Thoughts on a Roadmap to Consensus

First draft - 11-September-2022

We have been working for a while now, actually years, on a few thorny issues including how to better cover the area of cognitive language and learning disabilities. A lot of good work has been done, but we also seem to be spinning our wheels on some key areas that appear to be preventing us from making the progress we need to make - and coming to consensus together.

We are now embarking on creating a structure and format for WCAG 3.x. And I see that some of these same things are rising up and, I'm afraid that if we don't address some of these big issues up front, we will waste a lot of time and lose some good people. Already I am hearing from people who are worried that we have spun our wheels a lot in the last years, and are concerned that we will just continue to do so. And some are giving up and checking out.

This document is prepared in a hope that it can help facilitate our identifying

* where we have common ground,
* where there are differences of opinion (and understanding why)
* and how we might come to consensus on answers to key questions
	+ or at least paths that we can take in order to come together with a consensus position on key issues and how to proceed.

This document is broken up into several different documents or "chapter" so that different aspects can be explored separately or that people can agree with some parts but disagree with others.

Let me start out by saying that all of the good ideas in this document are from other people in the working group. You can attribute the bad ones to me.

# Chapter 1: Where do we agree?

Below are a series of statements that it might be useful to include in a straw poll on to see if there is or isn't wide consensus on these points by the members of the working group.

I believe this will be helpful in identifying where we have common ground. It will also be helpful in understanding whether we are all in agreement and if not, where the differences lie.

In order to make progress,

* it is important to understand whether we all agree on our goal and assumptions and agree on the methods or path to get there
* or whether we have some fundamental disagreements on goals or whether we have disagreement on goals, principles, or basic assumptions.

## Do we all agree on these? If not – where not?

1. WCAG 2.X provides better coverage for individuals with sensory and physical disabilities than it does for individuals with cognitive language and learning disabilities.
2. This poorer coverage is not because there are fewer provisions or because less effort has been expended for cognitive language and learning disabilities – but because it is a much harder area to cover, it is more diverse, and because the other disabilities rely heavily on assistive technologies (and browser features) that transform content into accessible form – whereas such features and assisted technologies do not exist to the same extent for people with cognitive language and learning disabilities.
	* A study by John Slayton (before he died) and more recently a study by one of the co-chairs found that there are as many or more provisions for cognitive language and learning than any other disability. However, that does not mean it is as well covered. Just that there are as many provisions.
3. We have a lot of knowledge and many strategies for making content more accessible for people with cognitive language and learning disabilities, however much of it is not easily measured or able to be reduced to an objective true/false statement or criterion. This is a problem.
4. We need to figure out some mechanism for increasing the focus and efforts made by authors to make their content more accessible to people with cognitive language and learning disabilities, including those practices that do not necessarily have objective outcome-based criteria.
5. It is worth exploring non-outcome-oriented criteria in order to try to find ways to increase focus on measures for cognitive language and learning disabilities.
6. We want the WCAG 3.X series to be adopted as widely as the WCAG 2.X series, including adoption by regulatory agencies, industry and other authors.
7. If regulatory agencies will accept something beyond outcome-oriented objective criteria, then we want to identify these and explore this as a mechanism for getting more cognitive language and learning disabilities information and advice into the guidelines.
8. If regulatory agencies will not accept anything except objective outcome-oriented criteria, then we need to figure out some other mechanism to get industry and authors to look beyond just the objective outcome-oriented provisions.
9. A key area for exploration should be which type of provisions (outcome or other) the different regulatory agencies and the different standards groups (these may be different) will accept as normative or required provisions.
10. We should also explore how much non-normative, required, provisions they will accept or include in their documents.
11. Once we figure out what they will or will not accept as normative provisions, we want to structure WCAG 3.0 within those constraints so that 3.0 can be widely adopted by both standards organizations and regulatory agencies in some fashion.

# Chapter Two: Organization.

I would like to make the case for keeping it organized under Perceive, Operate, Understand and Robust – and using **tagging** to sort them in different ways for different needs.

Some suggestions have been made about organizing WCAG 3.x in some fashion other than under the perceive, operate, understand, and robustness (robustness = compatibility and future proofing).

As other organizations have been proposed, I have looked at them with interest. However, it is not clear what the exact purpose or theory of the reorganization was. Reorganization for change-sake can sometimes yield new insights. It can also lead to more confusion.

The original structure was designed to focus around **basic human functions**. That is, all individuals must be able to perceive information in some fashion. They must be able to operate any controls in the interface. And both must be done in a fashion that is understandable by people with diverse abilities. The robust category was created to handle those few things that cross these areas, or that were important, but did not relate to any specific one of them.

By organizing things in this fashion, it helps to create guidelines that highlight the fact that individuals with disabilities need to carry out the same functions as everyone else, they just need some considerations in the content in order to do so. The ***Guidelines*** then provided the general guidance. The ***success criteria*** then gave the specific minimum things that needed to be done that could be identified for each area that were testable (and could therefore be required). This organization provides a stable, expandable, and future proof approach to organizing in that, no matter what new technologies we introduce, they will still have to be implemented so that these three-plus-one basic functions (perceive, operate, and understand) can be accomplished by all the individuals with different types, degrees, and combinations of disability.

Other approaches have looked at organizing things by component type, or technology, et cetera. Since these will change over time, I see problems here. They appeared to me to be more like sorting categories rather than fundamental needs or requirements of the interface. Although I will continue to explore these different organizations as they are proposed, it would be helpful to understand how they are better than the current organization, and how they can help us to remain technology agnostic and base the requirements on the fundamental needs that users have for their interfaces.

Re-sorting the provisions into categories to match different authors or creator types might at first seem to be a better way to sort, since it puts all of the similar provisions for the similar author type together. However, this doesn't lead to any better understanding for those authors as to the dimensions that are important for accessibility. I believe it will lead to a little bit more of ‘filling out the form’ or ‘paint-by-number’ behavior. If there is a desire to make it easy for individuals to find all of the provisions that apply to them, I would suggest that we *tag* the provisions with the different author types, or technology types, or whatever else we would like to. It is then easy to sort or filter the provisions so that only the provisions that are of interest to a particular individual, disability group, author type, technology, component type et cetera are visible. This also allows you to organize or filter the same provisions in many different ways to meet the needs of many different audiences.

# Chapter 3, Scoring.

I have spent what is perhaps a person year over the years, looking at the topic of scoring. It always looked promising to me, but each time I've tried it, I’ve failed. And each time I thought I had a new method for it not to fail, it failed anyway after additional work and exploration.

## The problems.

A couple of problems kept coming up over and over again.

**Weighting -**- The first was that not all provisions seemed to be as important as other provisions. However, any attempt to provide some provision with some type of weight as compared to another provision failed.

When we looked at showstopper versus "just make it easier" provisions, we soon discovered that what was "just make it easier" for one individual may in fact be a showstopper for a different individual. Particularly in the cognitive language and learning areas, (but also for low vision and other disabilities.) something that seemed to just make things easier to use, was in fact required for an individual with a more severe or a different type of disability to be able to use the website/page/software. Thus, all provisions turned out to be "showstoppers" for someone or another.

Weighting things by the number of people who have a disability (eg, weighing something less because there aren't that many people who have that problem as compared to other problems) also fails quickly on inspection. If we start deciding that the most important things are the things where the most people have the problem, we would quickly find that people who were totally blind are a much smaller group than other disabilities. So blind provisions should have low priority? Similarly, some types of cognitive language or learning disabilities may have much lower incidents than others, but should these be ignored or given lower priority along with blindness or any other low incident disability that could be addressed through design considerations?

Thus, we come to the conclusion that we can't some provisions more weight or priority than others without deciding which smaller groups of individuals are less important enough to rate or weight their showstopper problems lower than provisions that address showstoppers for larger numbers of people.

We therefore end up having to weight all provisions the same or decide who is important or not.

### Critical Errors

Note that critical errors categorizations (which I also thought was a great idea for a while) falls victim to the same problem. What makes something a "critical error"?

* The fact that it is a showstopper?
* That it is a showstopper for a lot of people?
* What about if it's a showstopper for a smaller number of people?
* How many people does it have to be a showstopper for before it becomes a critical error? (And how do we count the number of people who would be affected by each provision in a fashion which was critical to them)?

### Accessibility is a chain. Do we allow paper links.

Another thing to keep in mind is that accessibility is like a chain. It doesn't do any good for some links in the chain to be strong if other links are made of paper. It doesn't matter if 99% of the links in the chain are strong if 1% of the links fail because they are paper.

For example, one can have a 100 different provisions that make a building accessible to somebody in a wheelchair – including lack of steps at the door, low nap carpet, accessible bathrooms, roll-in showers, enough room around the toilet, grab bars, toilet paper and faucets within reach, et cetera. However, if a hotel implements 99 of the 100 but has three steps at the only entrance or three steps up to the elevators, then the person in a wheelchair will never get to see the other 99 nice things that were done in the hotel or up in the hotel rooms.

The first temptation here is to claim that steps up to the front door or elevator would be "critical errors". But we quickly descend into the "which items are critical for who" discussion above. In the end we either decide that all of the elements are critical elements, or we end up having to decide who’s provisions we are going to include as critical. (Because they are not critical showstoppers to enough people?).

### Suggestion:

That we continue to not rate, or list, or use scoring to allow authors to, "pass" without meeting all of the requirements in the standard.

* If some products meet many but not most of the provisions, then they can list the fact that they meet X number of the provisions and they can list the provisions that they do not meet.
	+ This is an admission that their product does not meet some of the accessibility requirements decided by the working group.
	+ But it allows them to get credit for the provisions that they did meet.
	+ And it allows consumers to look at the provisions that they do not meet and decide whether or not that is a problem for them personally.

Importantly, avoiding the use of scores would prevent authors from deciding themselves which provisions are important or not important, or skipping hard ones and stopping when they have reached the number which gives them a passing score.

If there are some provisions that we decide simply are not important enough to always be required, we might make them into recommendations rather than requirements. That is, if it's important enough to make it a requirement, then it's important enough that it not be able to be skipped over while hunting for other easier to complete provisions sufficient enough to score a "pass". And it if isn’t important enough to require- we make it a recommendation.

# Chapter 4: Bronze, Silver, Gold

I think the best opportunity we have for actually getting companies to go beyond what has been required in WCAG 2.X – and to go beyond what will be in any regulation or legal requirement – is the idea of having Bronze, Silver, and Gold levels.

However, how we designate things as being in the Bronze, Silver, or Gold levels is key to both creating a document which can be used by regulatory agencies and one which will push beyond what we've been able to do in the WCAG 2.X series.

After thinking about this for a very long time, I would like to suggest the following approach for consideration of the working group. (These are based on lots of people suggestions and also issues raised by different people).

## Bronze.

Bronze level includes outcome-oriented testable requirements.

* All of WCAG 2.X provisions as well as any other outcome based testable requirements would go into this category. If desired, the Bronze level could have sub-levels in the same fashion as WCAG 2.x series.

Since these requirements can be independently tested and confirmed by a third party, an organization can conform at the Bronze level without having to make any declarations.

* This is important for many companies who can “follow the law” and “comply with the standard” but do not feel like they legally can "declare" that they have "conformed" to the standard or the law.
* They have stated that their lawyers tell them they can follow laws or regulations but, for any number of reasons (including ISO 9000 considerations, bugs, undue burden, equivalent facilitation, origin of content, et cetera) they should not ever declare or affirm that they have met the law or standard.

Whether a company must conform to a standard or regulation also falls into the area of policy and is therefore outside of the scope of our guidelines.

* The WCAG Guidelines give a measure of (minimum) accessibility but it is outside of the scope of the guidelines to determine when it is appropriate to require that someone must conform to the guidelines.
* For example, it is typical that accessibilityguidelines are not required to be followed
	+ in the cockpits of fighter jets; or
	+ if an equivalent but different mechanism is used to achieve the same result ("equivalent facilitation"); or
	+ if it would cause a fundamental alteration in the function of the product; or
	+ if it would present an undue burden; or
	+ if it was technically infeasible.

These are all topics we are grappling with in our group. But some of them are outside of the scope of our group.

By sticking with WCAG as being the ***ruler*** to ***measure*** accessibility rather than as the "pass or fail" to make it accessible enough (which is what scoring does), we can continue to be of value to regulatory agencies without crossing over into their domain. For example, should small businesses of only half a dozen employees have to follow all of these guidelines or not.

* It's not that we should make an exception for them – and if they don't follow the guidelines, we would declare their pages as accessible, or accessible enough. They would simply be, by law or regulation, required or not required to meet the regulations. And whether they are required to meet the guidelines or not - their pages would still be inaccessible if they did not meet the requirements
* Ditto for archival content – it might not be required (by policy) to meet the guidelines – or all of the guidelines – but that would not make it pass. If it did not pass it would not be accessible. It just would not fail the regulation since it was excepted.
* Ditto for 3rd party or content that flows in too fast to process, etc.

 One thing we could do – is provide levels in the Bronze.

* Regulations might be written to require different sets of criteria for different groups or situations.
* In this case we create different measures of accessibility and leave it to regulators to determine which set should be met by whom.
* These could be levels like WCAG 2.x or some other set created by tagging.
* Again, this is not for us to debate or decide who should be required to do what, this is a regulatory issue rather than a standard of accessibility issue.
* Note: In deciding levels or groups we run the same risk as with “critical errors” or “weighting”. How do we decide which provisions are most important for these situations. This is not so say we shouldn’t look at this – just that we need to be careful of our biases.

## Silver.

This is where we get to push beyond what WCAG 2.x Has been able to achieve.

However, if we want this level to also be of interest to regulatory or other people setting requirements, we still have to make this level objective.

* That is, we can't require a company do something if there is no way for the company to reliably determine if they in fact have conformed to the requirement *before* some external judge is brought in.

The suggestion here is that Silver requirements be based on affirmations.

* For example, a company could affirm that it has carried out some process in order to make their content more accessible.
* They might
	+ affirm that they have reviewed their pages for complexity of language and tried to increase its readability and decrease the reading age.
	+ Affirm that they have included people with disabilities in the early product development stages for their websites.
	+ Affirm that they have included people with the following disabilities in their user testing.
	+ Affirming that they have …. [done anything else in the overall guidance documents that the W3C has produced for how to make web content more accessible – but where the outcome was not testable - (or it could have already been included a the Bronze level)]

Since the requirement for the elements in Silver level is to affirm something – it would be testable and could therefore be included in requirements.

* It is testable by simply determining whether or not they have made such an affirmation.
* While it's not possible to actually go into the company and to test whether or not they did it, this is not required since the only requirement was that they affirm that they did it. The requirement is for an affirmation.

Affirmations can be very powerful for most companies because of something called the ISO 9000 and 9,001. These are standards that are very important for companies. They basically state that if you affirm that you are doing something, you must in fact be doing it.

* [From Wikipedia entry for ISO 9000](https://en.wikipedia.org/wiki/ISO_9000)
	+ “The **ISO 9000** family of [quality management systems](https://en.wikipedia.org/wiki/Quality_management_systems) (QMS) is a set of standards that helps [organizations](https://en.wikipedia.org/wiki/Organizations) ensure they meet customer and other [stakeholder](https://en.wikipedia.org/wiki/Stakeholder_%28corporate%29) needs within statutory and regulatory requirements related to a product or service.[[1]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-1) [ISO](https://en.wikipedia.org/wiki/ISO) 9000 deals with the fundamentals of QMS,[[2]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-Tsim,_Yeung,_Leung_2002-2) including the seven quality management principles that underlie the family of standards.[[2]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-Tsim,_Yeung,_Leung_2002-2)[[3]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-iso.org-3)[[4]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-4) ISO 9001 deals with the requirements that organizations wishing to meet the standard must fulfill.[[5]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-5)
	+ “Third-party certification bodies provide independent confirmation that organizations meet the requirements of ISO 9001. Over one million organizations worldwide[[6]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-6) are independently certified, making ISO 9001 one of the most widely used management tools in the world today. However, the ISO certification process has been criticized[[7]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-clifford-7)[[8]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-wilson-8)[[*self-published source?*](https://en.wikipedia.org/wiki/Wikipedia%3AVerifiability#Self-published_sources)]as being wasteful and not being useful for all organizations.[[9]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-seddon-9)[[10]](https://en.wikipedia.org/wiki/ISO_9000?wprov=srpw1_0#cite_note-seddon2-10)”
* Thus an affirmation can be a very powerful incentive to doing something.

One might actually include affirmation that one meets the Bronze level as a requirement for a Silver. This however runs into the problem that we raised above. I would therefore suggest that we not require that they affirm that they have met Bronze as part of achieving Silver. One should also not require that somebody affirm that they have met Silver in order to meet the various affirmations in Silver.

Again, it may behoove us to have different levels of Silver, where we put some affirmations as being basic affirmations and put them at Silver 1. And then have Silver 2 be more difficult and Silver 3 be the most difficult types of affirmations. Or require a number of affirmations (but think about the comments above about scoring).

The key to this level being helpful in moving beyond the constraints of WCAG 2.x is that this Silver allows us to include a lot of things where you do not have a testable outcome. However, if the places affirm that they have carried out some process or procedure, it is easy to test whether or not they have made the affirmation.

 People wanting to say that they've gone beyond just the basic Bronze level could make Silver affirmations – and that would allow us to put more things on the table that would be in the focus and attention of developers than just those that have/are testable outcomes.

To get into this Silver level, they more likely to look at ***all*** of the Silver level. So Silver will be most effective ir we make the Silver level-1 easy enough that people will actually go for it.

* Once they are looking at silver-1 we have gotten them to look beyond just the testable outcome requirements and looking at (all of) the Silver requirements.
* They will then see the Silver2 and Silver3 levels and are more likely to be able to go for them. If we make the first level of Silver too hard, then it is likely that they will look and then just bounce off of Silver and stick with Bronze.
* Silver might be most effective if it does not have levels but rather you earn different levels by doing different numbers of Silver affirmations. In this way – even if they are only looking to do Silver1 they might look at *all* of the Silver items to see which they can most easily do (rather than only considering the level 1 items if we group them by levels.)

## Gold.

This is the highest level and something that we don't want to award easily. For someone to get to this level they have to really be doing something exceptional.

It could be that this is an even higher or tougher set of affirmations than Silver

It could also be something different.

* This would be in keeping with each of the three levels (Bronze, Silver, and Gold) being a different kind of achievement.
	+ Bronze - is that you have met the outcome achievements.
	+ Silver - is that you've gone beyond just those things that are testable outcome and have affirmed that you have done other processes to try to do even more than those things that are easily testable, but that are equally important.
	+ Gold - would mean that you have gone even further yet.

One idea is that Gold isn't something that you can earn, it can only be something that is awarded to you.

* That is, there is some independent body that judges and gives "Gold" awards.
* The danger here is that that could very quickly become political. Companies with lots of money, or lots of influence, or who have contributed in other ways, or who can apply lots of pressure, could influence the outcome.
* Another problem is that if a lot of places were doing great work, it could easily overwhelm whatever this ‘judging' committee is.
	+ And who is going to pay for the people to be doing this judging (it takes time to evaluate things for accessibility).
	+ Perhaps an independent lab can be set up like an underwriter laboratory that would do this?
* Although this sounds attractive in some ways I can see nothing but a continuous number of issues, or questions, or problems.

I toss these ideas out here just as a starting point for discussion or for people to think about how they can't work and come up with better ideas.

I had thought of things like user testing or things like that as being at the Gold level. But those could also be an affirmation done at the Silver level, perhaps at the higher level. Or as mentioned above – as a Gold level affirmation.

### Note on user testing

Unless you test with a lot of people representing the full spectrum of different disabilities, AND the different degrees of disability, user testing can be not very useful or even as useful as other requirements in establishing that something is accessible.

**Any type of user testing however is great.**

* It allows for users to actually try to use the whole site as an entity and to carry out functional use of the site.
* Whereas the Bronze requirements could pick out individual things that are likely to be barriers. It is possible for sites to meet all of those minimum requirements and still not be very usable.
* ***User testing is really important and should definitely be at the Silver level or above***.
* The biggest concern, however, is that if testing is done with a small number of individuals and sites are optimized for them rather than the full spectrum of people with different types, degrees, and combinations of disabilities, then the accessibility of the site might be limited.
* And usability testing by itself, of course, doesn't really assure very much.
	+ I have seen very bright blind people use sites that are very inaccessible to most blind people, but these bright blind users are just so good they are able to work their way around what would otherwise be show stoppers for ordinary users.
	+ Also in the area of cognitive language and learning disabilities testing with someone with one type or one degree of cognitive language or learning disability, of course is not representative at all of individuals with the other types of cognitive language learning or disabilities, and also not for people with more serious cognitive language or learning disabilities, or even less serious.

# The bottom line.

The bottom line is that I think we need to stop wrestling with all different approaches trying to find the final one -- and try to come together on some one approach to start with. That does not mean that at the end of the process, we won't decide to reorganize and change how we approach things. But right now we don't seem to be agreeing on any one approach and we don't seem to be able to "live with" anything except each of our different view of what might be ideal.

We need to come together on something and go ahead and try to build it. Then after we have gotten a bunch of the pieces in place, we may look at them and say, "Hey, we've got a better way."

My suggestion to the group is that we pick some method – this one or some other – and then say, "Okay, we're going to go with this unless we can think of a better one." And then we can try to think of better ones as we work on putting the different pieces together for this one.

The danger is that we don't make progress because we spend all of our time trying to think of the ideal. I'm afraid there isn't an ideal. Or at least nothing we all would agree on as ideal.

I hope that these thoughts are useful. I am looking forward to people looking at them, evaluating them, either poking holes in them, or even better, figuring out how to improve them or tune them or correct them, so that we have a basis for moving forward.

 It's going to be a lot of work to get to where we need to go to. We need to have some structure, orientation, and direction in order to do that. And a lot of patience and listening.

Humbly (and with some trepidation) submitted

Gregg Vanderheiden.