG), which is part of the Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C). The W3C launched the WAI in 1997 to work on guidelines, reports, and educational materials relating to Web Accessibility. The first version of WCAG was released in 1999. WCAG 2.0 was released in December 2008. WCAG 2.1 was released in 2018. WCAG 2.0 has also become an ISO standard (ISO/IEC 40500:2012). It has been translated into 28 languages and incorporated into laws in 10 countries and the European Union. Web Accessibility related lawsuits in the United States commonly use WCAG 2.x Level AA as settlement criteria.

Issues by WCAG Principle

The guidelines and Success Criteria are organized around the following four principles, which lay the foundation necessary for anyone to access and use Web content. Anyone who wants to use the Web must have content that is:

1. **Perceivable** - Information and user interface components must be presentable to users in ways they can perceive.
    - This means that users must be able to perceive the information being presented (it can't be invisible to all of their senses)
2. **Operable** - User interface components and navigation must be operable.
    - This means that users must be able to operate the interface (the interface cannot require interaction that a user cannot perform)
3. **Understandable** - Information and the operation of user interface must be understandable.
    - This means that users must be able to understand the information as well as the operation of the user interface (the content or operation cannot be beyond their understanding)
4. **Robust** - Content must be robust enough that it can be interpreted reliably by a wide variety of user agents, including assistive technologies.
    - This means that users must be able to access the content as technologies advance (as technologies and user agents evolve, the content should remain accessible)

If any of these are not true, users with disabilities will not be able to use the Web.

> Under each of the principles are guidelines and Success Criteria that help to address these principles for people with disabilities. There are many general usability guidelines that make content more usable by all people, including those with disabilities. However, in WCAG 2.0, we only include those guidelines that address problems particular to people with disabilities. This includes issues that block access or interfere with access to the Web more severely for people with disabilities.
>
> *(Introduction to Understanding WCAG 2.0)*

The chart & table below shows the orientation of issues against each of the four principles of WCAG. Over half of the issues logged are aligned with the “Perceivable” principle of WCAG, indicating that users with sensory impairments will be impacted quite a bit.

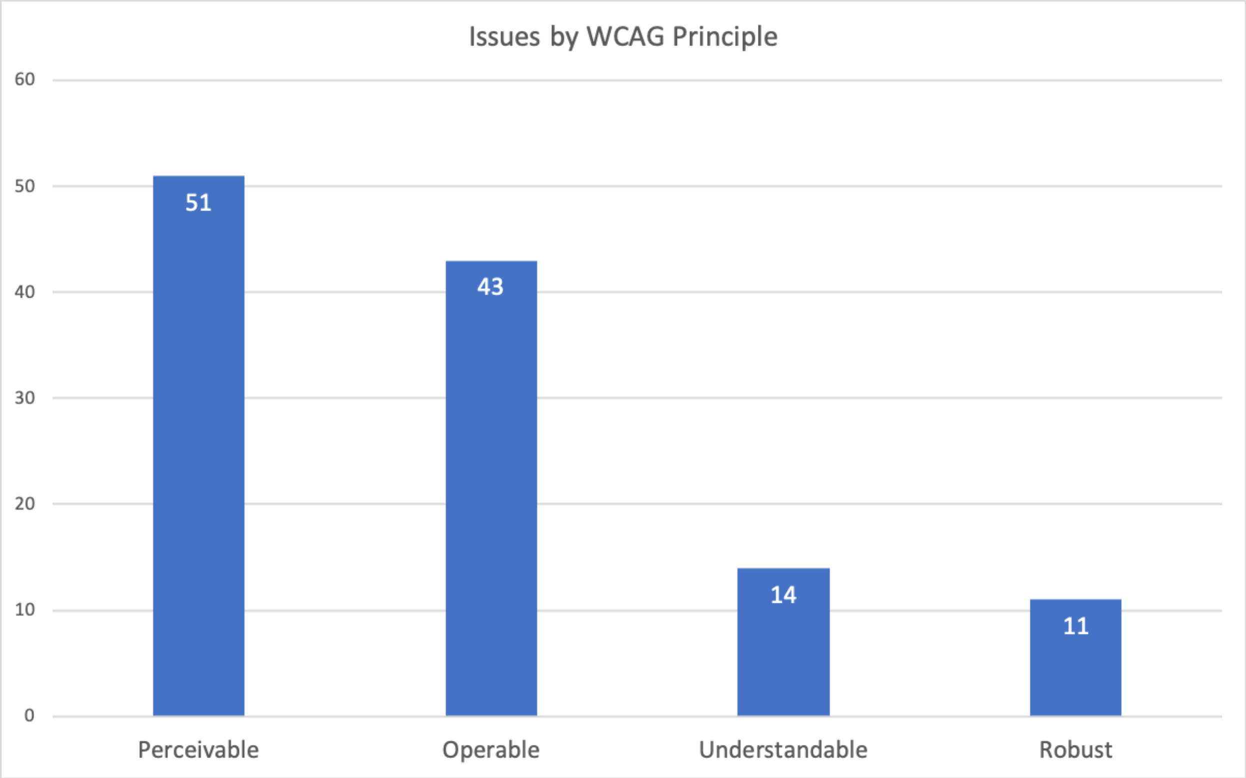


Figure 14: Chart, Issues by WCAG Principle

Table 12: Data, Issues by WCAG Principle

Principle	Issue Count

Perceivable	51
Operable	43
Understandable	14
Robust	11

Issues by WCAG Level

> For each guideline, testable success criteria are provided to allow WCAG 2.0 to be used where requirements and conformance testing are necessary such as in design specification, purchasing, regulation, and contractual agreements. In order to meet the needs of different groups and different situations, three levels of conformance are defined: A (lowest), AA, and AAA (highest).
>
> *(Understanding Conformance)*

Over 50% of the issues logged in the Technical Report are for Level A Success Criteria. 39% of the issues are for Level AA and 8% are for Level AAA. It is important for the reader to understand, however, that WCAG Level is not equal to either Severity or Priority. Although businesses and legislators often conflate these terms, Level is a bit more nuanced.

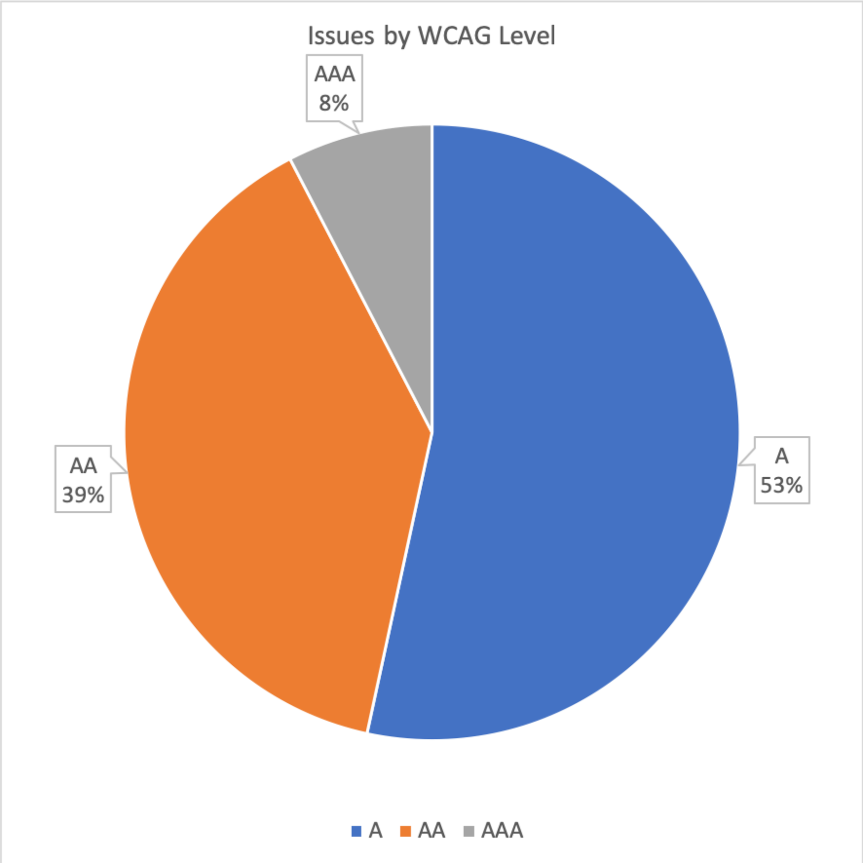


Figure 15: Chart, Issues by WCAG Level

Table 13: Data, Issues by WCAG Level

WCAG Level	Issue Count
A	63
AA	46
AAA	9

The very large number of Level A issues helps to illuminate the significant problems that users with disabilities will face when attempting to use Gutenberg. It also brings to light the potential legal and compliance risk that may arise from using Gutenberg.

Issues by WCAG SC

> Under each principle there is a list of guidelines that address the principle. There is a total of 12 guidelines. A convenient list of just the guidelines can be found in the WCAG 2.0 table of contents. One of the key objectives of the guidelines is to ensure that content is directly accessible to as many people as possible, and capable of being re-presented in different forms to match different peoples' sensory, physical and cognitive abilities.
>
> Under each guideline, there are Success Criteria that describe specifically what must be achieved in order to conform to this standard. They are similar to the "checkpoints" in WCAG 1.0. Each Success Criterion is written as a statement that will be either true or false when specific Web content is tested against it. The Success Criteria are written to be technology neutral.
>
> All WCAG 2.0 Success Criteria are written as testable criteria for objectively determining if content satisfies the Success Criteria. While some of the testing can be automated using software evaluation programs, others require human testers for part or all of the test.
>
> *(Introduction to Understanding WCAG 2.0)*

The following chart and table list each of the WCAG Success Criteria for which issues were logged. Gutenberg fails 30 to comply with 30 of the WCAG 2.1 Success Criteria. The top 4 most violated Success Criteria are:

- 1.3.1 Info and Relationships (Level A)
- 1.3.3 Sensory Characteristics (Level A)
- 2.1.1 Keyboard (Level A)
- 2.4.7 Focus Visible (Level AA)

This data aligns quite closely with the results from the usability testing.

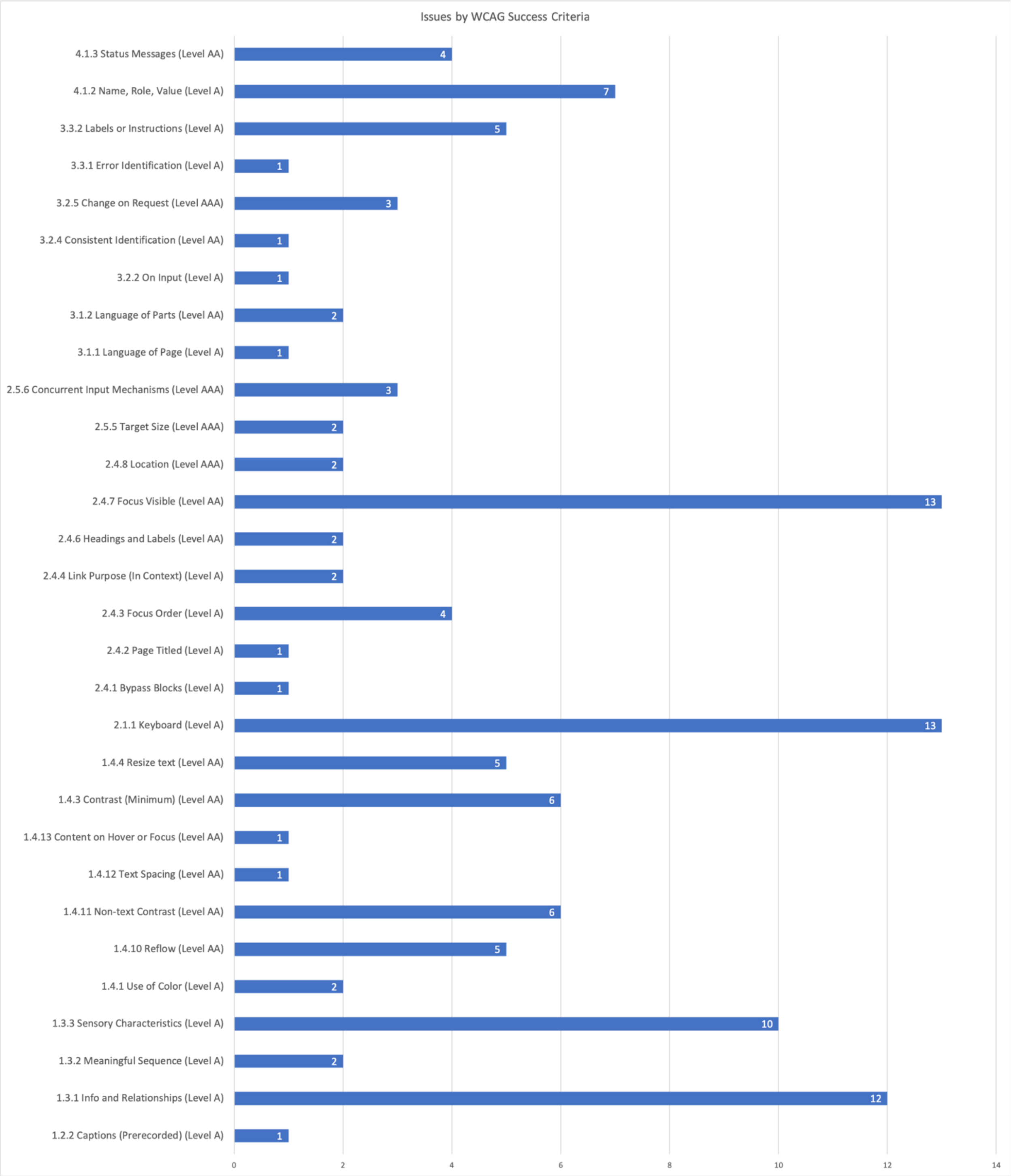


Figure 16: Chart, Issues by WCAG Success Criteria

Table 14: Data, Issues by WCAG Success Criteria

WCAG Success Criteria	Issue Count
1.2.2 Captions (Prerecorded) (Level A)	1
1.3.1 Info and Relationships (Level A)	12
1.3.2 Meaningful Sequence (Level A)	2
1.3.3 Sensory Characteristics (Level A)	10
1.4.1 Use of Color (Level A)	2
1.4.10 Reflow (Level AA)	5
1.4.11 Non-text Contrast (Level AA)	6
1.4.12 Text Spacing (Level AA)	1
1.4.13 Content on Hover or Focus (Level AA)	1
1.4.3 Contrast (Minimum) (Level AA)	6
1.4.4 Resize text (Level AA)	5
2.1.1 Keyboard (Level A)	13
2.4.1 Bypass Blocks (Level A)	1
2.4.2 Page Titled (Level A)	1
2.4.3 Focus Order (Level A)	4
2.4.4 Link Purpose (In Context) (Level A)	2
2.4.6 Headings and Labels (Level AA)	2
2.4.7 Focus Visible (Level AA)	13
2.4.8 Location (Level AAA)	2
2.5.5 Target Size (Level AAA)	2
2.5.6 Concurrent Input Mechanisms (Level AAA)	3
3.1.1 Language of Page (Level A)	1
3.1.2 Language of Parts (Level AA)	2
3.2.2 On Input (Level A)	1
3.2.4 Consistent Identification (Level AA)	1

WCAG Success Criteria	Issue Count
3.2.5 Change on Request (Level AAA)	3
3.3.1 Error Identification (Level A)	1
3.3.2 Labels or Instructions (Level A)	5
4.1.2 Name, Role, Value (Level A)	7
4.1.3 Status Messages (Level AA)	4

Conclusion and Recommendations

There is no question that WordPress has attained its goal of democratizing publishing. This is shown by its incredible popularity, as WordPress is used by nearly 60% of all the websites whose content management system is known. This popularity offers a unique leadership opportunity, especially when it comes to accessibility. The ease with which developers can create fully custom themes means that a sufficiently skilled developer can create a highly accessible experience for end users. Unfortunately, the Gutenberg editor presents considerable challenges for persons with disabilities who choose WordPress as their publishing platform.

Gutenberg has significant and pervasive accessibility problems, the likes of which amount to a step backwards for users with disabilities over the legacy editor. Our user-based testing – backed by data from our technical review – indicates that the accessibility problems are severe in nature. We feel concerned that Gutenberg's current accessibility issues will prove problematic for website owners who deploy Gutenberg to content creators in protected populations or for website owners who are themselves part of a protected population. Therefore, organizations which have high risk profiles should consult legal counsel before using it and may want to choose to use the legacy editor instead.

Based upon the size of the user base for WordPress, the issues uncovered in our Technical Report should be addressed aggressively by the developers of Gutenberg. The detailed guidance provided in our report should make addressing each issue very straight forward for sufficiently skilled developers.