"Monitoring Systems for Checking Websites on Accessibility"

A STUDY BY THE COMPETENCE CENTER FOR DIGITAL ACCESSIBILITY.

AUTHORS: ANDREAS BURKARD

PROF. DR. GOTTFRIED ZIMMERMANN

PROF. DR. BETTINA SCHWARZER





Accessible Screenshots and Tables

All **screenshots** included in this presentation are each described in a YouTube video. The corresponding YouTube video is linked to the screenshot so that you can start it by clicking on the screenshot. In addition, the transcript of the video is stored as alternative text for the screenshot.

All **tables** are alternatively available as Excel files. There is a link "Download Excel file" at the very beginning of each table (first focusable element).





Research Question

What would be the appropriate solution for monitoring websites in terms of accessibility for the websites of the "Hochschule der Medien Stuttgart" (English: "Stuttgart Media University")?





Advantages of an Accessibility Monitoring System







Siteimprove





📕 Hide menu		O Site	improve		? Help Center and Academy $ \smallsetminus $	🔅 Settings 🗸	💄 Andreas Burkhard 🗸
Search in menu Q	Hochschule der Medien Stuttgart ~ http://www.hdm-stuttgart.de	No group selected					O Search for a page
Dashboard	Accessibility						<u>↓</u> Export
< MAIN MENU							
Accessibility	Accessibility Overview						export
Accessibility Overview	Score details		Overall	score @	Accessib	ility Score progress	0
Summary ~	Progress in resolving Errors Ø	89,0 /100	Acces	sibility	100		
Issues	Progress in resolving Warnings 🖗	77,3 /100	61	9 (100	50 <u>r + </u>		
Guidelines	Percentage of pages with low Error rate @	0,0 /100	+(0,0	0	24.05.2019	07.12.2019
Pages					— Score — - Industry bend	chmark (Education)	P Annotations
PDFs			👬 Score b	preakdown			
Validation Y							
All Decisions V		● Level A Ø ● Level AA Ø	Levels of Conform	O Errors O	⊚ Warnings @ ⊙ Reviews @		
Chrome extension							
	Issues (all roles)	🦨 Editor	🧩 Web	omaster	Leveloper	Page	e error rate
	Fix These Issues To Impr	ove Your Score	⊘ Export	Fixed Issue	S		© Export
	Issue		Points you can gain	Issue			Points already gained
	Link identified only by color		0,82 points	WAI-ARIA i	mage is missing alternative text		0,83 points

Hide menu		O Siteimprove	? Help Center and Academy 🗸 🏚	Settings 🗸 🙎	Andreas Burkhard 🗸
Search in menu Q	Hochschule der Medien Stuttgart ~ http://www.hdm-stuttgart.de	ected V			О Search for a page
Dashboard	Issues / Specific issue under 1.1.1 Non-text Content				
< MAIN MENU	image with no alt attribute	5			
Accessibility	★ ① Image with no alt attribute			? help	options Export
Accessibility Overview					
Summary ~	Description The image does not have an 'alt' attribute (alt="").	How to t	fix it: ortant all images have the attribute for alternative text regardle	ss of whether an a	alternative text is

Issues

Guidelines

Pages

PDFs Validation All Decisions

Chrome extension

Where to find it in WCAG 2 🚱





added.

give no value to the end user.

attribute is added to all images.

A screen reader knows how to handle both an empty alt attribute and one with a text. If there is no attribute some screen readers will compensate and read the path to the image instead, which will often

If you are using a CMS (Content Management System), the default setting should be that an empty alt

Hide menu				O Siteimprove	• Help Center and Academy	🗸 🏟 Setti	ings 🗸 💄 And	dreas Burkhard $\!$
Search in menu	٩	Conformance level	No.	Guidelines and Success Criteria		Errors (!)	Warnings 🛆	Reviews 💿
			1	Perceivable				
Dashboard			1.1	Text Alternatives				
		А	1.1.1	Non-text Content 📀	~	5	~	3
A			1.2	Time-based Media				
Accessibility		Α	1.2.1	Audio-only and Video-only (Prerecorded) 📀	~	~	~	1
Accessibility Overview		Α	1.2.2	Captions (Prerecorded)	~	~	~	1
-		А	1.2.3	Audio Description or Media Alternative (Prerecorded)	~	~	\checkmark	1
Summary	~	AA	1.2.4	Captions (Live) 👩	\sim	~	~	1
Issues		AA	1.2.5	Audio Description (Prerecorded) 📀	~	~	~	1
0.11.1		AAA	1.2.6	Sign Language (Prerecorded) 💿	~	~	~	1
Guidelines		AAA	1.2.7	Extended Audio Description (Prerecorded) 👩	~	~	~	1
Pages		AAA	1.2.8	Media Alternative (Prerecorded)	~	~	~	1
DDEo		AAA	1.2.9	Audio-only (Live) 👩	~	~	~	1
			1.3	Adaptable				
Validation	~	Α	1.3.1	Info and Relationships 🔞	~	16	4	4
	~	A	1.3.2	Meaningful Sequence 💿		-	-	-
		Α	1.3.3	Sensory Characteristics @		-	-	-
Chrome extension		AA	1.3.4	Orientation 📀		-	-	-
		AA	1.3.5	Identify Input Purpose @		-	-	-
		AAA	1.3.6	Identify Purpose @		-	-	-
			1.4	Distinguishable				
		А	1.4.1	Use of Color 🔞	\sim	1	~	~
		А	1.4.2	Audio Control 📀	~	~	~	1
		AA	1.4.3	Contrast (Minimum) 📀	~	1	~	~



my2.siteimprove.com/QualityAssurance/Inspector/957804/23462961248/PageDetails/Report...

Siteimprove - Summary







axe Monitor

DEQUE





WorldSpace Comply

Welcome, Andreas Burkard <u>Laccount</u> <u>Logout</u>

2

х

 \mathbf{v}

🕀 Home 🔄 Dashboards - 📃 Pages 🛕 Issues 📀 Scans 🤽 Settings -

Project Dashboard for: Hochschule der Medien, hdm-stuttgart.de -

Date Range 05-20-2019 - 06-19-2020 Go!

Update now Show potential issues Scans Export to excel Share report

Project snapshot - effective 18.06.20 22:13

			Pages				
Snapshot for-	Score	Issues Per Page	Total	Critical	Serious	Moderate	Good
hdm-stuttgart.de	26%	10	<u>1070</u>	<u>416</u>	<u>634</u>	<u>0</u>	<u>20</u>

View Issue Assignments

Automated issues

- Grouping	¢ Standard	Priority	Description	+ Issues	Pages	% of Pages
Color				<u>2684</u>	<u>759</u>	
Color	Success Criterion 1.4.3 Contrast (Minimum)	Serious	Ensures the contrast between foreground and background colors meets WCAG 2 AA contrast ratio thresholds	<u>2684</u>	<u>759</u>	71%
Forms				<u>96</u>	<u>24</u>	
Forms	Success Criterion 3.3.2 Labels or Instructions Success Criterion 1.3.1 Info and Relationships	Critical	ures every form element has a label ¹		<u>24</u>	2%
Keyboard				<u>129</u>	<u>129</u>	
Keyboard	Success Criterion 2.4.1 Bypass Blocks	Serious	Ensures each page has at least one mechanism for a user to bypass navigation and jump straight to the content 1	<u>129</u>	<u>129</u>	12%
Language				<u>172</u>	<u>171</u>	
Language	Success Criterion 3.1.1 Language of Page	Serious	Ensures every HTML document has a lang attribute 0	<u>170</u>	<u>170</u>	16%
Language	Success Criterion 3.1.2 Language of Parts	Serious	Ensures lang attributes have valid values 0	2	1	0%
Name Role Value				<u>2551</u>	<u>788</u>	
Name Role Value	Success Criterion 4.1.2 Name, Role, Value Success Criterion 1.3.1 Info and Relationships	Serious	Ensures aria-hidden elements do not contain focusable elements 🛈	2	2	0%
				40	40	4.07

Automated issues

Assigned To	Label	Description	Severity	Element source code	Тад Туре
		Ensures the contrast between foreground and background colors meets WCAG 2 AA contrast ratio thresholds	Violation	Mehr erfahren	A
gzimmermann@acm.org		Ensures elements have alternate text or a role of none or presentation ①	Violation	<pre><img https:="" scontent.cdninstagram.com="" src="https://scontent.cdninstagram.com/v/t51.2885-
15/e15/s150x150/828287</pre></td><td>IMG</td></tr><tr><td></td><td></td><td>Ensures elements have alternate text or a role of none or presentation ①</td><td>Violation</td><td><pre><img_src=" t51.2885-<br="" v=""/>15/e15/s150x150/808056</pre>	IMG
		Ensures links have discernible text	Violation	<pre><img src="/news/news201912181057</pre></td><td>A</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td><pre><img src="/news/news202001091107</pre></td><td>A</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td><pre><img src="/news/news201912181057</pre></td><td>A</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td><pre><img src="/news/news202001091202</pre></td><td>A</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td></pre>	А
		Ensures links have discernible text	Violation		А
		Ensures links have discernible text	Violation		А
		Ensures links have discernible text	Violation		А
		Ensures links have discernible text	Violation	<pre><a cla<="" class="see_more" href="/kontakt" https:="" id="manage_icon" index_html="" manage"="" pre="" style="background-color:
rgb(100,100,100,</pre></td><td>A</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td><pre>Ka href=" www.hdm-stuttgart.de=""></pre>	A
		Ensures links have discernible text	Violation	<pre> <img src="</pre></td><td>Α</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td></pre>	А
		Ensures the contrast between foreground and background colors meets WCAG 2 AA contrast ratio thresholds	Violation	<pre><a onclick="var d = new Date(); d = new Date(d.getTime() +1000*60*60*24*730);</pre></td><td>A</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td><pre><img "="" b9rmcvao2oi="" https:="" p="" src="/news/news202003300931</pre></td><td>А</td></tr><tr><td></td><td></td><td>Ensures links have discernible text</td><td>Violation</td><td></pre>	А
		Ensures elements have alternate text or a role of none or presentation ①	Violation	<pre><img_src="https: scontent.cdninstagram.com="" t51.2885-<br="" v="">15/e15/s150x150/929511</img_src="https:></pre>	IMG
		Ensures elements have alternate text or a role of none or presentation ①	Violation	<pre><img https:="" scontent.cdninstagram.com="" src="https://scontent.cdninstagram.com/v/t51.2885-
15/e35/s150x150/907021</pre></td><td>IMG</td></tr><tr><td></td><td></td><td>Ensures elements have alternate text or a role of none or presentation ①</td><td>Violation</td><td><pre><img src=" t51.2885-<br="" v=""/>15/e15/s150x150/904253</pre>	IMG
		Ensures elements have alternate text or a role of	Violation	<pre><img src="https://scontent.cdninstagram.com/v/t51.2885-</pre></pre>	IMG

٠

Ŧ

Coor Wor	rldSpace Com	ply			×	Welco	me, Andreas Burkard	Logout ?
Home	Dashboards -	Pages	▲ Issues	🕲 Scans	🤹 Settings 🗸			
Issue De	tails for: Hochscl	hule der M	edien, hd	m-stuttgai	rt.de			
Date created:	Freitag, 24. Januar 2020 14:2	5 Uhr UTC						<u>Delete issue</u>
Practice:	Accessibility							
Issue description:	Ensures elements have presentation	e alternate text or a i	role of none or				0	
Common Issue	No							
Severity	Violation ~							
Issue comments:							11	
Suggested remediation:							4	
Page:	https://www.hdm-stuttgart.de/					<u>View Page [</u>	Details	
Element:								
Selector:	1.#insta > li:nth-child(2) > a[target="_	blank"] > img					
Element source code:	<pre><img src="https://scont
_nc_ht=scontent.cdninst</pre></th><th>ent.cdninstagram
agram.com&_n</th><th>.com/v/t51.2885
c_ohc=qumATWKtl</th><th>-15/e15/s150x150
o8AX9XOtts&o</th><th>/82828709_1012903080519
h=f77fc63730d4682ff2a50</th><th>906_5243237493452959773_n.
c4f4b474b9db&oe=5EBE56</th><th>jpg?
41"/></pre>							

Suggested emediation:	
Page:	// https://www.hdm-stuttgart.de/
	View Page Details
Element:	
Selector:	1.#insta > li:nth-child(2) > a[target="_blank"] > img
Element source code:	<pre></pre>
Status:	Open 🗸
signed to:	ab246@hdm-stuttgart.de 🗸
Labels:	(separate multiple labels with a comma ",")
Additional	Summary:
formation:	Fix any of the following:
	Element does not have an alt attribute
	aria-label attribute does not exist or is empty
	aria-labelledby attribute does not exist, references elements that do not exist or references elements that are empty
	Element has no title attribute or the title attribute is empty
	Element's default semantics were not overridden with role="presentation"
	Element's default semantics were not overridden with role="none"

.

DEIEN A STUDIEREN. WISSEN. MACHEN. **f y o o in** FÜR UNTERNEHMEN \bigcirc HOCHSCHULE STUDIENINTERESSIERTE **STUDIERENDE** HOCHSCHULE **DER MEDIEN IRGENDWAS MIT MEDIEN** INSTAGRAM ABSCHLUSS Bitte auswählen V SUCHEN VERANSTALTUNGEN NETZWERK ERFOLGE Diese Website verwendet Cookies. Durch die Nutzung dieser Website erklären Sie sich damit einverstanden, dass Cookies gesetzt werden. Mehr erfahren 🛛 3 🗛 1 🛛 🗱 : X Elements Console Sources Network Performance Memory Application Security Lighthouse Adblock Plus WorldSpace ARC Toolkit axe Landmarks WorldSpace Attest Images must have alternate text i ≪ < 1 of 7 > ≫ Analyze -Issue description To solve this violation, you need to: All issues found 60 -D Run again Impact: critical Ensures elements have alternate text or a role of none or presentation CLearn more Fix at least one (1) of these issues: · Element does not have an alt attribute 14 Elements must have sufficient color contrast Element location · aria-label attribute does not exist or is empty · aria-labelledby attribute does not exist, references elements that do not exist or references elements that are Images must have alternate text 7 #insta > li:nth-child(1) > a[target="_blank"] > img empty Links must have discernible text 37 Element has no title attribute or the title attribute is empty Flement source

Issue tags: category: text-alternatives wcag2a wcag111 section508 section508.22.a

widoos alamanta muat hava a

axe Monitor - Summary







ARC Monitoring

THE PACIELLO GROUP











.

Ŧ

Show 100	✓ entries						Search:	
Page Count ↓≣	Engine 🕸	Assertion	.↓↑	Violation Count	l↑ Cat	ategory ↓†	WCAG 2.1 Checkpoint $\downarrow\uparrow$	Checkpoint Level
164	ARC	lineBreakUsed ` `element is used to create visual line breaks. This may indicate that content is not correctly marked up / semantically structured. Ensure that appropriate structural markup is used.		1624	AL	LERTS	1.3.1 Info and Relationships	А
164	ARC	titleOnNonActiveElement A non-interactive/content element has a `title=""` attribute. This is not consistently exposed by browsers / announced by assistive technologies.		535	AL	LERTS	3.3.2 Labels or Instructions	А
164	ARC	onlyOneListItem Found an ` ` ordered list or `` unordered list that contains only a single list items. Ensure that this structure is logical (for instance, if there are situations where this list may contain more than one item), as otherwise this structure may be unnecessary and potentially confusing for users.		427	AL	LERTS	1.3.1 Info and Relationships	А

Ŧ

ARC

ARC Home

Main Navigation

Domains

Document Library

m TPG Tutor

KnowledgeBase

🛋 Rules Engine

🖹 Reports

🐣 Profile

Account Administration

📽 General Settings

Sccess Rights

Crawler Settings

m Tutor Settings

🗋 Page Assertions

=

image lacking alt attribute

The element lacks an alt="..." attribute. As a result, the meaning/purpose of the image won't be conveyed to users who can't see the image, including users of assistive technologies.

WCAG 2.1 - Errors

Guideline 1.1 Text Alternatives: Provide text alternatives for any non-text content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language.

1.1.1 Non-text Content (A)

Engine: ARC Rules v3.2.1 Key: noAlt

KnowledgeBase Resources

- Desktop and Responsive Web: Deciding what type of text alternative to provide
- Desktop and Responsive Web: Provide a text alternative using the alt attribute

Contact us about premium content

Commentary

No commentary available for this assertion.

Add new commentary

Related HTML

Found on https://www.hdm-stuttgart.de/hochschule/aktuelles/presse

 $l. \quad < \text{img src="/news/news20200302165157/thumbstart"} > \\$

2.

3.

4.

5.

L¥.			8 3	A 1	्य	7	:	X	_
Re	DO IIIayes	ø							
-	Rich media	1							
-	Headings	33		3	1				
	Landmarks	7		1					
	Lists	14		4	88		46		
	Paragraphs	39		1	97				
Ire	Pseudo content	1			21				
ructu	Tables	0							
st	Forms	4	1	2	10	5	5		
	Frames	0							
	Titles	28		9	35		9		
	Text formatting	5		5					
	Language	2							
	Links	125	14	20	301	163			
	Internal links	0			20	20	20		
oard	Buttons	7							
Keyb	Access keys	0							
	Tabindex	6	6		106	106			
	✓Tab order	134	4						
	ARIA UI	0							
KIA	ARIA live	0							
AF	ARIA usage	2							
	ARIA hidden	1		1					
Color	Color contrast	145	2	8					
DS	IDs	26			24				
Show and track focus Check page reflow (WCAG 2.1) ⑦ Check text spacing (WCAG 2.1) ⑦ Send DOM to validator									
Se	nd URL to validat	tor							



BEWERBUNG FÜR MASTERSTU25ENGÄNGE AB SOFORT MÖGLICH

HOCHSCHOLE

Ab sofort können sich Studieninteress ierto für die Masterstudiengänge an der Hochschule der Medien (HdM) bewerben. Insgesamt zehn Masterstudiengänge stehen zur Ausvahl. Neu im Angebot: Die Vertiefungsrichtung Sportkommunikation im Master Crossmedia P ... mehr

STUC#5M

Q

......

Die Hochschule der Medie intornhiert, wie sich die Covid-19-

Pandemie auf den Hochschulbetrieb auswirkt:

Informationen zum Studium

Informationen für Erstsemester

Informationen für Beschäftigt

Allgemeine Informationen zu Coronavirus

Informationen für Studieninteressierte





...................

.....

ARC Monitoring - Summary







Pope Tech





$\equiv \mathbf{P}$ pope**tech**

Stuttgart Media Univ... × 👻

>



Accessibility	

Dashboard

Update

-

Websites

📐 Dashboard

🚠 Groups	
Lsers	>

Settings

Result Typ	bes			Most Com	imon Issues				Chart	Tab
×	Errors 26.65k	0	Contrast 2.39k	Errors	Alerts					
	Alerts	-	Structure			26.645				
	100.26k	A	73.33k			Total				
						13.487	12.162	613	309	
	Features					Linked image missing	Empty link	Missing form label	Missing alterna	ative .
_	14.3K		TT.JK			30%	43%	۷%	1 %	
Scan Deta	ils				2 ≣	Errors By Group			Chart Table	¢
Name	Frrors -	Contras	t 🌢 🛛 Alerts 🚖	Pages 🚖	Details		- it i			

Name	Errors 🗸	Contrast 🗢	Alerts 🗢	Pages ≑	Details
https://www.hdm-stu	9,862	800	39,652	227	>>
HdM (depth 1)	8,083	794	30,215	159	>>
HdM (comparison 19	8,082	794	30,218	159	>>
https://digitalisierung	409	1	147	93	>
Digitalisierung (Dept	77	1	9	19	>>



					? 🛛 Andreas Burkard 🔻
Stuttgart Media Univ × 💌					Dashboard Deta
📥 Dashboard	Recult Det				
Accessibility >	Nesult Det				
Websites	Filters				
🚠 Groups	Website				
Lusers >		LIIOIS			
Settings					
	Results				
	Errors				
	Info	Name	Count	▪ Websites ≑	Details
	P	Linked image missing alternative text	13,48	7 7	>
	(max)	Empty link	12,162	2 7	>
		Missing form label	613	7	>
		Missing alternative text	309	3	>
	h-	Empty heading	26	7	>

>>>

$\equiv \mathcal{P}$ pope**tech**

		Andreas Burkard 🔻
Stuttgart Media Univ × 💌		Dashboard Detail
📥 Dashboard	Result Details	
Accessibility >		
Websites	Filters	
🚠 Groups	Website Category	
Lusers >	The second secon	
Settings		
	Results	
	Errors	



Missing alternative text

What it means:

Image alternative text is not present.

Why it matters:

Each image must have an alt attribute. Without alternative text, the content of an image will not be available to screen reader users or when the image is unavailable.

How to fix it:

Add an alt attribute to the image. The attribute value should accurately and succinctly present the content and function of the image. If the content of the image is conveyed in the context or surroundings of the image, or if the image does not convey content or have a function, it should be given empty/null alternative text (alt="").

The Algorithm... In English

An image does not have an alt attribute.

WCAG 2.1 Success Criteria:

1.1.1 Non-text Content (Level A)







Diese Website verwendet Cookies. Durch die Nutzung dieser Website erklären Sie sich damit einverstanden, dass Cookies gesetzt





VERSTANDEN

Pope Tech - Summary







Definitions

- A true positive (TP) is an error that was indicated as an error by the accessibility monitoring system and which is found to be an actual error by manual checking.
- A false positive (FP) an error that was reported as an error by the accessibility monitoring system, but which is not found to be an accessibility error by manual checking.
 - If an accessibility monitoring system identifies a finding as a true accessibility error, but it does not break any success criteria, is only a usability error or it is best practice, then it is considered a FP.
 - If found accessibility errors are no barriers, for example because an error is located in an element that is always hidden for every user, then this is also evaluated as FP.





Definitions

• A false negative (FN) - a true error that was not recognized as an error by the accessibility monitoring system.

Note: whether it is a TP or a FP has been checked manually and in case of controversial questions has been discussed with two accessibility experts.





Coverage of webpages (weight: 10,49 %) (Abduganiev, 2017; Vigo, 2013)

- Number of crawled websites. No duplicate websites are counted or jump marks to the same page. Only pages with new content count.
- Automatic authentication
- Evaluation of processes.

Coverage of success criteria (w: 10,84 %)

Number of violated success criteria found.





Completeness (w: 9,42 %) (Abduganiev, 2017; Vigo, 2013)

• Relation between true positives and false *negatives*.

Correctness (w: 9,59%) (Abduganiev, 2017; Vigo, 2013)

• Relation between true positives and false *positives*.

Support for localization of errors (w: 10,49 %)

• How well are the errors localized on a webpage?

Support for manual checks (w: 9,20 %)

• How well is manual reporting of errors supported?





User experience (empirical) (w: 14,96 %)

• User study results with User Experience Questionnaire (Laugwitz, Held, and Schrepp, 2008).

Gamification Patterns (w: 5,62 %)

• Number and types of gamification patterns (Majuri, Koivisto, and Hamari, 2018).

Input formats (w: 5,50 %)

• Range of file formats that can be evaluated.





Report formats (w: 3,54 %)

Range of report file formats for export.

Methodology Support for WCAG-EM (w: 4,61 %)

 Support for the Website Accessibility Conformance Evaluation Methodology 1.0 (WCAG-EM) methodology (Eric Velleman, 2014).

Methodology Support for German BITV-Test (w: 5,73 %)

• Support for the BITV-Test methodology (BITV-Test, 2019).




Meeting of Experts

Some of the evaluation criteria were derived from the literature or directly adopted, others were newly defined.

In order to ensure that the evaluation criteria were based on a scientific foundation, six experts (professors working in appropriate fields and accessibility experts) were invited in the context of this work to discuss and vote on the weighting of the evaluation criteria.

These experts also voted on the weighting of the six scales of the "User Experience Questionnaire" (UEQ) (Laugwitz, Held, and Schrepp, 2008) for the evaluation criterion: User Experience (empirical), which are necessary to combine the values into a single value, the "Key Performance Indicator" (KPI).





Samples for Evaluation Criteria

Samples from selected websites are used for the following evaluation criteria:

- Coverage of Success Criteria
- Completeness
- Correctness
- Support for Manual Checks

Samples of HdM website:

- <u>https://www.hdm-stuttgart.de/</u>
- <u>https://www.hdm-stuttgart.de/hochschule/profil/qm</u>
- <u>https://www.hdm-stuttgart.de/science</u>

Samples of HdM Digitization website:

- <u>https://Digitization.hdm-stuttgart.de/</u>
- <u>https://Digitization.hdm-stuttgart.de/barrierefreiheit/barrieren-melden/</u>





Settings and Rules Part 1

• The evaluation criteria of the accessibility monitoring system and the pertaining browser extension were combined.

- For the evaluation criterion "coverage of webpages" the scan depth was 1, i.e. main page + one level of subpages.
- The conformance level tested against was WCAG 2.1 AAA.
- The scan of May 19, 2020 was used.





Settings and Rules Part 2

- If a true positive error violates several success criteria, then this error counts as an error also per violated success criterion.
- We have also manually checked which errors fit which success criteria and which do not.
- "Manual checks" are the manual checks and findings of all accessibility monitoring systems combined.





Data of the Evaluation Criteria

- The values of the results of the evaluation criteria are either in a range from 0 to 1 or are converted to this value range by normalization.
- For the final result, the normalized results of all evaluation criteria are multiplied by their respective weightings and then summed up.





<u>Download Excel file</u> Evaluation Criteria	Weights	Siteimprove	axe Monitor	ARC Monitoring	Pope Tech
Coverage of webpages	10,49%	0,74	<u>1,00</u>	0,75	0,75
Coverage of success criteria	10,84%	<u>1,00</u>	0,67	0,87	0,83
Completeness	9,42%	0,97	0,62	<u>1,00</u>	0,70
Correctness	9,59%	0,85	<u>1,00</u>	0,65	0,83
Support for localization of errors	10,49%	<u>1,00</u>	0,60	0,80	0,80
Support for manual checks	9,20%	0,23	<u>1,00</u>	0,88	0,36
User experience (empirical)	14,96%	<u>1,00</u>	0,09	0,07	0,73
Gamification Patterns	5,62%	<u>1,00</u>	0,24	0,48	0,12
Input formats	5,50%	<u>1,00</u>	<u>1,00</u>	0,50	0,50
Report formats	3,54%	<u>1,00</u>	<u>1,00</u>	<u>1,00</u>	<u>1,00</u>
Methodology Support for WCAG-EM	4,61%	0,75	<u>1,00</u>	0,80	0,70
Methodology Support for BITV-Test	5,73%	<u>1,00</u>	<u>1,00</u>	<u>1,00</u>	0,87
<u>Result Index:</u>	100 %	#1: <u>0,87</u>	#2: 0,71	#3: 0,69	#3: 0,69

Coverage of Webpages (Scan Depth: 1)

Download Excel file	Siteimprove	Pope Tech	ARC Monitoring	axe Monitor
Pages crawled (HdM) w: 25%	155 pages	157 pages	158 pages	160 pages
	= 0.97	= 0.98	= 0.99	= 1
Pages crawled (Digitization)	19 pages	19 pages	19 pages	19 pages
w: 25%	= 1	= 1	= 1	= 1
Automatic Authentication?	Yes.	Yes.	Yes.	Yes.
w: 25%	= 1	= 1	= 1	= 1
Can scans monitor processes?	No.	No.	No.	Yes with recorded scripts.
w: 25%	= 0	= 0	= 0	= 1
Result (normalized)	0.74	0.75	0.75	<u>1</u>

Coverage of Success Criteria

Violated success criteria found with at least one true positive.

Formula:

coverage of success criteria = $\frac{a}{b}$ where:

SUCCESS

ЧO

a : violated success criteria found by a specific tool.

b: total found violated success criteria by all tools and manual checks combined.

COVERAGE OF SUCCESS CRITERIA



Coverage of Success Criteria (Table View)

Download Excel file	axe Monitor	Pope Tech	ARC Monitoring	Siteimprove	Manual checks
Number of found violated success criteria (HdM).	14	18	16	26	48
Violated success criteria found in % (HdM). w: 50 %	29%	38%	33%	<u>54%</u>	100%
Number of found violated success criteria (Digitization).	5	6	7	6	14
Violated success criteria found in % (Digitization). w: 50 %	36%	43%	<u>50%</u>	43%	100%
Result (normalized)	0,67	0,83	0,87	<u>1,00</u>	

Completeness

Relation between true positives and false negatives.

Formula:

Completeness = $\frac{h}{i}$

where:

h : true positives found by a single tool.

i : total number of true positives found.



Completeness (Table View)

Download Excel file	axe Monitor	Pope Tech	Siteimprove	ARC Monitoring	Manual checks
True Positives (HdM)	152	149	260	146	417
False Negatives (HdM)	561	564	453	567	0
Completeness (HdM) w: 50 %	21%	21%	<u>36%</u>	20%	100%
True Positives (Digitization)	11	15	15	29	41
False Negatives (Digitization)	61	57	57	43	0
Completeness (Digitization) w: 50 %	15%	21%	21%	<u>40%</u>	100%
Result normalized	0.62	0.70	0.97	<u>1.00</u>	

Correctness

Relation between true positives and false positives. This criterion can only be checked manually.

Formula:

Correctness = $\frac{j}{j+k}$

where:

j : true positives found by a single tool.

k : false positives found by a single tool.



Correctness (Table View)

Download Excel file	ARC Monitoring	Pope Tech	Siteimprove	axe Monitor
True positives (HdM)	146	149	260	152
False positives (HdM)	290	56	24	0
Correctness (HdM)	33 %	73 %	92 %	100 %
True positives (Digitization)	29	15	15	11
False positives (Digitization)	1	1	4	0
Correctness (Digitization)	97 %	94 %	79 %	100 %
Result normalized	0.65	0.83	0.85	<u>1.00</u>

Support for Localization of Errors

Download Excel file	Opens directly the page with error in the browser extension	Highlights errors	Scrolls to the error	Image preview	Jump to error code	Sum	Result normalized
Siteimprove	1	1	1	1	1	<u>5</u>	<u>1.0</u>
Pope Tech	1	1	1		1	4	0.8
ARC Monitoring		1	1	1	1	4	0.8
axe Monitor		1	1		1	2	0.6





Support for Manual Checks

Evaluates the reporting of possible errors, which helps the user to find true positives.

Formula:

Support for manual checks $= \frac{m}{o}$

where:

m : potential errors found by a specific tool, which result in a true positive error.

o : all found true positives.

HDM (A, AA AND AAA)

Possible errors to check them manually True positives that have emerged from these checks \mathbf{H} NUMBER OF FINDINGS 99 251 234 53 96 27 52 σ 4 0 SITEIMPROVE POPETECH ARC MONITORING AXE MONITOR MANUAL CHECKS DIGITALISIERUNG (A, AA AND AAA) Possible errors to check them manually True positives that have emerged from these checks 171



Support for Manual Checks (Table View)

Download Excel file	Siteimprove	Pope Tech	ARC Monitoring	axe Monitor	Manual checks
Possible errors to check them manually (HdM)	266	234	251	96	0
True positives that have emerged from these checks (HdM)	49	77	153	62	713
Support for manual checks (HdM) w: 50%	7 %	11 %	<u>21 %</u>	9 %	
Possible errors to check them manually (Digitization)	171	0	43	20	0
True positives that have emerged from these checks (Digitization)	0	0	3	13	72
Support for manual checks (Digitization) w: 50%	0 %	0 %	4 %	<u>18 %</u>	
Result (normalized)	0.23	0.36	0.88	1	

User Experience (Empirical) – User Study

• We conducted a user study with 15 participants. The target group for the user test consisted of website administrators from the HdM and students who had attended at least one lecture on accessibility and web development.

• The User Experience Questionnaire (UEQ) (Laugwitz, Held, and Schrepp, 2008) was used for the evaluation. The weighting of the scales of the UEQ by the meeting of experts was as follows:

Scale		Weight	t
Attractivene	ess	5,17	
Efficiency		6,83	
Perspicuity		6,67	
Dependabil	ity	5,83	
Stimulation		3,50	
Novelty		2,50	
Г	Download I	Excel file	





User Study – Methodologies Part 1

- Free Exploration Test (Goodman, 2012)
 - Instead of concrete tasks, users had 15 minutes per accessibility monitoring system and browser extension used by the company to look at the tool and check its functionality.
- Within-Subjects (Nielsen, 1994)
 - Every participant has tested every tool.
- Counterbalancing (Albert, 2013)
 - To avoid fatigue or learning effects, counterbalancing was used, which in this case means that the order in which the users tested the tools was changed for each session. Care was taken to ensure that each tool was tested as often as possible on each position.





User Study – Methodologies Part 2

- Coaching (Nielsen, 1994)
 - The participants were allowed to ask questions about the tools, which person who performed the experiment answered as well as possible and in as equal detail as possible for all tools.
- Think-aloud protocol (Nielsen, 1994)
 - The users were asked to think out loud during the user test.
- Participants who gave too contradictory information according to the UEQ's instructions were excluded from this study.
 - This was the case with two participants, so there are now only 13 evaluated data sets.







User Experience (Empirical) – Siteimprove

KPI = 1.33 \rightarrow normalized result = <u>1.0</u>

Distribution of Answers for Siteimprove per Item (Table View) Part 1

Item	1	2	3	4	5	6	7	Scale
annoying / enjoyable	0	1	0	3	5	2	2	Attractiveness
not understandable / understandable	0	0	1	1	3	5	3	Perspicuity
dull / creative	0	1	1	4	3	3	1	Novelty
difficult to learn / easy to learn	0	0	2	1	6	1	3	Perspicuity
inferior / valuable	0	0	0	2	2	8	1	Stimulation
boring / exciting	0	0	1	1	7	4	0	Stimulation
not interesting / interesting	0	0	0	2	5	5	1	Stimulation
unpredictable / predictable	0	0	0	2	4	4	3	Dependability
slow / fast	0	0	2	5	3	2	1	Efficiency
conventional / inventive	1	0	2	1	5	4	0	Novelty
obstructive / supportive	0	0	0	1	2	6	4	Dependability
bad / good	0	0	0	1	4	6	2	Attractiveness
complicated / easy	0	1	1	1	3	5	2	Perspicuity
unlikable / pleasing	0	0	0	2	5	4	2	Attractiveness
usual / leading edge	0	2	1	3	5	2	0	Novelty
unpleasant / pleasant	0	0	0	2	4	4	3	Attractiveness
not secure / secure	0	1	0	5	2	5	0	Dependability
demotivating / motivating	0	0	0	3	1	5	4	Stimulation
does not meet expectations / meets expectations	0	0	1	1	3	4	4	Dependability





Distribution of Answers for Siteimprove per Item (Table View) Part 2

Item	1	2	3	4	5	6	7	Scale
inefficient / efficient	0	0	1	1	3	5	3	Efficiency
confusing / clear	0	1	0	2	2	6	2	Perspicuity
impractical / practical	0	0	1	1	3	6	2	Efficiency
cluttered / organized	1	0	1	1	2	5	3	Efficiency
unattractive / attractive	0	0	1	3	2	4	3	Attractiveness
unfriendly / friendly	0	0	0	3	5	3	2	Attractiveness
conservative / innovative	0	0	4	1	2	6	0	Novelty





Benchmark for Siteimprove (Table View)

Scale	Mean	Comparison to benchmark	Interpretation
Attractiveness	1,41	Above Average	25% of results better, 50% of results worse
Perspicuity	1,35	Above Average	25% of results better, 50% of results worse
Efficiency	1,27	Above Average	25% of results better, 50% of results worse
Dependability	1,52	Good	10% of results better, 75% of results worse
Stimulation	1,46	Good	10% of results better, 75% of results worse
Novelty	0,60	Below Average	50% of results better, 25% of results worse







User Experience (Empirical) – Pope Tech

KPI = 0.93 \rightarrow normalized result = <u>0.73</u>

Distribution of Answers for Pope Tech per Item (Table View) Part 1

Item	1	2	3	4	5	6	7	Scale
annoying / enjoyable	0	0	3	2	3	3	2	Attractiveness
not understandable / understandable	0	1	0	1	4	5	2	Perspicuity
dull / creative	0	0	2	2	4	4	1	Novelty
difficult to learn / easy to learn	0	0	1	3	5	3	1	Perspicuity
inferior / valuable	0	0	0	3	5	4	1	Stimulation
boring / exciting	0	0	3	3	5	2	0	Stimulation
not interesting / interesting	0	0	0	5	4	3	1	Stimulation
unpredictable / predictable	0	1	1	2	6	3	0	Dependability
slow / fast	0	0	1	5	4	3	0	Efficiency
conventional / inventive	0	0	0	3	6	3	1	Novelty
obstructive / supportive	0	0	1	1	5	5	1	Dependability
bad / good	0	0	0	1	3	7	2	Attractiveness
complicated / easy	0	0	2	1	5	2	3	Perspicuity
unlikable / pleasing	0	0	1	4	5	2	1	Attractiveness
usual / leading edge	0	0	1	5	5	2	0	Novelty
unpleasant / pleasant	0	0	0	6	5	2	0	Attractiveness
not secure / secure	0	0	1	4	4	4	0	Dependability
demotivating / motivating	0	0	2	0	6	4	1	Stimulation
does not meet expectations / meets expectations	0	0	1	2	4	5	1	Dependability



Distribution of Answers for Pope Tech per Item (Table View) Part 2

Item	1	2	3	4	5	6	7	Scale
impractical / practical	0	0	1	2	7	1	2	Efficiency
cluttered / organized	0	2	2	2	1	5	1	Efficiency
unattractive / attractive	0	0	1	2	3	5	2	Attractiveness
unfriendly / friendly	0	0	1	2	7	2	1	Attractiveness
conservative / innovative	0	0	1	4	3	4	1	Novelty





Benchmark for Pope Tech (Table View)

Scale	Mean	Comparison to benchmark	Interpretation
Attractiveness	1,10	Below Average	50% of results better, 25% of results worse
Perspicuity	0,96	Below Average	50% of results better, 25% of results worse
Efficiency	0,87	Below Average	50% of results better, 25% of results worse
Dependability	1,02	Below Average	50% of results better, 25% of results worse
Stimulation	0,96	Below Average	50% of results better, 25% of results worse
Novelty	0,94	Above Average	25% of results better, 50% of results worse







User Experience (Empirical) – axe Monitor

KPI = 0.11 \rightarrow normalized result = 0.09

Distribution of Answers for axe Monitor per Item (Table View) Part 1

Item	1	2	3	4	5	6	7	Scale
annoying / enjoyable	0	1	6	3	2	1	0	Attractiveness
not understandable / understandable	0	2	4	2	3	1	1	Perspicuity
dull / creative	1	6	3	2	1	0	0	Novelty
difficult to learn / easy to learn	0	3	1	5	3	1	0	Perspicuity
inferior / valuable	0	0	1	4	5	3	0	Stimulation
boring / exciting	0	4	2	4	3	0	0	Stimulation
not interesting / interesting	0	2	3	2	5	1	0	Stimulation
unpredictable / predictable	0	0	1	2	7	2	1	Dependability
slow / fast	0	0	2	7	0	4	0	Efficiency
conventional / inventive	1	5	4	2	1	0	0	Novelty
obstructive / supportive	0	0	2	2	6	2	1	Dependability
bad / good	0	1	2	2	4	3	1	Attractiveness
complicated / easy	0	3	3	5	1	1	0	Perspicuity
unlikable / pleasing	0	2	3	7	0	1	0	Attractiveness
usual / leading edge	1	3	4	3	2	0	0	Novelty
unpleasant / pleasant	0	3	2	2	5	1	0	Attractiveness
not secure / secure	0	0	0	4	6	3	0	Dependability
demotivating / motivating	1	1	5	4	2	0	0	Stimulation
does not meet expectations / meets expectations	0	0	3	1	4	3	2	Dependability



Distribution of Answers for axe Monitor per Item (Table View) Part 2

Item	1	2	3	4	5	6	7	Scale
inefficient / efficient	0	0	3	3	3	3	1	Efficiency
confusing / clear	0	1	6	2	2	1	1	Perspicuity
impractical / practical	0	2	2	1	4	3	1	Efficiency
cluttered / organized	0	2	2	3	3	2	1	Efficiency
unattractive / attractive	2	3	4	1	2	1	0	Attractiveness
unfriendly / friendly	0	1	4	4	3	1	0	Attractiveness
conservative / innovative	1	6	4	0	1	1	0	Novelty





Benchmark for axe Monitor (Table View)

Scale	Mean	Comparison to benchmark	Interpretation
Attractiveness	-0,18	Bad	In the range of the 25% worst results
Perspicuity	-0,17	Bad	In the range of the 25% worst results
Efficiency	0,50	Bad	In the range of the 25% worst results
Dependability	0,94	Below Average	50% of results better, 25% of results worse
Stimulation	-0,10	Bad	In the range of the 25% worst results
Novelty	-1,15	Bad	In the range of the 25% worst results







User Experience (Empirical) – ARC Monitoring

KPI = 0.09 \rightarrow normalized result = <u>0.07</u>

Distribution of Answers for ARC Monitoring per Item (Table View) Part 1

Item	1	2	3	4	5	6	7	Scale
annoying / enjoyable	1	4	0	5	2	1	0	Attractiveness
not understandable / understandable	1	4	1	0	6	1	0	Perspicuity
dull / creative	0	3	4	3	2	0	1	Novelty
difficult to learn / easy to learn	1	4	1	2	3	1	1	Perspicuity
inferior / valuable	0	1	1	3	5	2	1	Stimulation
boring / exciting	0	3	5	3	2	0	0	Stimulation
not interesting / interesting	1	1	3	3	4	1	0	Stimulation
unpredictable / predictable	1	2	0	3	4	2	1	Dependability
slow / fast	0	0	0	7	5	0	1	Efficiency
conventional / inventive	2	0	5	2	3	1	0	Novelty
obstructive / supportive	0	1	3	3	3	1	2	Dependability
bad / good	0	1	3	1	5	1	2	Attractiveness
complicated / easy	0	3	2	1	3	3	1	Perspicuity
unlikable / pleasing	0	4	2	1	5	1	0	Attractiveness
usual / leading edge	1	1	4	4	3	0	0	Novelty
unpleasant / pleasant	1	2	2	1	5	2	0	Attractiveness
not secure / secure	0	0	0	6	4	3	0	Dependability
demotivating / motivating	1	2	2	1	4	2	1	Stimulation
does not meet expectations / meets expectations	1	1	1	1	5	3	1	Dependability



Distribution of Answers for ARC Monitoring per Item (Table View) Part 2

Item	1	2	3	4	5	6	7	Scale
inefficient / efficient	1	0	4	0	5	1	2	Efficiency
confusing / clear	0	4	3	1	2	3	0	Perspicuity
impractical / practical	0	3	1	4	2	2	1	Efficiency
cluttered / organized	2	1	3	1	2	3	1	Efficiency
unattractive / attractive	2	1	3	3	2	2	0	Attractiveness
unfriendly / friendly	0	3	2	2	3	3	0	Attractiveness
conservative / innovative	2	1	3	3	4	0	0	Novelty





Benchmark for ARC Monitoring (Table View)

Scale	Mean	Comparison to benchmark	Interpretation
Attractiveness	-0,08	Bad	In the range of the 25% worst results
Perspicuity	-0,13	Bad	In the range of the 25% worst results
Efficiency	0,31	Bad	In the range of the 25% worst results
Dependability	0,54	Bad	In the range of the 25% worst results
Stimulation	0,00	Bad	In the range of the 25% worst results
Novelty	-0,46	Bad	In the range of the 25% worst results





User Experience (Empirical) - Results

Download Excel file	Siteimprove	Pope Tech	axe Monitor	ARC Monitoring
КРІ	1,33	0,97	0,11	0,09
Result (normalized)	-> <u>1,00</u>	-> 0,73	-> 0,09	-> 0,07




Weighting of the Gamification Patterns Part 1

- In the literature review by Majura et al. (Majuri, Koivisto, & Hamari, 2018) studies for each Gamification Pattern were quantified in a table.
 - We added the "weighting" column, which is calculated as follows:
 - weighting = number of positive papers / (positive papers + equal papers + negative papers).

Affordance <u>Download Excel file</u>	Mainly positively oriented	Null or equal positive and negative	Mainly negatively oriented	Sum	Weighting
Points, score, XP	38	13	1	52	73.08%
Leaderboards, ranking	27	13	3	43	62.79%
Badges, achievements, medals, trophies	25	12	2	39	64.10%
Challenges, quests, missions, tasks, clear goals	27	8	2	37	72.97%
Levels	19	7	2	28	67.85%
Cooperation, teams	17	2	2	21	80.95%
Quizzes, questions	15	3	18		83.33%
Progress, status bars, skill trees	13	2	1	16	81.25%
Social networking features	11	1	2	14	78.57%
Performance stats, performance feedback	13	1		14	92.86%

Weighting of the Gamification Patterns Part 2

Affordance	Download Excel file	Mainly positively oriented	Null or equal positive and negative	Mainly negatively oriented	Sum	Weighting
Timer, speed		12			12	100%
Narrative, narratio	n, storytelling, dialogues, theme	10	1		11	90.91%
Avatar, character, v	rirtual identity	8	1		9	88.89%
Competition		7	1		8	87.50%
Virtual currency		3	1		4	75%
Full game (also boa	ard games), also undescribed	1	2		3	33.33%
commercial gamific	cation systems					
Reminders, cues, n	otifications, annotations	1	1		2	50%
Real world/financia	al reward	1	1		2	50%
Role play		1			1	100%
Game rounds		1			1	100%
Motion tracking		1			1	100%
Penalties		1			1	100%
Total		297	74	15	386	

Gamification Patterns

• The evaluation criterion "Gamification Patterns" is the number of gamification patterns multiplied by the respective weighting. The "weighting" (w) corresponds to the previous table.

Download Excel file		Siteimprove	ARC Monitoring	axe Monitor	Pope Tech
Points, score, XP (w: 0,73)		2		1	
Progress, status bars, skill trees (w: 0,81)		2	1	1	1
Competition (w: 0,88)		1	1		
Challenges, quests, missions, tasks, clear		1	1		
goals (w: 0,73)		-	1		
Performance stats, performance		2	1		
feedback (w: 0,93)		Z	T		
Result		<u>6,55</u>	3,15	1,54	0,81
Result (normalized)		<u>1,00</u>	0,48	0,24	0,12





Input Formats

Download Excel file	axe Monitor	Siteimprove	ARC Monitoring	Pope Tech
HTML	1	1	1	1
PDF	1	1	0	0
Result	2	<u>2</u>	1	1
Result (normalized)	<u>1</u>	<u>1</u>	0.50	0.50





Report Formats

Download Excel file	Siteimprove	ARC Monitoring	Pope Tech	axe Monitor
HTML	1	0	1	1
PDF	1	1	1	0
XLSX	0	1	0	1
CSV	1	1	1	1
Result	<u>3</u>	<u>3</u>	<u>3</u>	<u>3</u>
Result (normalized)	<u>1.0</u>	<u>1.0</u>	<u>1.0</u>	<u>1.0</u>





Methodology Support for WCAG-EM Teil 1

Download Excel file	axe Monitor	Siteimprove	Pope Tech	ARC Monitoring	Comments
Step 1.a: Define the Scope of the Website	1	1	1	1	Each of the tools can define which pages should be checked.
					Siteimprove: Can check websites for conformance level AAA.
Sten 1 h. Define the Conformance Target	0.5	1	0	0.5	Pope Tech: The conformance level is not changable.
Step 1.5. Denne the comormance rarget	0.5	1	U	0.5	ARC Monitoring and axe Monitor: They have no "AAA"
					conformance level.
Stop 1 a. Dafina an Accessibility Sympost Basalina	1	0	0.5	0	Pope Tech: The Viewport is customizable.
Step 1.C. Denne an Accessionity Support Baseline	T	0	0.5	U	axe Monitor: Agent is selectable.
Step 1.d: Define Additional Evaluation Requirements	1	1	1	1	Each of the tools can sort by errors and show how to fix them
(Optional)	T	T	T	1	Lach of the tools can solt by enors and show how to fix them.
Step 2.a: Identify Common Web Pages of the Website	0	0	0	0	None of the tools has a function that supports this.
Step 2.b: Identify Essential Functionality of the Website	0	0	0	0	None of the tools has a function that supports this.
Step 2.c: Identify the Variety of Web Page Types	0	0	0	0	None of the tools has a function that supports this.
Step 2.d: Identify Web Technologies Relied Upon	0	0	1	0	Pope Tech: Shows which technologies are used on which pages
Step 2.e: Identify Other Relevant Web Pages	0	0	0	0	None of the tools has a function that supports this.
Step 3.a: Include a Structured Sample	0	0	0	0	None of the tools has a function that supports this.
Step 3.b: Include a Randomly Selected Sample	0	0	0	0	None of the tools has a function that supports this.
Step 3.c: Include Complete Processes	1	0	0	0	axe Monitor: Recordable scripts for performing processes.
Step 4.a: Check All Initial Web Pages	1	1	1	1	Each of the tools can check the initial pages.
Step 4.b: Check All Complete Processes	1	0	0	0	axe Monitor: Recordable scripts for performing processes.
Step 4.c: Compare Structured and Random Samples	1	1	1	1	With each of the tools pages can be compared.
Ston E as Desument the Outcomes of Each Ston	1	1	1	1	Each of the tools can generate reports that offer various
Step S.a. Document the Outcomes of Each Step	T	T	T	1	possibilities for documentation.
Step 5.b: Record the Evaluation Specifics (Optional)	0	0	0	0	None of the tools has a function that supports this.

Methodology Support for WCAG-EM Teil 2

Download Excel file	axe Monitor	Siteimprove	Pope Tech	ARC Monitoring	Comments
Step 5.c: Provide an Evaluation Statement (Optional)	0	0	0	0	None of the tools has a function that supports this.
Step 5.d: Provide an Aggregated Score (Optional)	1	1	0	1	Pope Tech: Has no score.
Step 5.e: Provide Machine-Readable Reports (Optional)	0.5	0.5	0.5	0.5	None of the tools support EARL
Result	<u>10</u>	7.5	7	7	
Result normalized	-> <u>1.0</u>	-> 0.75	-> 0.70	-> 0.70	





Download Excel file	axe Monitor	Siteimprove	ARC Monitoring	Pope Tech	Comments
Evaluation Methodology:					
	2	1			Siteimprove: PDFs
Evaluation Methodology:	2	1			axe Monitor: Agents, PDFs
1.1 What belongs to the test item?	2	2	1	1	Siteimprove: PDFs
	۷	Z	1	1	axe Monitor: PDFs
6.1. Analysis of the web presence	1	1	1	1	Different barriers can be detected with each tool.
6.2.3. Cover all harriers	1	1	1	1	With each tool the user can identify different page types based
	_	-	-	1	on the error list.
6.2.4. Include different page types	1				axe Monitor: Recordable scripts for performing processes.
6.2.5 Include different name states					None of the tools can automatically search for pages with
0.2.3. Include different page states					different functions.
Evaluation Steps:					
1.1.1a Alternative texts for control elements	1	1	1	1	
1.1.1b Alternative texts for graphics and objects	1	1	1	1	
1.1.1c Empty alt attributes for layout graphics	1	1	1	1	
1.1.1d Alternatives for CAPTCHAs					
1.2.1a Audio-only and Video-only (Prerecorded)	1	1		1	
1.2.2a Captions (Prerecorded)	1	1	1	1	
1.2.3a Audio Description or Media Alternative	1	1	1	1	
1.2.4a Captions (Live)		1			
1.2.5a Audio Description (Prerecorded)	1	1	1	1	
1.3.1a Info and Relationships for Titles		1	1	1	
1.3.1b Info and Relationships for Lists	1		1		
1.3.1c Info and Relationships for Citations					

Download Excel file	axe Monitor	Siteimprove	ARC Monitoring	Pope Tech	Comments
1.3.1d Content Structured	1	1	1	1	
1.3.1e Data Tables Correctly Structured	1	1	1	1	
1.3.1f Assignment of table cells	1	1	1	1	
1.3.1g No Structure Markup for Layout Tables	1	1	1	1	
1.3.1.h Labeling of Form Elements can be Determined Programmatically	1	1	1	1	
1.3.2a Meaningful Sequence			1	1	
1.3.3a Sensory Characteristics			1		
1.3.4a Orientation	1		1		
1.3.5a Identify Input Purpose	1		1		
1.4.1a Use of Color	1	1	1		
1.4.2a Audio Control	1	1		1	
1.4.3a Contrast (Minimum)	1	1	1	1	
1.4.4a Resize text	1		1		
1.4.5a Images of Text		1			
1.4.10a Reflow					
1.4.11a Non-text Contrast					

Download Excel file	axe Monitor	Siteimprove	ARC Monitoring	Pope Tech	Comments
1.4.12a Text Spacing	1		1		
1.4.13a Content on Hover or Focus					
2.1.1a Keyboard	1		1	1	
2.1.2a No Keyboard Trap				1	
2.1.4a Character Key Shortcuts					
2.2.1a Timing Adjustable	1	1	1	1	
2.2.2a Pause, Stop, Hide	1		1	1	
2.3.1a Three Flashes or Below Threshold					
2.4.1a Bypass Blocks	1	1	1	1	
2.4.2a Page Titled	1	1	1	1	
2.4.3a Focus Order		1	1	1	
2.4.4a Link Purpose (In Context)	1	1	1	1	
2.4.5a Multiple Ways		1			
2.4.6a Headings and Labels		1	1	1	
2.4.7a Focus Visible		1			
2.5.1a Pointer Gestures					
2.5.2a Pointer Cancellation					
2.5.3a Label in Name	1		1		
2.5.4a Motion Actuation					
3.1.1a Language of Page	1	1	1	1	
3.1.2a Language of Parts	1	1	1		
3.2.1a On Focus					
3.2.2a On Input		1		1	

Download Excel file	axe Monitor	Siteimprove	ARC Monitoring	Pope Tech	Comments
3.2.3a Consistent Navigation					
3.2.4a Consistent Identification					
3.3.1a Error Identification		1		1	
3.3.2a Labels or Instructions	1	1	1	1	
3.3.3a Error Suggestion		1	1		
3.3.4a Error Prevention (Legal, Financial, Data)					
4.1.1a Parsing	1	1	1		
4.1.2a Name, Role, Value	1	1	1	1	
4.1.3a Status Messages				1	
Result	<u>39</u>	<u>39</u>	<u>39</u>	34	
Result (normalized)	<u>1</u>	<u>1</u>	1	0.87	





<u>Download Excel file</u> Evaluation Criteria	Weights	Siteimprove	axe Monitor	ARC Monitoring	Pope Tech
Coverage of webpages	10,49%	0,74	<u>1,00</u>	0,75	0,75
Coverage of success criteria	10,84%	<u>1,00</u>	0,67	0,87	0,83
Completeness	9,42%	0,97	0,62	<u>1,00</u>	0,70
Correctness	9,59%	0,85	<u>1,00</u>	0,65	0,83
Support for localization of errors	10,49%	<u>1,00</u>	0,60	0,80	0,80
Support for manual checks	9,20%	0,23	<u>1,00</u>	0,88	0,36
User experience (empirical)	14,96%	<u>1,00</u>	0,09	0,07	0,73
Gamification Patterns	5,62%	<u>1,00</u>	0,24	0,48	0,12
Input formats	5,50%	<u>1,00</u>	<u>1,00</u>	0,50	0,50
Report formats	3,54%	<u>1,00</u>	<u>1,00</u>	<u>1,00</u>	<u>1,00</u>
Methodology Support for WCAG-EM	4,61%	0,75	<u>1,00</u>	0,80	0,70
Methodology Support for BITV-Test	5,73%	<u>1,00</u>	<u>1,00</u>	<u>1,00</u>	0,87
<u>Result Index:</u>	100 %	#1: <u>0,87</u>	#2: 0,71	#3: 0,69	#3: 0,69

Thank you for your attention!





- Abascal, J. a. (2019). Tools for Web Accessibility Evaluation. In Y. a. Yesilada (Ed.), *Web Accessibility: A Foundation for Research* (pp. 79-503). London: Springer London.
- Abduganiev, S. G. (2017). Towards Automated Web Accessibility Evaluation: a Comparative Study. *Int. J. Inf. Technol. Comput. Sci. (IJITCS)*(9), 18-44.
- Albert, W. a. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics.* Newnes.
- BIK-Projekt. (2019). *BIK BITV-Test*. Retrieved September 5, 2020, from https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/verfahren.html





Deterding, S. a. (2011). From game Design Elements to Gamefulness: Defining Gamification. In *Proceedings* of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9-15).

European Commission. (2010). European Disability Strategy 2010-2020: A Renewed Commitment to a Barrier-Free Europe. European Commission Brussels.

European Commission and others. (2016). Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies (Text with EEA relevance). Retrieved from https://eurlex.europa.eu/eli/dir/2016/2102/oj





European Telecommunications Standards Institute. (2019). Draft EN 301 549 - V3.1.1 - Accessibility requirements suitable for public procurement of ICT products and services in Europe. Retrieved September 6, 2020, from https://www.etsi.org/deliver/etsi_en/301500_301599/301549/03.01.01_20/en_301549v030101a.pdf

Fahrmeir, L. a. (2015). *Multivariate statistische Verfahren*. Walter de Gruyter GmbH \& Co KG.

Schmutz, S. a. (2016). Implementing Recommendations From Web Accessibility Guidelines: Would They Also Provide Benefits to Nondisabled Users. (C. SAGE Publications Sage CA: Los Angeles, Ed.) *Human Factors, 4*, pp. 611-629.

Goodman, E. a. (2012). Observing the user experience: A Practitioner's Guide to User Research. Elsevier.

Hamari, J. a. (2014). Does gamification work? - A Literature Review of Empirical Studies on Gamification. 2014 47th Hawaii international conference on system sciences, pp. 3025 - 3034.





Laugwitz, B. a. (2008). Construction and Evaluation of a User Experience Questionnaire. (Springer, Ed.) Symposium of the Austrian HCI and usability engineering group, 63-76.

Majuri, J. a. (2018). Gamification of education and learning: A review of empirical literature. *Proceedings of the 2nd international GamiFIN conference, GamiFIN 2018*.

Nielsen, J. (1994). Usability Engineering. (M. Kaufmann, Ed.)

Pădure, M. a. (2019). Exploring the Differences Between Five Accessibility Evaluation Tools.

Solovieva, T. I. (2014). Monitoring for Accessibility and University Websites: Meeting the Needs of People with Disabilities. *Journal of Postsecondary Education and Disability*, 27(27), pp. 113-127.

Velleman, E. a.-Z. (2014). Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0. W3C Working Group Note. http://www.w3. org/TR/WCAG-EM.





Vigo, M. a. (2013). Benchmarking Web Accessibility Evaluation Tools: Measuring the Harm of Sole Reliance on Automated Tests. *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, 1-10.

World Wide Web Consortium and others. (2018). Web Content Accessibility Guidelines (WCAG) 2.1. World Wide Web Consortium.





Special Thanks

- We thank the vendors of Deque, Pope Tech, Siteimprove, and The Paciello Group for providing us with free trial versions of their tools and for their technical support.
- We also thank the numerous voluntary participants in the user study who took part despite the difficult conditions of the COVID-19 pandemic.



