Library Accessibility Alliance Impact and Analysis Report 2024

Executive Summary

Data analysis of 132 accessibility testing reports commissioned by the Library Accessibility Alliance shows high rates of nonconformance with WCAG AA success criteria for e-resource platforms licensed by libraries. Platform developers should strive for high conformance, and librarians should take accessibility conformance into consideration when deciding to license an eresource.

Analysis of 132 testing reports shows that no platforms tested were fully in conformance with the WCAG AA standard, and some WCAG success criteria were in nonconformance in up to 93% of tested platforms. Some platforms in the sample were tested more than once, and retested platforms generally improved between test and retest. Additionally there is a general trend towards greater conformance over time, however some specific criteria also show increases in nonconformance. The average eresource platform showed nonconformance in 26% of WCAG AA criteria.

Introduction/Background

The Library Accessibility Alliance (LAA) is a multi-consortial organization dedicated to improving the accessibility of libraries and library content. One of the ways that LAA accomplishes this goal is by providing accessibility testing of eresource content providers. The testing committee identifies platforms frequently used by library systems to provide digital content to their patrons, and partners with testing vendors to provide reports on the accessibility of the resource. The testing vendors conduct automated and manual accessibility testing using screen readers. Each platform is evaluated based on Web Content Accessibility Guidelines (WCAG) standards. These accessibility audits aim to provide a high-level accessibility review of the product by performing primary tasks on the homepage, the search and search result pages, and an individual record. Each test results in a report, which identifies areas of nonconformance and the corresponding WCAG criteria that needs to be adjusted for the issue to be resolved. Testing is not comprehensive and will not identify all instances of nonconformance, but gives a high-level overview of the issues an end user is likely to encounter. All testing reports are added to a publicly available database.

The Web Content Accessibility Guidelines (WCAG) provide a set of best practices for making web content accessible for people with a range of auditory, visual, physical, cognitive, speech,

and neurological disabilities. While the Guidelines cannot address all requirements or preferences for every disability, they allow for improved usability for many. Guidelines are organized according to four principles: perceivable, operable, understandable, and robust. Each guideline encompasses testable success criteria organized into three levels: A, AA, and AAA. (source: https://www.w3.org/WAI/standards-guidelines/). Criteria that are ranked at the A and AA levels are considered more essential to providing an accessible product. Most eresource contracts require vendors to comply with the latest version of WCAG at the AA level, meaning that the product is in conformance with all A and AA success criteria.

In 2021, the LAA Impact and Analysis committee undertook a quantitative analysis of the reports then available. This analysis resulted in a white paper that identified trends in the WCAG violations identified in the testing reports. The current report expands that analysis to two additional years of testing reports, and builds on the data analysis of the LAA testing reports.

Methodology

The Impact and Analysis group incorporated the data from the the 58 testing reports included in the 2021 Whitepaper, and added data from 74 additional test reports conducted between 3/27/2020 and 8/8/2023 (covering all testing from the beginning of LAA testing in 2017 through 8/8/2023), for a total sample of 132 reports. For each testing report, the group recorded each WCAG criterion that was flagged in the report, as well as metadata about the report including whether the report was conducted based on WCAG 2.0 or 2.1. Note that data coding of the reports only recorded the first instance of a WCAG criterion being flagged in the report - so the data captures the breadth of WCAG criteria that had violations, but not the overall volume of violations for that product. If a report notes, for example, that there are violations of WCAG 1.4.3 (insufficient color contrast) on multiple webpages or multiple violations on the same page, that is coded in the data as a binary “yes” for issues found with 1.4.3, rather than counting the number of instances where insufficient color contrast was found.

LAA testing reports are not comprehensive, and are a snapshot in time. The time that the testing vendors spend on each assessment is also limited, and many reports note that more advanced or manual testing could not be completed within the allotted time. Additionally, LAA’s testing vendors are only looking for WCAG violations up to the AA level, although a AAA issue will be included if it happens to be seen by a tester.

Testing up to the AA level presented new challenges in this dataset relative to the 2021 data, because some tests occurred under WCAG version 2.0 and some under 2.1. Criteria that were added in WCAG 2.1 obviously were not checked for in testing done under 2.0, making comparison between tests conducted under the different versions difficult. To account for these variations we recorded three possible values for each WCAG criterion: “issues found, “no issues found,” or “no data.” All AAA criteria were coded as “no data” unless a criterion violation was affirmatively identified in the test. Additionally, criteria that were added in WCAG 2.1 were coded as “no data” if the testing occurred under 2.0. Percentages for the rate of issues found in

a given test excluded criteria for which there was no data, so as not to skew the results by conflating absence of evidence with evidence of absence.

This dataset also included retest data. In 2021[1] LAA began retesting products to assess updates and progress in the product and to continue to provide relatively current data for each product. The dataset includes 18 retests. These 18 retests were included in the general dataset, and were also paired with their prior test in order to assess the change over time in a given product.

The full dataset can be found at https://doi.org/10.5281/zenodo.15376295.

Results

Many WCAG criteria had a high percentage of tests reporting violations. 80% of A and AA level criteria had violations in at least one platform.[2] 83% of A level criteria had violations in at least one platform.

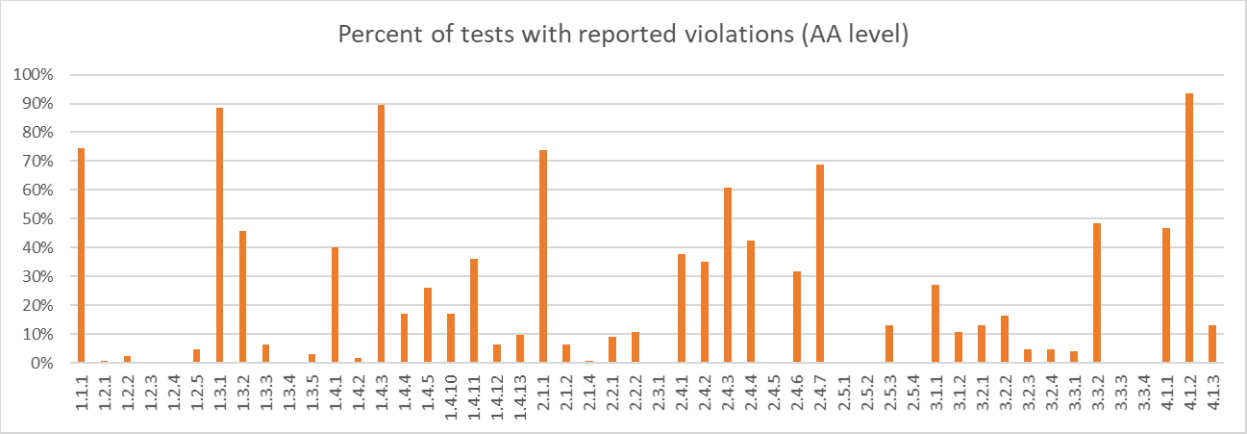

Figure 1: The percent of testing reports showing nonconformance with each A and AA level WCAG criterion.

The most common violations were in the following criteria:

Most common violations (20% frequency or higher)				
WCAG criterion	WCAG label	level	Number of tests reporting violations on	Percent of tests reporting violations on

[1] With an outlier retest in 2019.
[2] 59% of WCAG criteria across all levels had reported violations, however we do not have data on AAA criteria unless a violation was noticed and specifically reported. Lack of a violation in AAA criteria should therefore not be interpreted to mean that no issues were present in those criteria - it just indicates that those criteria were not tested.

			this criterion	this criterion
4.1.2	Name, Role, Value	A	114	93%
1.4.3	Contrast (Minimum)	AA	109	89%
1.3.1	Info and Relationships	A	108	89%
1.1.1	Non-text Content	A	91	75%
2.1.1	Keyboard	A	90	74%
2.4.7	Focus Visible	AA	84	69%
2.4.3	Focus Order	A	74	61%
3.3.2	Labels or Instructions	A	59	48%
4.1.1	Parsing	A	57	47%
1.3.2	Meaningful Sequence	A	56	46%
2.4.4	Link Purpose (In Context)	A	52	43%
1.4.1	Use of Color	A	49	40%
2.4.1	Bypass Blocks	A	46	38%
1.4.11	Non-text Contrast	AA	44	36%
2.4.2	Page Titled	A	43	35%
2.4.6	Headings and Labels	AA	39	32%
3.1.1	Language of Page	A	33	27%
1.4.5	Images of Text	AA	32	26%

Table 1. The WCAG criteria that had the highest frequency of nonconformance. All criteria with 20% or higher nonconformance across testing reports are included.

The rate of WCAG violations at the A and AA level is unfortunate. Most vendors agree to contract terms obligating them to maintain conformance with WCAG AA 2.1, or have released accessibility statements committing to attaining this standard. The evidence from this data shows that actually achieving this goal is difficult and that these accessibility promises have largely not been met. Accessibility is an evolving goal, and testing reports are a tool that helps vendors know where they can improve, but clearly improvements are necessary and the list above suggests areas in need of particular focus.

A brief explanation of the top 5 problematic success criteria may help convey how failure to meet these criteria presents significant barriers for people with disabilities:

- 4.1.2 Name, Role, Value (A): Ensures that user interface components have names and roles that can be programmatically determined. Example: if a drop down list on a page is not announced as such to a screen reader user they are unlikely to be able to use the form.
- 1.4.3 Contrast (Minimum) (AA): Requires a minimum contrast ratio of 4.5:1 between text and background colors to ensure readability for users with visual impairments. Example: a person with color vision deficiency (color blindness) may be unable to identify important links on a page if color is the only means of identifying the link when nested within text of a different color and the person cannot perceive a difference in the colors. Meeting minimum color contrast would ensure the text of links is discernible.
- 1.3.1 Info and Relationships (A): Requires that information, structure, and relationships conveyed through presentation can be programmatically determined or are available in text. Example: A page that includes obvious headings formatted with large text and a bold treatment are not programmatically coded as headings using HTML heading elements. A screen reader user would be unable to easily distinguish between headings and paragraph text.
- 1.1.1 Non-text Content (A): Ensures that all non-text content has a text alternative that serves an equivalent purpose. Example: An image on a page containing important information and lacking alternative text would not be perceivable by a person who is deafblind and using a Braille display.
- 2.1.1 Keyboard (A): Mandates that all functionality is operable through a keyboard interface. Example: Someone using a switch device is unable to activate a form control necessary for completing an important task.

This list is comparable to the criteria found to most frequently be at issue in the 2021 Whitepaper. All of the criteria cited in 50% or higher of the tests examined in 2021 still appear in the table above. A comparison of the criteria reported on in the 2021 report and the 2023 data can be found below:

Comparison of WCAG criterion violation frequency between 2021 and 2023 data

			2021		2023	
WCAG criterion	WCAG label	level	Number of tests reporting violations on this criterion	Percent of tests reporting violations on this criterion	Number of tests reporting violations on this criterion	Percent of tests reporting violations on this criterion
1.3.1	Info and Relationships	A	50	86%	108	89%
1.4.3	Contrast (Minimum)	AA	49	84%	109	89%
2.4.3	Focus Order	A	48	83%	74	61%

2.1.1	Keyboard	A	48	83%	90	74%
1.1.1	Non-text Content	A	47	81%	91	75%
4.1.2	Name, Role, Value	A	46	79%	114	93%
2.4.7	Focus Visible	AA	46	79%	84	69%
2.4.4	Link Purpose (In Context)	A	30	52%	52	43%

Table 2. Comparison data showing the frequency of nonconformance in the 2021 and 2023 datasets for the criteria showing 50% or greater nonconformance in 2021.

The following graph shows the change in frequency of the top eight criteria from 2021 between the 2021 and 2023 datasets.

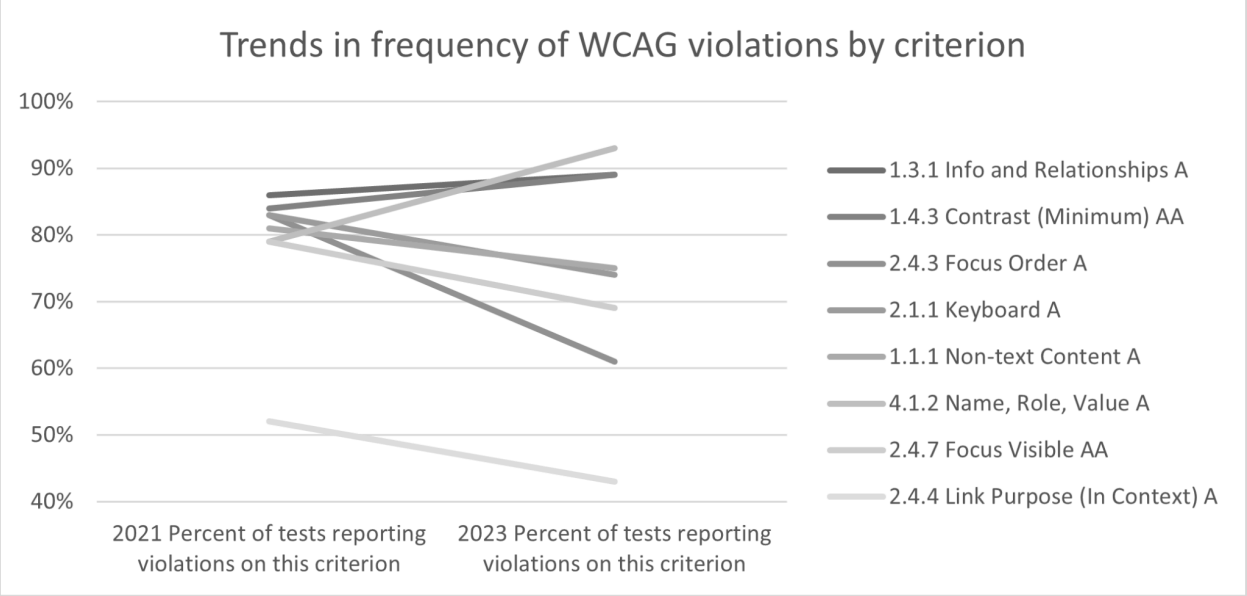

Figure 2. A line graph showing the frequency of WCAG violation by criterion in the 2021 dataset and the 2023 dataset for the criteria with the eight highest nonconformance rates.

Three criteria have increased in violation frequency (1.3.1, 1.4.3, and 4.1.2), suggesting particular areas for product developers to focus on to improve product accessibility. These three criteria were also the most consistently problematic in the 2023 data. Most dramatically, the number of tests reporting issues under 4.1.2 has increased from 79% to 93% to become the most frequently cited issue in the 2023 data. This increase may be attributed to increasing complexity of web applications and the rise of JavaScript frameworks being used by developers to create advanced user interfaces that often do not implement appropriate accessibility features due to lack of awareness and training.[3] Failure of this criterion is particularly impactful

[3] See for example, Claire Kearney-Volpe and Amy Hurst. 2021. Accessible Web Development: Opportunities to Improve the Education and Practice of web Development with a Screen Reader. ACM Trans. Access. Comput. 14, 2, Article 8 (June 2021), 32 pages. https://doi.org/10.1145/3458024; Martins,

to low vision, blind and deafblind people because it often results in interface controls being completely unusable to this population.

There are, however, indications that overall WCAG violations are decreasing over time. Figure 3 shows a mild trend towards fewer WCAG violations the more recent the testing.

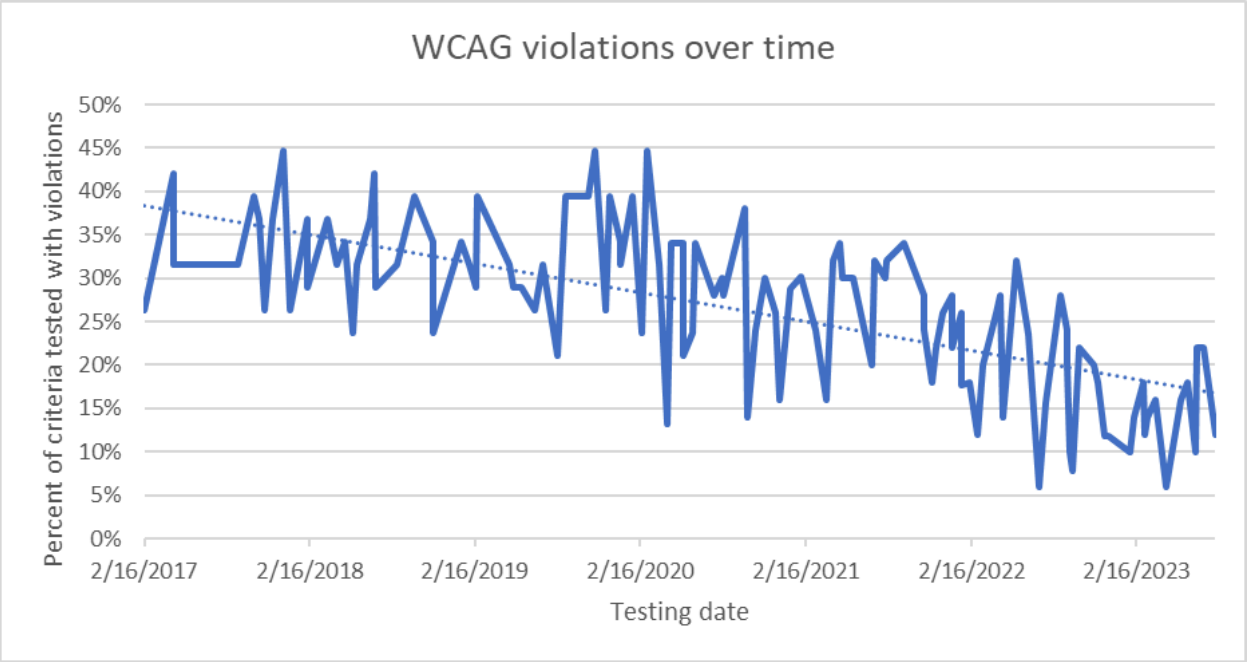

Figure 3. A line graph of WCAG nonconformance rates by testing date, showing a general trend toward fewer issues over time. (Data is shown from the beginning of LAA testing, 2/16/17 through the end of data collection for this project, 8/8/23; one later test from 5/27/24 was accidentally coded and included in the dataset, but is omitted here for visual clarity.)

Additionally, the data from the 18 products that have been retested show that the majority of retested products have improved between test and retest.

B., Duarte, C. A large-scale web accessibility analysis considering technology adoption. Univ Access Inf Soc 23, 1857–1872 (2024). https://doi.org/10.1007/s10209-023-01010-0.

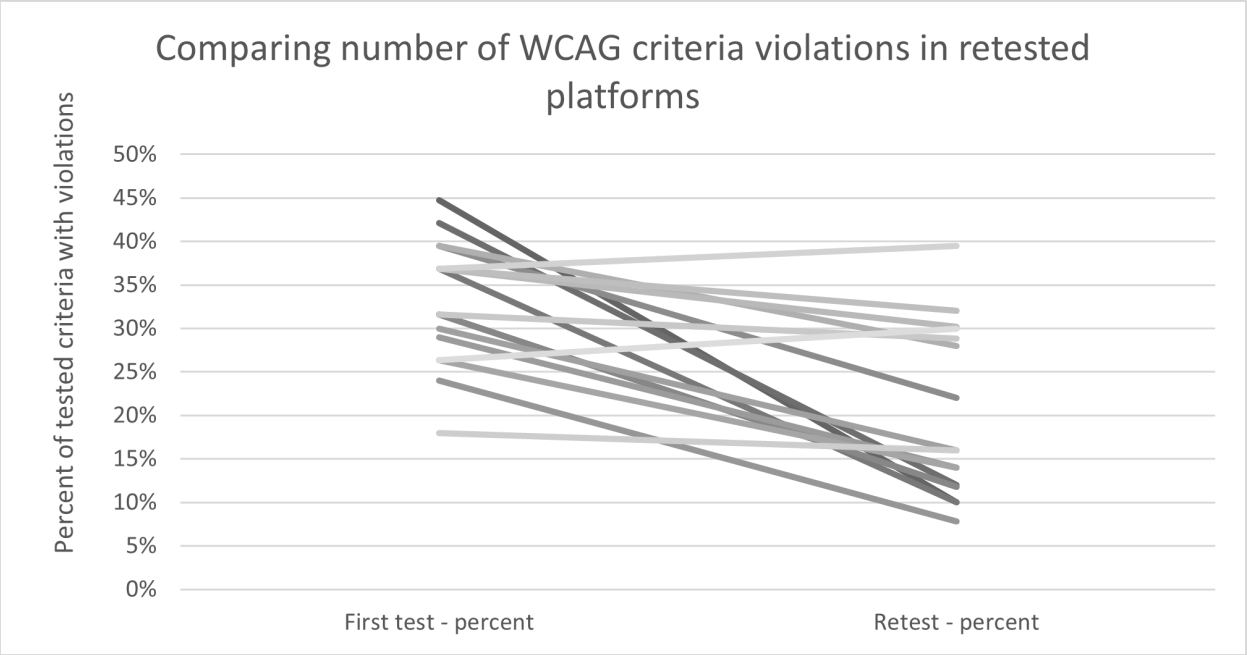

Figure 4. Trendline comparison showing platform performance in retested products. For all products that were retested, the percent of WCAG criteria nonconformance in test and retest are shown.

Products that showed the most improvement between test and retest were: Sage ejournals (45%, down to 10%), Springerlink (42% to 12%), JSTOR (27% to 10%), Cambridge Core (20% to 12%), and Elsevier ScienceDirect (32% to 12%).

A greater period of time between the test and retest was also significantly correlated with a greater difference in test and retest results ($r = .630$, $p = .005$).

Product by product results show that the average product tested had violations in 26% of the WCAG AA criteria that they were tested for. Products ranged from issues noted in 45% of WCAG AA criteria down to 6%.

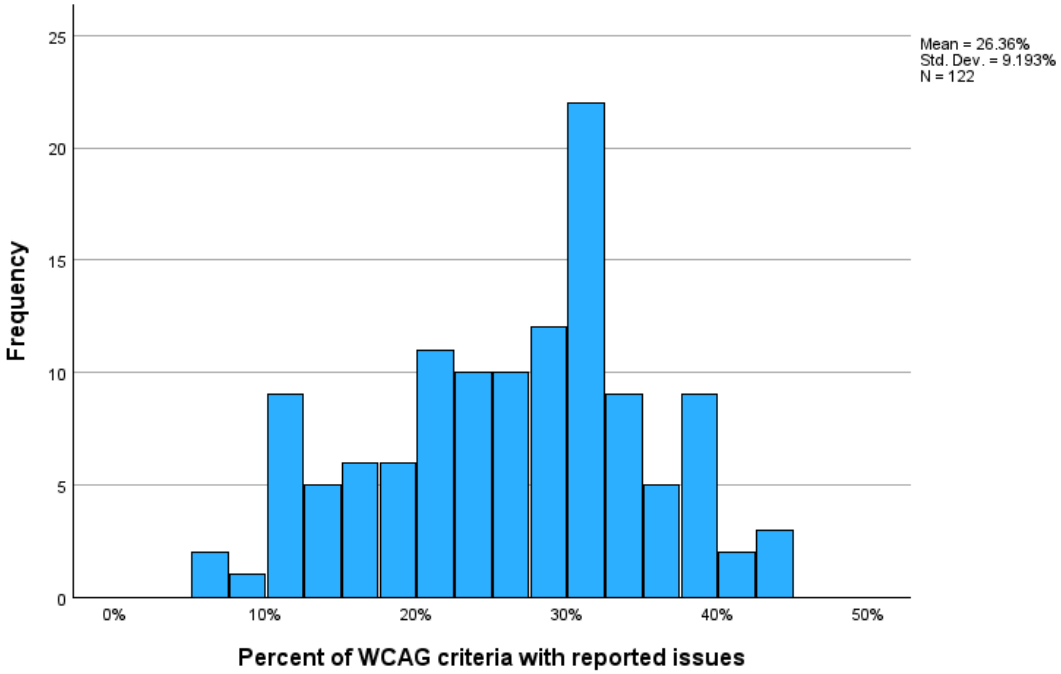

Figure 5. A frequency histogram of the percent of nonconforming WCAG criteria by platform. The average number of nonconforming criteria across platforms was approximately 26%.

Product-specific results can be found in the associated data set at https://doi.org/10.5281/zenodo.15376295, or by checking the product in the LAA testing database.

No product tested was completely WCAG compliant. Products with WCAG violations in less than 10% of the WCAG criteria that they were tested for are:

Product	First test or retest	Percent of WCAG criteria tested that had criteria violations
Africa Commons	First test	6%
Emerald Insights	First test	6%
Walter de Gruyter	retest	8%
Sage ejournals	retest	10%
JSTOR	retest	10%
New Play Exchange	First test	10%
OverDrive	First test	10%

Table 3. Platforms with the fewest instances of WCAG nonconformance.

It seems notable that half of this list are retested products, and that the WCAG violations in the earlier tests for each of these products were significantly higher (24%, 45%, and 37%, respectively). This reflects well on these products and suggests a concerted and largely successful effort has been made to improve accessibility on these products. The data on all tested vendors is available at https://doi.org/10.5281/zenodo.15376295.

Conclusions

It is encouraging that re-test data and testing over time show improved accessibility support over time. Most of the platforms that have been tested more than once show improvement between tests, and more recent tests tend to show better results than older tests. This suggests a general trend of increased support and attention to accessibility concerns in eresource platform development. At the same time, increasing nonconformance within specific success criteria (1.3.1, 1.4.3, and 4.1.2) shows areas in need of further support. More work and subsequent testing is needed to ensure continued improvement.

It's important to note that the focus of LAA's testing efforts to date have been on the user interfaces of the platforms, and not the accessibility of the content itself (articles, media, or other materials that are part of the platform offering). Some of the reports did capture this information, but most did not, so it is not possible for us to generalize about the accessibility status of content from the reports. We do believe that in general platforms that offer content, particularly content in PDF format, do not provide that content in an accessible format. For PDF, for example, that would mean delivering PDFs that are compliant with WCAG 2.1 AA and the PDF-UA standards. Again, the reporting to date has not consistently verified that content like PDFs are meeting those standards, but based on the experience of LAA members, in general platforms fall far short of these standards for content. Platforms that aggregate content from other publishers often do not take responsibility for assuring the accessibility of that content, but library contractual relationships require that all such content also complies with accessibility standards. Platforms should consider including accessibility clauses in their agreements with content providers.

Overall, while the data shows improvement over time, the overwhelming finding remains that eresource platforms struggle to meet web accessibility guidelines. No platform in the testing sample was free of conformance issues, and some WCAG criteria were nonconforming in up to 93% of the tested platforms. As the ADA Title II changes come into effect, mandating that resources licensed by public institutions comply with WCAG 2.1 at the AA level, it will become even more imperative that library vendors meet this standard. Eresource vendors and libraries should continue to prioritize accessibility, and clearly there is still progress to be made.