Issue Severity

Availability

Who	Week 1 (12th Oct)	Week 2 (19th Oct)	Week 3 (26th Oct)	Week 4 (2nd Nov)	Week 5 (9th Nov)	Week 6 (16th Nov)	Week 7 (23rd Nov)	Week 8 (30th Nov)
Alastair	Regrets	Regrets						
Azlan				Regrets				
Francis								
Sarah							Regrets	
Shawn								

Goal

Create a proposal for rating outcomes for severity (user-impact).

- How do we define a scale of severity? (e.g. critical/high/med/low)
- How do we apply those ratings to the outcomes?

NB: The source of the sub-group was the 'real world accessibility' questionnaire, however, the sub-group thought that to tackle those problems we need to establish a rating based on user-impact first.

"Homework"

Week 1

1. Each member reviews current WCAG 3 FPWD content and documents 2 low, 2 medium, 2 high, 2 critical issues
    a. Note that some tests and outcomes include defined critical issues (see Critical issues, below). (Only one, but there are others but they were removed for the FPWD that we might be able to recapture.)
    b. I created a worksheet with the current WCAG 3 FPWD content for us to use and added ratings there. Feel free to add yours there, too! I've hidden mine so they don't influence others, and you can do the same once you're done. I used a share link — do we have a shared folder?
    c. There's this Phase 4 Silver Content space where we could add a folder. I (Francis) can't do that as my account permissions don't allow sharing outside

of my employer, so if someone else wanted to create a folder, that would be useful.

2. Bring those to week two's meeting.

Alastair's issues

Francis' issues

Sarah's issues

See note under Week 1, 1b

Shawn's issues

Todd's issues

Week 2

Using the worksheet, we'll work on the rows where we all agree: there are 2 rows of Critical and (text alternatives, captions) two High (structured content, visual contrast of text).

Week 4 (subgroup 2)

Considering context for issues at the test level.

Scenarios to consider:

- When a usually critical issue isn't. E.g. a lone input on a search page isn't critically required to have a label. You could consider it labelled by the page?
-

Example, labels for inputs:

- Where there is more than one input for similar info, you have to differentiate inputs.
- Where there is no context around the input to derive the meaning of the input.
-

Current methods of assessing user-impact

What are the key aspects of each method? What attributes of a guideline could we look at to make the assessment? What outside of a guideline affects the user-impact?

Usability testing

Generally a very good way of assessing user-impact as it is specific to the product and the context. However, it is difficult to scale and can vary by participant-sample and tasks conducted.

“Barrier score”

From a particular organisation, the barrier score is drawn from WCAG issues of an audit which is a sample of pages/components.
The score is evaluated by looking at the instances of issues and assessing what impact they would have on the relevant user-journey. (Divided by category then totalled into a percentage.)
Important that it is task based, the main thing being: how likely is an issue going to affect the journey?
Gives a percentage.

Issue severity matrix

Start with an audit in a spreadsheet to WCAG SCs, pass/fail.
Audit scoped to tasks (e.g., register, log in, find product/information, contact), so the task aspect is built-in already.
Define severity ratings, crit/high/med/low (derived from various sources, including Barrier Walkthrough Method)
Assign severity ratings for SC fails for different user groups (derived from functional performance criteria)

- Separated SCs where it impacted severity, e.g. has-captions vs accurate-captions.

Severity ratings not based on A/AA/AAA, but research based, data from audits, user research, evaluator expertise

Axe rules

https://dequeuniversity.com/rules/axe/4.4

Functional needs

Main Functional Needs - Main User Needs - Main Outcomes

Deceptive Pattern Outcomes
https://docs.google.com/document/d/1WKGPMiDL8CKcXWr0md09ZLwZEPdoxeNTNQsf8IuszqQ/edit#heading=h.qfsuwy1j0ruu

Categorization exercises from Guidelines Breakdown Exercise

Categorization exercise 2 Aug.xlsx

Joint SC Breakdown Exercise

Barrier Walkthrough Method

https://users.dimi.uniud.it/~giorgio.brajnik/projects/bw/bw.html

- Critical problem: the barrier is so big that very often users give up, and they do not reach their goals. This can happen after users have spent considerable time and effort to try to overcome the barrier, perhaps with many errors. There are no alternative ways (known to the users) that can be followed to achieve the goals. The barrier has a strong negative impact on effectiveness, and consequently also on productivity, satisfaction and safety.
- Significant problem: the barrier is detected and it heavily affects the task execution. To overcome the barrier the user has to back-up, follow a trial-and-error strategy, guess the proper action, repeat an action several times; the user may incur in errors. In many cases it is not possible to avoid the barrier, which reduces effectiveness and/or security; even if it can be avoided, this requires a substantial knowledge and/or memory (to recall that there is the barrier and on how to avoid it). The barrier affects effectiveness, productivity, satisfaction and also safety.
- Minor problem: the barrier is detected by the user, but there are simple ways to overcome it or to avoid it; it is easy to remember it, to learn how to avoid or get around it. This barrier affects marginally productivity or satisfaction, but not effectiveness nor safety.

Research Report on Web Accessibility Metrics

https://www.w3.org/TR/accessibility-metrics-report/

Critical issues

In the initial FPWD, where a ‘critical issue’ is a big problem.
https://www.w3.org/TR/wcag-3.0/#error-prevention
“Critical errors” are part of each guideline, see the error prevention example.
Presence of any critical error scores 0 for the outcome.
Critical errors are identified at the test level (e.g., https://www.w3.org/WAI/GL/WCAG3/2021/methods/instruction-for-completing-task/#tests-button)

- Test that instructions are present
    - Test Procedure
        1. Identify inputs that only accept data in a given format.
        2. Check that the page or view provides details about the required format.
    - Expected Results
        - #2 is true.
        - Critical Error if #2 is false.

Categories / axes

- Accessibility supported, e.g. only works in Chrome & Jaws.
- How is the testing/audit structured? Pages or tasks?
- How are the issues distributed over components / pages?
- Availability of work-arounds.

Issue Severity worksheet: Issue Severity worksheet

Potential approaches

Slide deck with test evaluation work from last year. Deck includes links to three sites that were used for testing:

1. Forked version, with accessibility defects, of an accessible escape room site
2. Deque's "Gefälscht" web testing course site
3. A short series of pages that resemble a login system.

Severity Assessment

- Combined is all the tests in one, per method
- Crit-severity worksheet:
    - Rating column hidden.
- The spreadsheet is a mechanism for establishing which tests should be critical.
-

Overall:

- Per test (within methods), identify a critical issue condition.
- Links with functional needs so you know which groups impacted. Show who is impacted & how.
- Requires you to review all instances of failures.
- Would lead to a failure if any critical issues are found, for that guideline.
- Means that you'd need some form of scoring (at least within the guideline), or perhaps priority (high/medium/low) per instance.

Issue prioritisation, post-testing

- Do all the testing first, then assess how big a barrier each instance is, in context.
- Per "scope" of the conformance statement, need to identify all known issues;
- Walk through an assessment of severity;
- Base the severity on various factors, including:
    - WCAG level (A/AA etc)
    - Impact on task (need to identify tasks)
    - Functional needs impacted;

    - (Can include organisation oriented things like which component is impacted, how big a task it is to fix etc.)
    - https://users.dimi.uniud.it/~giorgio.brajnik/projects/bw/bw.html
- Needs to be a process that happens once issues are identified;
- Could contribute to conformance, but not as part of the guideline/method layer.

Presentation: Issue Severity subgroup update

Drafting text for editor's draft

Current text

Process: A sequence of steps that need to be completed in order to accomplish an activity / task from end-to-end.

Critical error: An accessibility problem that will stop a user from being able to complete a process.

Critical errors include:

- Items that will stop a user from being able to complete the task if it exists anywhere on the view (examples: flashing, keyboard trap, audio with no pause);
- Errors that when located within a process means the process cannot be completed (example: submit button not in tab order);
- Errors that when aggregated within a view or across a process cause failure (example: a large amount of confusing, ambiguous language)."

Dependencies

Functional needs / User needs - need a useful subset to use for this purpose.
For functional needs SG: Can we have a simpler set of headings please? What makes a difference in terms of experience?

Scoping - how to incorporate process/context.
For scoping SG: How can we define a process/task that can be used to apply importance to a component?

Scoring - how to use the output of severity.
For scoring, critical tests should have more weight, particularly for a process.

Equity - how to maintain inclusivity when applying issue severity categories.

PR

- Demonstrated that it is possible to categorise tests by severity. (Worksheet)
- Best done at the test level. The higher the level the impact is assessed at, the less it aligns with the experience.
- Incorporate task/context for the best alignment would be best, but that depends on scoping and how to define the task/process.
- Will be a lot of work to categorise each test with an impact level & functional needs affected. Maybe best to focus on the critical issues. Would also be best to do as we go along when creating tests.
- Could contribute towards scoring, but there are many other questions to do with scoring.
- Could contribute towards prioritisation, which could replace A/AA/AAA (at the test level).

Proposed updates

For the process of creating this: Focus on critical issues. Don't try to rate high/medium/low.

Use the Critical severity matrix to categorise tests. (At least setup that process to define each test by functional need & impact.)

Could use Critical/High for a level of conformance (e.g. "Bronze"), and then silver if there are no crit/high/medium etc.

Could use the severity scales as input to a post-testing process where you prioritise issues based on your context and tasks.

Example

Functional images method/tests. TBC, Sarah will add at the bottom.

Next steps

Need to carry on grouping functional needs (that share critical user-needs), and have similar failed critical tests.

Create a process for adding the severity to tests as they are created, rather than as one big task later.

Those should then go in row-1 of the worksheet. Then for the critical tests, are these functional needs appropriate.

Once that is done, we'd need to use the matrix to:

- Support scoring, coordinate with the scoring sub-group.
- Provide data for content (e.g. how-to content, method content which both speak to how people are affected.)
- About performance, we need to interrogate whether a 'severe/critical' issue affects performance or prioritisation. E.g. can a regulator say that all the critical issues need to be done first?

Questions

What to do with non-critical issues?

Different people may have different ideas on what is critical.

How do we incorporate context / process?

Can matrix inform designation of functional categories, e.g., https://www.w3.org/WAI/GL/WCAG3/2021/outcomes/text-alternative-available

How to include context/task. Is that part of scoping, or issue severity. Both are important to the end result.

Edits to PR Sept 2022

PR Source: https://w3c.github.io/silver/guidelines/

Outcome: Text alternative available

Method: Functional images

https://www.w3.org/WAI/GL/WCAG3/2022/methods/functional-images/#tests-button

Procedure for Technology Agnostic

1. Examine each image in the content.
2. Check that each functional image that conveys meaning has its text alternative.
3. If the image contains text that is not purely decorative, the text alternative contains the same text.
4. If it is within a link together with text, check that it is implemented to be ignored by assistive technology or the text alternative describes the image and supplements the link text.
5. If it is a button, check that the text alternative indicates the button's function.

Expected Results

Checks #2 and #3, or #2 and #4, or #2 and #5 are true.

Critical issue if #2, #3, and/or #5 is false and functional image is needed to complete a task.

Outcome: Translates speech and non-speech audio

Method: Provides text equivalents of speech and non-speech audio

https://www.w3.org/WAI/GL/WCAG3/2022/methods/text-equiv/#tests-button

Procedure for each test:

1. Turn on the closed caption feature of the media player
2. View the synchronized media content with the closed caption feature turned on
3. Check that the captions can be turned on and off
    - Closed captions can be turned on and off
    - Open captions cannot be turned off

Rating:

- Each video with no captions is rated a 0
- Each video with open captions is rated a 1
- Each video with closed captions is rated a 2

Critical issue if video with no captions is needed to complete a task.

Subgroup 2 PR content

EDITOR'S NOTE

This section is exploratory.

Severity rating could contribute towards scoring and prioritization. This is a potential way to replace how A/AA/AAA levels represented severity by incorporating a mechanism to evaluate severity as a part of testing.

Outstanding questions that need to be addressed include:

1. What to do with non-critical issues?
    - Before non-critical issues can be worked on, it will be essential to understand scope and the conformance model.
2. How best to assign severity, particularly if testers have different ideas on what is critical?
    - The aim is to have the tests determine the severity, not the tester.
    - A next step would be to examine the tests to see if they can reliably predict the severity of impact on different functional needs.
    - Can critical tests be paired with a conformance model that uses prescriptive requirements to make severity rating more straightforward.
3. How do we incorporate context/process/task? Is that part of scoping, or issue severity? Both are important to the end result.
4. If included, how will situations where issue severity depends on context be handled?
5. How can the Framework for Accessible Specification of Technologies (FAST) be integrated and used to further this work, for example, to inform the designation of functional categories associated with outcomes?
6. How will issue severity fit into levels? For example:
    - "Bronze" could be an absence of any critical issues;
    - "Silver" could be an absence of any critical or high issues.
7. How to account for cumulative issues becoming critical?
8. Would another approach be more effective, for example assigning critical issues after testing is complete based on task or type of task rather than by test.

Tests will include critical issues. Each test will have a category of severity, so some tests will be flagged as causing a critical issue. Examples of critical issues in tests are at Text Alternative Available and Translates Speech And Non-Speech Audio.

https://w3c.github.io/silver/guidelines/#bronze

Bronze is the minimum conformance level and requires that there are no critical issues. Content that does not meet the requirements of the bronze level does not conform to WCAG 3.0. While there is a lot of overlap between WCAG 2.X and WCAG 3.0, WCAG 3 includes additional tests and different scoring mechanics. As a result, WCAG 3.0 is not backwards compatible with WCAG 2.X.

—----------

Outcome: Text alternative available

(See https://github.com/w3c/silver/pull/656/files)

Method: Functional images

https://www.w3.org/WAI/GL/WCAG3/2022/methods/functional-images/#tests-button

Procedure for Technology Agnostic

1. Examine each image in the content.
2. Check that each functional image that conveys meaning has its text alternative.
3. If the image contains text that is not purely decorative, the text alternative contains the same text.
4. If it is within a link together with text, check that it is implemented to be ignored by assistive technology or the text alternative describes the image and supplements the link text.
5. If it is a button, check that the text alternative indicates the button's function.

Expected Results

Checks #2 and #3, or #2 and #4, or #2 and #5 are true.

Critical issue if #2, #3, and/or #5 is false and functional image is needed to complete a task.

Outcome: Translates speech and non-speech audio

Method: Provides text equivalents of speech and non-speech audio

https://www.w3.org/WAI/GL/WCAG3/2022/methods/text-equiv/#tests-button

Procedure for each test:

4. Turn on the closed caption feature of the media player

5. View the synchronized media content with the closed caption feature turned on
6. Check that the captions can be turned on and off
    - Closed captions can be turned on and off
    - Open captions cannot be turned off

Rating:

- Each video with no captions is rated a 0
- Each video with open captions is rated a 1
- Each video with closed captions is rated a 2

Critical issue if video with no captions is needed to complete a task.