

IBM comments on the WCAG 3 FPWD

Thank you for all of the work you have put into the WCAG 3 draft. Coming up with a different approach to accessibility requirements, making them easier to understand and addressing the needs of many different types of consumers of a standard is a difficult task. We applaud your efforts in working toward a model where efforts of developers and authors of web content and software applications can have a single source for accessibility requirements and guidance.

IBM has documented our thoughts and concerns in several topical areas, noted in level 2 headings below. We have also incorporated responses to some of the questions you posed as that apply to the topical areas, as noted. If you have any questions regarding our comments, please contact me directly.

Mary Jo Mueller

Processes

Adopting a concept of **processes** for testing and validation is sound. It reflects the reality of how many organizations currently use sampling for testing and automated regression testing. This gives authors the freedom to identify processes they believe are aligned with primary user outcomes on their web site/application.

Scoring results against processes in addition to sites/pages/views holds the potential for meaningful granularity in results that would allow for a better assessment of accessibility and usability.

There seem to be opportunities for explaining and incorporating process-based information:

- More clarity is needed in the definition of processes, identifying what is and what is not part of a process. It appears that the draft uses the term process in regard to *user* process. Is the intention to focus on the user's journey through the content? If so, we recommend changing the term to "user process." If there are other types of processes, be sure to differentiate the term in those cases.
- WCAG 3 could identify common recommended processes to test, perhaps based on the type of content.
- **Blog question: Is this approach of using complete processes as the smallest unit of conformance workable for different types of organizations? These include organizations with very large, dynamic, or complex content, medium-sized organizations relying on external accessibility resources, and very small organizations with limited resources?**

Response: While this approach is reasonable as a unit of conformance, we are unsure of how this will work when reporting conformance of a product in a procurement or bid situation. How would these tested processes map into the documentation of product accessibility in an Accessibility Conformance Report? Would a product need to

document which processes were tested to give an accurate representation of conformance? (See prior comments for possible approaches.) In cases of large-scale applications with robust functionality, choosing the right processes to test could pose some challenges, as the most-used processes vary depending on the user's role when using the application. For example, a timecard application may have different views/functionality available to the administrator, regular employee, managers, human resources, payroll, finance, etc.

Precision and testability

- The spec needs to be more explicit on the guidelines, outcomes and methods whether this is a user agent requirement, operating system requirement, authoring tool requirement, or author requirement.

For example the Editor's Note in 1.1 about WCAG 3.0 guidelines includes an example method from UAAG ([Reflow of captions and other text in context](#)) where it is not clear who (the author, platform owner, browser owner, combination of stakeholders) is responsible for meeting the 7.3 Captions guideline. The standard also needs to address how to meet guidelines that is more explicit of the responsibility or coordination between platform owners (operating system vs user agent).

- There should also be an easy way to "gather" requirements/methods based on their applicability to the user agent, operating system, authoring tool, or authored content. This will help each different application of the standard be easily obtained without having to peruse through every guideline, outcome, and method to find what is applicable or how to do it.
- Has the concept of "**accessibility supported**" introduced in WCAG 2.1 been removed? Is there a similar concept in WCAG 3? If so, it needs to be more clearly articulated.
- **Blog question: We would like constructive feedback on the testing approach, and examples of why you would or would not implement it in your organization.**

Response: Without automated ways to identify, count and test instances of content applicable to an outcome and aggregate results for each test, it will be difficult to accurately score compliance to each objective in a repeatable way – especially for large-scale websites and applications. This will be especially problematic if the outcomes are not defined using readily measurable, testable parameters.

Any single "critical error" results in a score of "very poor" for that outcome and also results in an inability to reach bronze level as an overall score. This seems unreasonable for large-scale websites and applications and negates all of the work that may have been done to meet the outcome. There should be a very narrow scope for what are deemed "critical errors" and the percentage score should prevail to mathematically calculate the overall rating for each functional category.

- For the critical errors documented in the current draft, it takes a person's judgement whether or not an issue is considered critical and will likely lead to inconsistencies between reviewers/testers and inconsistent final scores. A more precise definition of "[critical error](#)" is needed. The list should only include those guidelines from WCAG 2.x Level A that cause harm (e.g., flashing). There should be limited allowances for 3rd party content, as an author may have no ability to correct the accessibility of such content – especially if it is protected under copyright laws. Critical errors should **not** include any vague criteria that is subjective such as the current example of a "large amount of confusing, ambiguous language". What is considered a large amount? What is considered confusing or ambiguous language?
- With the expansion of WCAG 3 to cover more broad requirements and elimination of Levels A, AA, and AAA for criteria, it is difficult to imagine that any website or application (except for small, non-complex web pages with non-technical content) would be able to meet the Bronze level of conformance. This is especially true if the current bar is a score of 3.5 for EVERY functional category with no critical errors anywhere. (See the Scoring and Conformance section in these comments for more information.)

Not achieving a Bronze level of conformance that demonstrates some significant level of accomplishment will dissuade teams from performing any holistic testing – an unintended consequence and is not consistent with the intended goal of WCAG 3 to enable teams to be able to show some level of accomplishment.

Scoring and Conformance

IBM applauds the working group's [intent of moving away from a "no-defects" conformance](#) model to have conformance better able to reflect real-world accessibility, including making allowances for bugs and other low-impact issues. However, we think there is still work to do in this area.

- One way of adding granularity into the scoring and conformance reporting system could be to provide reporting by functional category. Combined with scoring by process, this would allow for assessors to judge sites by the priorities of the procurer/user and be scoped to particular user roles.
- **Question from Status section: How well will this proposal of conformance work in your type of organization and why?**

Response: It is really difficult to assess that. First, it is a little difficult to fully understand how the scoring and conformance will work in practice. Our current tools, test methods, and conformance reporting would have to change significantly to support the data needed to accurately calculate conformance using this model. Changes needed in tools:

- Not all test tools/test methods currently count all instances of elements tested or aggregate into a percentage score.

- Test rules would have to change for the requirements changes (e.g. text contrast) and the changes may require different splits between automated/manual testing.
- Test rules would have to change from reporting pass/fail on the previous WCAG 2.x success criterion (often by Level A, AA & AAA) to break down to the specific WCAG 3 outcomes.
- The flattening of levels within WCAG to a single level means a significant number of new test rules would need to be developed. There are a great number of Level AAA criteria (28 in WCAG 2.1, more than 1/3 of the total number of WCAG criteria) that have no test cases, automated testing, or conformance reporting in the tools. If Silver also adds a large number of brand-new requirements (e.g., Clear words), then the cost of implementing all of these criteria could increase the cost and time of accessibility implementation by a factor of 35% or more. Having this large number of additional criteria would reduce the number of products and websites that can meet bronze level and may dissuade authors from trying.
- Depending on the number of outcomes that require manual testing, the ability to count instances (manually) of applicable content, and appropriately test all of the outcomes could cost significant time.
- Aggregating test results for very large, complex applications that have multiple components, multiple user roles, and multiple technologies and use the scoring methodologies could be a challenge, at least initially. Remember, the end-game for a product is to report the accessibility to potential customers and give enough granularity to enable meaningful comparisons between competing products.
- ***Does the model for scoring and aggregated ratings work? Why or why not? If not, please propose an alternative solution.***

Response: The current proposal is a bit difficult to understand exactly how it works without a good example. It would help to show how this could work in real-world situations, for both large-scale and small-scale application of the scoring methodology. Suggest that there be an explainer document developed that contains the WCAG 3 document structure and other background (removing that material from the standard), with an example of how the scoring methodology could be used.

- The scoring and conformance are quite different from WCAG 2.x to WCAG 3. This means that persons who are trying to procure accessible products may find it difficult to compare products working to meet WCAG 2.x vs. WCAG 3. I know this is outside of the standard's purview but must be considered when thinking of the roadmap to adoption.

Usability and understandability of the standard itself

- Diagrams of the structure and organization of the document could use some updates for usability:

- Section **2 Structure of these guidelines** Figure 1 doesn't have all of the parts of the methods as shown in Figure 2. Missing "Introduction" and "Resources". Is this intentional?
- Section **2.3 Methods structure** is missing a couple of sub-sections based on Figure 2. There should be a subsection on the Introduction tab and what you can find there as well as a subsection on the Resources tab.
- Section **3 Additional Documentation and Scoring**: The changes in terminology between WCAG 2.x and WCAG 3 are sometimes confusing. Should there be a readily linkable section or inline definition that shows equivalent language, similarities in language, or a new definition/use of a WCAG 2.x term. Or, avoid unnecessarily creating new terms.
- Figure 3 is confusing for the following reasons:
 - Why do some of the boxes in the figure have heavy lines and others don't. Those aren't the documents, so what is the visual treatment trying to convey?
 - From this diagram it seems there are three kinds of scoring: test scoring, outcome scoring and functional category scoring. Is that correct, or are we misreading the diagram?
 - Combining document structure/relationships with the scoring relationship is confusing, especially since there's no indication in the diagram of what items represent documents vs. what items are needed for scoring. Separating scoring from documentation relationships into two diagrams may help - with a full text explanation of each.
 - The diagram is also confusing because How-to's have an "Outcomes" tab, which lists the outcomes, but has no connection in the diagram to outcomes.
 - If guidelines are normative, Functional Needs and functional categories seem out of place, as they are supporting information providing supportive context of the requirement. Perhaps they should be moved to the non-normative part of the documentation.
- **Question from status section: Are there usability improvements that would make WCAG 3 easier to use and find information? Is the structuring of the How-To, outcomes, and methods clear? Are there improvements to structure and style that you would like to see? Is there another way we can make content more usable?**

Response: There are a lot of resources, which does confuse things a bit. There are Functional Needs, Guidelines, Outcomes, Methods, and How-to's. The navigation between the documents is a little confusing and having this number of documents may make maintenance of WCAG 3 more time consuming. It is difficult to communicate to those unfamiliar with them which document(s) are best to fully understand a requirement and how to implement it. Wondering if it might be better to combine methods and how-to's into one resource and potentially guidelines and outcomes into one resource. Then there's less chance for getting lost or confused about which resource you're in or how to get to the related resource.

- **Section 4. Normative requirements** should be earlier in the document since this term is used in sections above with no pointers/links to this section where it describes what the

term means. Also suggest that the normative items be provided in a list format rather than prose. Lists are quicker and easier to parse.

Functional categories

We like the use of functional categories to drive the rationale behind the guidelines. We have a few comments on the functional categories:

- Functional categories defined in WCAG 3 do not have equivalent levels of granularity. For example, Vision and Visual is one functional category, which encompasses a number of subcategories which may have non-aligned and even have opposing needs. Even ignoring those, the needs of blind and low vision users are significantly different, both in modes of operation and in functional needs. Low Vision alone can encompass users with such disparate needs as visual acuity, light sensitivity, contrast sensitivity, shape perception, field of vision, color vision and motion tracking. It seems likely that the diversity of needs within the vision category is greater than that of other categories, as it relates to author considerations, yet it is encompassed by a single functional category in WCAG 3. As a second example, the Cognitive category is split into 7 sub-categories in WCAG 3, while 508 only has one combined category for [302.9 With Limited Language, Cognitive, and Learning Abilities, similar to the one category for EN 301 549's 4.2.10 Usage with limited cognition, language or learning](#). Suggest at the minimum to adopt a harmonized list of functional categories similar to the Revised 508 Standards Functional Performance Criteria and EN 301 549 Functional Performance Statements.
- References to the European EN 301 549 standard should be to the [latest V3.1.1](#), not the earlier V1.1.1, freely available from the ETSI.org website.
- It is unclear if photosensitivity and vestibular triggers are included under Vision and Visual or maybe under one of the intersection categories.
- Not sure if working on a separate functional needs document is necessary when the [ISO/IEC 29138-1:2018](#) document contains an extensive list already, granted it's behind a paywall. Addressing functional needs is important, and these can be mapped to the functional categories documented for the guideline. The ISO standard used to have a mapping of user needs to functional categories, but the latest update eliminated that mapping.
- Wonder if functional categories should be documented on the individual outcomes (or how-to's) vs. in the guideline.
- In the list of functional needs in the [Cognitive category](#): Attention sub-section, the "Use with limited ability to direct attention" seems synonymous with "Use with limited ability to focus attention". Most people aren't experts on cognitive accessibility, so they'll need some explanation of the differences between these two. Otherwise, consider simplifying the two items to a single functional need.

Terminology & Naming

- In general, there are a number of differences in terminology between WCAG 2.x and WCAG 3. This creates a good bit of confusion around the meaning and use of the terms.

It is difficult to remember the new term vs the old one. There are also a number of new terms which leads to more confusion.

The terms aren't all defined & compared to WCAG 2.x in one place. Some are in Appendix B, are hidden within paragraphs of information, or are stated in the definitions – it is inconsistent. This makes drawing mental parallels between the two standards harder. It might be helpful to gather those comparisons together into one place. Accompanying diagrams that show how WCAG 3 is structured (with the naming convention) compared to the WCAG 2.x structure (and names) could also be a helpful quick reference.

- WCAG has been re-defined as W3C Accessibility Guidelines, where WCAG used to mean Web Content Accessibility Guidelines. The change in name is not well explained in the introduction, especially since the acronym "WCAG" was retained. Add an explanation that the old WCAG acronym is still being used and why, and then why the "Content" part was removed from the long name. Otherwise, it may look like an oversight that these don't match up and will further confuse readers/adopters of the standard because this is a big departure from the structure, testing and conformance of previous WCAG versions.
- **5.3 Types of tests:** The ACT TF has two terms for test cases: Atomic tests and composite tests. It seems the use of atomic tests here is a different use of that term from what the ACT Rules Format defines, especially when composite tests aren't also included.
- **5.3 Types of tests:** Holistic test makes sense as a term when focusing testing on a particular disability type. Terms like "Automated tests", "Manual tests", "Assistive technology tests", "Usability tests" are more familiar in the space of WCAG 2.x so would prefer similar terms be used. Otherwise, it will be confusing moving from WCAG 2.x to WCAG 3. "Functional test" is also a familiar test term for testing particular scenarios or workflows. Suggest the TF works with some test professionals, both within and outside of the accessibility specialty, to help find consistently familiar terms.
- **Section 5. Testing** links to a "views" definition when there's a **5.1 Views** section that describes a view very differently. The definition is less understandable than the 5.1 Views section, but both need to be more clearly described. It seems that a "view" is the content a user has access to at a single point in time, whether it be a "web page" or scrollable content, or a dialog that is displayed due to some user action (opening a web page or document), using an interactive component, etc. Visual examples could also help relay the concept of "views".

Potential for blockers or delayed adoption of WCAG 3

The following can cause delays or be blockers to migration to WCAG 3:

- **Effort to consume and understand the new model**
- **Need for transition support materials:** Mapping of requirements from WCAG 2.2 to 3.0 is needed, with clear indication of what requirements and verification methods have changed or are new since WCAG2.2, ATAG, & UAAG. This will help product and website

developers, as well as test tool developers create a roadmap toward adopting the new standard.

- **Dependency on automated testing retooling before any larger-scale adoption can occur:** It would require a significant effort for existing tools to modify current automated testing to return results required for WCAG 3 including development of test rules for new and changed requirements, changes to the structure & granularity (where WCAG 2.x guidelines are split into multiple outcomes), reporting, and mapping to the new outcome structure.

When new methods of test or new metrics are introduced, there needs to be test rules with sample test cases (from the work of the ACT Task Force) to support automating them, as much as possible, to promote adoption. Otherwise, there will be a significant delay in adoption as teams will not be able to afford the cost and time of 100% manual testing until tools are developed.

- **WCAG 3 appears that it will have significantly more criteria:** The number of new requirements (and flattening of levels) would significantly increase the number of requirements, not to mention effort to implement them. (See also the bullet “Question from the Status section” in the [Scoring and Conformance](#) and [Requirements](#) sections of our comments for more thoughts on this topic.) Logically, that means WCAG 3 will be harder to do. If the effort to meet bronze is significantly greater than the effort to meet WCAG 2.x AA, it could slow or halt adoption.

Plain language summaries

We support the inclusion of plain language summaries and think these are helpful short descriptions of the topics covered in WCAG 3. We do have a few concerns, questions, and suggested improvements for plain language summaries.

- Including the Plain Language Summary at the beginning of each section, interspersed, and in-between other normative and informative content is a bit confusing. If the primary intent is to serve a specific audience, perhaps a consolidated view of only Plain Language content should be made available using a “plain language” filter. For example, the opposite of what is shown now with all the Plain Language sections collapsed and the complex language shown/expanded by default.
- **Introduction:** It is a little confusing that the plain language summary is titled “Introduction” yet the actual introduction section has no text other than, “This section is non-normative.” It instead summarizes the content in the sub sections below, none of which have “Introduction” in their title. Perhaps it should include an explanation that it summarizes sub-sections 1.1-1.4. The plain language summary for the [structure](#) sections is clearer since the bulleted list with one bullet for each sub-section that contains the distinctive word from the headings of the sub-section AND there is text with the diagram that gives a good introduction to the topic as well.

- In an effort to keep the plain language sections, a lot of sub-section details aren't present. Wonder if this may cause some confusion or missed information. In other words, the plain language summary cannot be relied upon on its own to understand the full picture. An example is Section 2. **Structure of these guidelines where** the topics of critical errors, outcome ratings and the structure of methods is not included.
- The Figure 1 core structure diagram could potentially be combined with the plain language summary for better comprehension. Some people understand pictures better, and some understand plain language better, and the combination may be better yet.
- Does each plain language summary need to have the last line announcing the "End of summary..."? Seems that the call-out box styling is sufficient to indicate that, especially if it is also defined with a landmark region role for AT programmatic identification of the block of information.

Requirements

The following comments have to do with the Guidelines and outcomes, and methods:

- We support the incorporation of ATAG and UAAG methods, provided the applicability to user agent vs authoring tool vs author is crisp and clear. This helps provide a bigger picture of how all of the accessibility pieces fit together, so a reader can get an understanding of how all of the parts need to be in place for the accessibility to work properly. However, in the current examples it isn't clear what aspects of the requirements are the responsibility of content/application owner vs. browser/user agent owners.
- Flattening the hierarchy of Levels (A, AA, AAA) into single set of requirements is contradictory to previous WCAG 2.x [conformance level](#) guidance that says: "It is not recommended that Level AAA conformance be required as a general policy for entire sites because it is not possible to satisfy all Level AAA Success Criteria for some content."

In addition, not having a hierarchy of Levels (A, AA, AAA) for guidelines or outcomes gives them all equal weight in the scoring system, and it shouldn't. This means that teams won't know what criteria are most impactful, which to implement first, etc. that are implied by having the levels. It may cause adoption of AAA success criteria to the detriment of other important criteria to try to score well in certain functional categories to obtain a Bronze level.

- In response to the questions regarding captions:
 - **Should the scoring of essential videos have more important/higher weight than captions on advertising and promotional videos?**
Response: There is some concern whether everyone viewing or testing a site would deem the same videos as essential. In addition, many advertising or promotional videos are 3rd party content that are not controlled by content authors of a web page or site. It seems unfair that the site or service be held accountable for the potential score of such content.
 - **Are burned-in captions equivalent to closed captions or should they be scored lower?**

Response: Burned-in captions, if they are of high quality, are nearly equivalent for the requirement of having captions, so should be scored either the same or slightly lower (as they are not available to users who are deaf-blind). Other requirements for being able to adjust the font, color, or placement should be separately judged and reported. Obviously, burned-in captions wouldn't support any user adjustments for captions. One way to handle this is to have Bronze level be provision of accurate captions, in whatever form. Silver and/or Gold support for this requirement could be providing the personalization.

- **Structured content:** Structured content is helpful to everyone, but not all content lends itself to having headings. For example, dialogs, forms, etc. In such cases, headings aren't needed so requiring all content to be organized into headings is not always a desired direction. When there are application controls, these also don't need to have a heading. As an example, if the controls are grouped into a UI Control (e.g., menus or toolbars) a heading isn't necessary. However, we do agree that where larger quantities of information are contained in the view, headings are one good way to visually and programmatically organize the content. Be careful that the standard doesn't require headings everywhere, when there are clearly situations that are perfectly accessible and understandable without their presence. Also, technically the example shows menu "headings" but, in reality, it's a structure of menu with sub-menus where the higher level menu is a higher level "category" of the sub-menus. There may also be "categories" or "sub-categories" used in lists of items, but these aren't technically called "headings".

Review Process for WCAG 3

- Appreciate ability to communicate that we needed more time for our review of WCAG 3, and that our input is still welcomed.
- **Response to question in 1.2 Background:** *As you evaluate this document, please consider whether there are ways the Working Group can better support your review, feedback, or inclusion within the process of creating this standard. We welcome feedback on this question as part of your comments.* It is unclear how to provide positive feedback of things we like. These don't seem to warrant opening an issue in GitHub, as there isn't anything to correct due to the comment. Assume sending an email would be most appropriate but making that explicit might be helpful to reviewers. It would also be helpful to always ask questions where specific input is needed on a topic rather than make a statement. Otherwise, one has to read carefully through editor's notes to find them. Be sure that if the same (or similar) questions are posed in the blog vs the standard's **Status of this document** vs editor's notes in particular sections that the verbiage is the same. Otherwise, it is harder to determine if these are new and different questions or a restatement of a question that's already been asked. Another way to handle is to gather all questions into one place and provide cross-references (or links) to the specific section where the review content is located.