Monitoring methodologies for web accessibility in the European Union

FINAL REPORT:
A study prepared for the European Commission
DG Communications Networks, Content & Technology
by:

Funka  empirica  EUROPEAN DISABILITY FORUM

Digital Single Market
This study was carried out for the European Commission by Funka, empirica, and the European Disability Forum.

This report was compiled and edited by Susanna Laurin, Andreas Cederbom, Jose Martinez-Useo (Funka), Lutz Kubitschke (empirica), Alejandro Moledo (EDF), Bart Simons (EDF), with the technical advice of Shadi Abou-Zahra (W3C).

Internal identification:

Contract number: 30-CE-0679566/00-74
SMART 2014/0061

Disclaimer:

By the European Commission, Directorate-General of Communications Networks, Content & Technology.

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission’s behalf may be held responsible for the use which may be made of the information contained therein.

doi:10.2759/483693

© European Union, 2016. All rights reserved. Certain parts are licensed under conditions to the EU.

Reproduction is authorised provided the source is acknowledged.
# Table of contents

## CHAPTER 1. INTRODUCTION

## CHAPTER 2. CURRENT MONITORING INITIATIVES IN EUROPE

2.1. CROSS-COUNTRY OVERVIEW IN WEB ACCESSIBILITY MONITORING

2.2. FURTHER INSIGHTS FROM SELECTED MONITORING INITIATIVES

2.2.1. Initial considerations

2.2.2. Web accessibility monitoring initiatives selected

2.2.3. In-depth analysis of the monitoring initiatives

## CHAPTER 3. VALIDATION EXERCISE OF A SET OF MONITORING METHODOLOGIES

3.1 PURPOSE OF THE VALIDATION EXERCISE

3.2. SELECTED MONITORING METHODS

3.3. DESCRIPTION OF THE VALIDATED “AVERAGE” METHODOLOGY

3.4. DESIGN OF THE VALIDATION EXERCISE

3.5. ANALYSIS OF THE EXTENT OF METHODS

## CHAPTER 4. RECOMMENDATIONS FOR THE DEVELOPMENT OF A EUROPEAN WEB ACCESSIBILITY MONITORING METHODOLOGY

4.1. INTRODUCTION

4.2. OVERALL DISCUSSION AND EVIDENCE BASE RECOMMENDATIONS

4.2.1. NUMBER OF WEBSITES TO BE ASSESSED

4.2.2. FREQUENCY OF ASSESSMENT

4.2.3. TESTING AND DOCUMENTATION

4.2.4. FORMAT OF REPORTING

4.2.5. COST OF WEB ACCESSIBILITY MONITORING

4.3. TOWARDS IMPLEMENTING A SUSTAINABLE WEB ACCESSIBILITY MONITORING SYSTEM AT EUROPEAN LEVEL

4.3.1. Conceptual considerations

4.3.2. Practical considerations
ANNEX 1. LIST OF STAKEHOLDERS CONSULTED FOR THE PURPOSES OF THE STUDY .................................................................91

ANNEX 2. DESCRIPTION OF THE NORWEGIAN METHODOLOGY "STATUS I RIKET" ..............................................................93

Step 1 .................................................................................................................................................................................93
Step 2 .................................................................................................................................................................................94
Step 3 .................................................................................................................................................................................94

USING THE RESULTS ......................................................................................................................................................95
LIST OF SUCCESS CRITERIA AND TESTS COVERED ........................................................................................................96

ANNEX 3. SUCCESS CRITERIA TESTING APPROACH TODAY AND FORESEEABLE IN THE FUTURE ........................................................................99
List of figures

Figure 1 – Overview of Type I initiatives according to repetition intervals.................13
Figure 2 – Overview of Type I initiatives according to types of websites covered in relation to website ownership..................................................................................................14
Figure 3 - Overview of Type I initiatives according to assessment technique applied.....14
Figure 4 - Overview of Type I initiatives according to assessment criteria applied.......15
Figure 5 - Overview of Type I initiatives according to parties performing the assessment ........................................................................................................................................16
Figure 6 - Overview of Type I initiatives according to numbers of websites assessed.....16
Figure 7 - Overview of Type I initiatives according types of publicly available feedback ..................................................................................................................................17
Figure 8 – Overview of Type II initiatives according to repetition intervals...............18
Figure 9 – Overview of Type II initiatives according to types of websites covered in relation to website ownership..................................................................................................19
Figure 10 - Overview of Type II initiatives according to assessment technique applied.19
Figure 11 - Overview of Type II initiatives according to assessment criteria applied......20
Figure 12 - Overview of Type II initiatives according to parties performing the assessment ........................................................................................................................................21
Figure 13 - Overview of Type II initiatives according to numbers of websites assessed.21
Figure 14 - Overview of Type II initiatives according types of publicly available feedback ..................................................................................................................................22
Figure 15 – Overview of Type III initiatives according to repetition intervals.............23
Figure 16 – Overview of Type III initiatives according to types of websites covered in relation to website ownership..................................................................................................23
Figure 17 - Overview of Type III initiatives according to assessment technique applied24
Figure 18 - Overview of Type III initiatives according to assessment criteria applied....25
Figure 19 - Overview of Type III initiatives according to numbers of websites assessed26
Figure 20 - Overview of Type III initiatives according types of publicly available feedback ..................................................................................................................................26
List of tables

Table 1 – Overview of web accessibility monitoring initiatives identified according to type.......................................................... 12
Table 2 – “Average” methodological approach observable according to type of initiative ................................................................. 28
Table 3 – Web accessibility monitoring initiatives selected for further investigation................................................................. 30
Table 4 - WCAG 2.0 requirements applied during the assessment procedure ................................................................. 37
Table 5 - Average effort required to validate a given website against WCAG 2.0 AA ................................................................. 41
Table 6 – Validated “average” methodology ................................................................................................................................. 51
Table 7 – Success Criteria covered .................................................................................................................................................. 55
Chapter 1. Introduction

The international evidence base on web accessibility monitoring has remained quite patchy and loosely organised up until now. Due to their thematic and conceptual foci, earlier comparative studies on measuring web accessibility (e.g. the WCAG 2.0 study and the MeAC studies) were only able to touch upon the monitoring theme to a certain extent and at a certain level of depth. In the last years, a confluence of key initiatives and policy developments in the field of web accessibility that can contribute to the development and implementation of a framework for web accessibility monitoring at national and EU level has occurred. These include:

- the ratification of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) by the EU in December 2010;
- the political agreement between the European Parliament and the Council of the European Union on the content of a “Directive of the European Parliament and of the Council on the accessibility of the websites and mobile applications of public sector bodies”, hereinafter called ‘the Directive’, and the subsequent adoption of the draft Directive by the Council on 18th July 2016;\(^1\) the finalisation of the work on standardization Mandate 376 to incorporate accessibility in public ICT procurements (February 2014) and the publication of the European Standard EN 301 549 V1.1.2, “Accessibility requirements suitable for public procurement of ICT products and services in Europe”;\(^2\) and
- the progressive adoption of Web Content Accessibility Guidelines 2.0 across Member states.

Article 8 on “Monitoring and Reporting” of the Directive requires the development by the European Commission, through implementing acts, of monitoring in a coordinated way and with a harmonised methodology at EU level. This opens new discussions and space for debate on the scientific and practical orientation of a monitoring methodology that must now be developed and used by all Member States.

The study has been carried out to satisfy three main objectives:

- Provide an up-to-date description of the state of web accessibility monitoring in the countries covered. This objective is satisfied through comprehensive data gathering from the countries covered and an empirical analysis of the data gathered to detect main trends and compose an overall picture of web accessibility monitoring initiatives in Europe.
- Validate a set of existing national monitoring methodologies for verifying how actually they measure web accessibility and how reliable these methods are. This objective is implemented through a practical exercise for

---


\(^2\) [http://www.etsi.org/deliver/etsi_en/301500_301599/301549/01.01.02_60/en_301549v010102p.pdf](http://www.etsi.org/deliver/etsi_en/301500_301599/301549/01.01.02_60/en_301549v010102p.pdf)
cross-checking 5 existing monitoring methodologies with WCAG Evaluation Methodology.

- Introduce a set of practical recommendations for a web accessibility monitoring methodology. The recommendations are based on the data gathered, analysed and validated, taking also into account policy, technical and financial implications.

The study covers three generic types of monitoring measures as defined by the European Commission for the purposes of the current study, namely:

- Initiatives taken in order to verify compliance with national provisions in the field of web-accessibility, such as regulation or guidelines (Type I);
- Initiatives taken under the initiative of public administrations even if there are no legal obligations in the field (Type II);
- Initiatives taken under the initiative of non-governmental actors including organisations of persons with disabilities and private sector companies (Type III).

These definitions do not include existing website inspection schemes where an accessibility related assessment is available to individual website owner organisations on request. Such schemes have emerged in some countries, for example, for certification purposes, that in many cases are developed for commercial reasons and managed by private companies. Other than those web accessibility monitoring initiatives falling under one of the three definitions provided above, they are not directed towards assessing a sample of websites at a certain point in time in order to understand levels of accessibility achieved across a certain geographic area or service domain.

It is important to note, that when this study was made, the Web Accessibility Directive did not include mobile applications. That is why only website interfaces are covered in this report.

**Policy context**

In December 2012, the European Commission presented a proposal for a Directive on the accessibility of public sector bodies' websites. On 3rd May 2016 negotiations were finalised between the European Parliament, the Council of the European Union and the European Commission, and on 18th July 2016 the Council formally adopted the "Directive of the European Parliament and the Council on the accessibility of the websites and mobile applications of public sector bodies".

The recommendations in Chapter 4 have been produced taking into consideration the current text of the draft Directive. However, two important issues in relation to this study within the context of the Directive are highlighted here.

Firstly, the Directive contains several references to implementing acts on the development of a monitoring and reporting methodology for Member States. In addition, several of the Directive's Articles are relevant to the aim and
recommendations contained in this report. The most relevant Articles are those on scope, presumption of conformity (standards) and monitoring and reporting. The content of these Articles and others are taken into consideration when formulating the report’s recommendations. One important development on the scope of the technologies covered by the Directive that was finalised towards the end of the negotiations was the inclusion of public sector mobile applications. Mobile applications were outside the scope of the current study and this report does not cover monitoring methodologies or make recommendations on a European-wide monitoring methodology for mobile applications.

A second important aspect of the Directive that needs highlighting at the outset are the standards referenced within Article 6 on presumption of conformity. This Article refers to the European Standard EN 301 549 version 1.1.2 “Accessibility requirements suitable for public procurement of ICT products and services in Europe”, published in 2015. Chapter 9 of this EN includes all level A and level AA success criteria from the Web Content Accessibility Guidelines 2.0 (WCAG 2.0), from the W3C. All the monitoring initiatives examined as part of this study refer to WCAG 2.0, and its Success Criteria. Therefore, for ease of readability and comparability between the different monitoring initiatives examined, this study refers to WCAG 2.0 when discussing accessibility requirements and making recommendations for a European-wide monitoring methodology. This has no material impact on the recommendations made in Chapter 4 on an EU-wide monitoring methodology for web accessibility.

Sections of the report

In Chapter 1 an overview of the whole study is presented, summarising the overall approach, scope, methodology, policy context and sections of the report.

In Chapter 2, a two-stage information gathering approach has been adopted. In a first step, an initial screening of the 30 countries to be covered by the current study was undertaken to identify any web accessibility monitoring initiatives conducted during the last five years. In the second step, this was followed by an investigation of 12 selected initiatives with a view to gaining a deeper understanding of current approaches adopted in web accessibility monitoring across Europe.

In Chapter 3, a validation exercise based on five selected monitoring methodologies has been carried out. The goal of this exercise was to obtain a picture of how well different types of methodologies actually measure web accessibility. The reliability and validity of the methodologies is a crucial part of the development of the framework. In the process it was crucial to weight quality

3 http://www.etsi.org/deliver/etsi_en/301500_301599/301549/01.01.02_60/en_301549v010102p.pdf

4 https://www.w3.org/WAI/intro/wcag
and resources needed to find the perfect balance where it is possible to assess a large number of websites in a way that is resource effective.

In Chapter 4 some recommendations on a web accessibility methodology are proposed. These recommendations are oriented to cover the key methodological aspects with reference to the European Commission’s proposal for a Directive on the accessibility of public sector bodies’ websites.
Chapter 2. Current monitoring initiatives in Europe

2.1. Cross-country overview in web accessibility monitoring

This chapter presents outcomes of the initial country screening. Overall, 43 initiatives of relevance to the current study were identified and described by national correspondents according to a common research template. Together, these cover three generic types of monitoring measures as defined by the European Commission for the purposes of the current study, namely:

1. Initiatives initiated in order to verify compliance with national provisions in the field of web-accessibility, such as regulation or guidelines (Type I);
2. Initiatives taken under the initiative of public administrations even if there are no legal obligations in the field (Type II);
3. Initiatives taken under the initiative of non-governmental actors including organisations of persons with disabilities and private sector companies (Type III).

These definitions do not include existing website inspection schemes where an accessibility related assessment is available to individual website owner organisations on request. Such schemes have emerged in some countries, for example, for certification purposes. Other than those web accessibility monitoring initiatives falling under one of the three definitions provided above, they are not directed towards assessing a sample of websites at a certain point in time in order to understand levels of accessibility achieved across a certain geographic area or service domain.

In the following subsections, outcomes are reported according to the three types of web accessibility monitoring initiatives defined by the European Commission for the purposes of the current study. Table 1 presents an overview of the number of web accessibility monitoring initiatives identified in each of the countries covered according to the three types. In line with this study’s overall objective, the analysis presented in the reminder of this chapter does, however, place the focus on a comparison of the identified initiatives by methodological approaches adopted for the purposes of web accessibility monitoring, rather than on a country-by-country comparison.

---

5 Please note that the current study focuses on web accessibility monitoring initiatives rather than on accessibility related certification schemes.

6 Examples of such initiatives include, for example, the so called BITV-Test in Germany, the AGID accessibility certification scheme in Italy and the AnySurver labelling scheme in Belgium. Information available on such schemes will be presented in a dedicated annex to the final study report.
Table 1 – Overview of web accessibility monitoring initiatives identified according to type

<table>
<thead>
<tr>
<th>Country</th>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>BE</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>BG</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>CY</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CZ</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>DK</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>EE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FI</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>FR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DE</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HU</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>IE</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LV</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>LT</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>LU</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>NL</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>NO</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>PL</td>
<td></td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>PT</td>
<td></td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>RO</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>SK</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ES</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Type I - Initiatives initiated in order to verify compliance with national provisions in the field of web-accessibility

In total, 25 web accessibility monitoring initiatives could be identified which have been initiated during the last five years to verify compliance with existing web accessibility provisions. As can be seen from Figure 1 overleaf, the overwhelming majority of these initiatives were designed as a non-regular monitoring effort, i.e. either they represent a one-off monitoring initiative (10) or repetition occurring at irregular intervals (8). Overall, annual (5) and bi-annual (1) web accessibility monitoring was less frequently applied. In one case ("other"), a tool was developed that monitors public websites on a continuous basis according to a confined number of automatically performed tests, thereby displaying statistics on outcomes on a publicly available website. Outcomes are hoped to stimulate further in-depth evaluation in cases where accessibility barriers have been detected by the automatic tool.
When it comes to different types of websites covered by the monitoring initiatives identified under this category, public websites across different governance levels were most frequently addressed by the identified monitoring initiatives, when compared with websites operated by non-governmental public services/bodies or commercial websites (Figure 2 overleaf). This result may not come as a surprise as existing legal or regulatory provisions on web accessibility in most countries concern websites and online services operated by public bodies rather than commercial parties. It is also worth noting that in a given country different types of websites may be addressed by different web accessibility monitoring initiatives, which may again vary in terms of methodological approaches adopted on a case-by-case basis.

The letter “n” represents the number of Type I initiatives identified for the purposes of this study.

When it comes to particular assessment techniques applied for the purposes of web accessibility monitoring (Figure 3) about half of the Type I initiatives relied upon a combination of automatic and manual assessment procedures (13), whereby manual end users assessment can be less frequently observed (4) when compared with manual expert assessment (9). A minority of initiatives relied upon automatic assessment techniques exclusively (4). In addition to this, other
approaches were pursued as well (6). These include, for example, surveys of website owners on whether or not a web accessibility monitoring has been conducted for their websites and if so what the outcome was. In such cases details on particular assessment techniques applied or on other methodological aspects were usually not available. Furthermore, some initiatives relied upon qualitative approaches which were not immediately directed towards quantifying the proportion of web-sites which do or do not comply with a given web accessibility standard or relevant guidelines. These focussed, for example, on assessing whether any procedures and workflows had been put in place by the website owner organisations to ensure that the design of a given website and/or content provided electronically were actually accessible to users with disabilities.

The majority (15) of Type I web accessibility monitoring initiatives relied upon WCAG 2.0, which is also ISO International Standard “ISO/IEC 40500”, as a benchmark against which selected websites were assessed (Figure 4). In some cases, (9) dedicated national standards/guidelines were applied, whereby individual assessment criteria may more or less overlap with WCAG 2.0. Other (1) benchmarks were applied for the purposes of the assessment as well, for example, procedural and workflow related guidelines/standards.

Only a minority (5) of the identified Type I monitoring initiatives relied on a self-assessment of compliance with relevant web accessibility standards/guidelines by the website owners themselves (Figure 5). In contrast, a clear majority of the initiatives (20) relied upon third-party assessment. When it comes to the amount of websites included in the assessment procedure numbers vary a lot across the individual monitoring initiatives identified, ranging from 11 to 3100 websites assessed per initiative. On average, 488 websites were assessed across all
monitoring initiatives. The majority of cases – about three-quarters of them – the number of websites assessed ranged, however, below the statistical average (Figure 6).

**Figure 5 - Overview of Type I initiatives according to parties performing the assessment**

![Figure 5](image)

n=25

**Figure 6 - Overview of Type I initiatives according to numbers of websites assessed**

![Figure 6](image)

9 For illustrative purposes, each of the Type I monitoring initiatives analysed for the purposes of this study are plotted against the number of websites assessed in the framework of the initiative under investigation. Each of the blue coloured rhombi represents one initiative.
Some level of information on outcomes was made publicly available by almost all monitoring initiatives identified (Figure 7), either online, in a printed format or both. Only in one case was it found that no information on outcomes was published. Synthesis reports (22) and selected statistics (16) represent the preferred content published by the monitoring initiatives. Raw data were made publicly available only by few initiatives (3). In one case a handbook was generated on the basis of monitoring outcomes to share experiences on how identified failures in complying with relevant web accessibility standards/guidelines can best be avoided.

Figure 7 - Overview of Type I initiatives according types of publicly available feedback

Earliest research points into the direction that website owner organisations seem to perceive a risk that country-wide or even EU-wide monitoring might turn out to be as a bureaucratic exercise rather than an operationally useful endeavour.¹⁰ In light of this, it is worth noting that few monitoring initiatives provided feedback specifically tailored to the owner organisations of the websites assessed. In some cases, detailed metrics were provided for individual websites as part of the synthesis report, for example, in terms of a dedicated data annex. In the case of a one-off monitoring initiative identified, a detailed outcome report was prepared for each website tested. The expert who had conducted the assessment was potentially available to provide up to 30 hours counselling to the website owner organisation in order to support the implementation of required improvements. This was then followed by a second audit (expert audit only). In another case a dedicated tool was made available to website owner organisations for reanalysing their website on a continuous basis.

Type II – Initiatives taken under the initiative of public administrations even if there are no legal obligations in the field

When it comes to web accessibility monitoring initiatives which have been undertaken by public administrations where no legal obligation exists (Type II), six initiatives could be identified across the 30 countries covered by the current study. As can be seen from Figure 8, most of them were conducted more than once (5), albeit not always according to regular repetition intervals (2).

Figure 8 – Overview of Type II initiatives according to repetition intervals

When it comes to different types of websites covered by the monitoring initiatives identified under this category, the majority (5) of initiatives addressed public websites across the national and/or regional governance levels (Figure 9). Websites operated by commercial parties and by non-governmental public bodies and/or local governments were addressed in one case each.
Figure 9 - Overview of Type II initiatives according to types of websites covered in relation to website ownership

- commercial
- public, non-government
- local government
- regional government
- national government

n=7

Figure 10 - Overview of Type II initiatives according to assessment technique applied

- automatic only
- manual expert only
- manual user only
- combined automatic & manual expert
- combined automatic & manual expert & end user
- other

n^{11}=6, missing^{12}=1

---

11 The abbreviation “n” stands for the number of Type II monitoring initiatives identified for the purposes of the current study.

12 The abbreviation “missing” stands for the number of initiatives for which no information was available in relation to the particular feature illustrated by the figure. This reads as follows: For one of the six Type II initiatives identified no information was available on the assessment technique applied.
When it comes to particular assessment techniques applied most initiatives relied upon a combination of automatic and manual assessment procedures (4), whereby only one initiative included manual end-users assessment (Figure 10). None of the identified initiatives relied exclusively upon automatic assessment techniques. Another approach observed (1) includes manual assessment by instructed students.

The majority of Type II web accessibility monitoring initiatives relied upon voluntary national standards/guidelines (3) or WCAG 2.0 (1) as a benchmark against which websites were assessed, whereby the former may overlap with WCAG 2.0 requirements to greater or lesser extent (Figure 11). Dedicated assessment criteria were developed in two cases, whereby it is not clear to what extent these referred to WCAG 2.0.

Figure 11 - Overview of Type II initiatives according to assessment criteria applied

![Bar chart showing assessment criteria applied]

n=6

Only a minority (2) of the identified Type II web accessibility monitoring initiatives relied on a self-assessment by the website owner organisations (Figure 12). Most initiatives (4) relied upon third-party assessment when it comes to checking compliance with relevant web accessibility standards or guidelines. The amount of websites included in the assessment procedure varied less strongly when compared with the Type I initiative under investigation. Actual numbers of websites sampled per initiative range from 11 to 307. On average, 151 websites were assessed across all Type II monitoring initiatives. In three quarters of the Type II initiatives identified, however, volumes of websites assessed do not exceed 100 sites (Figure 13).
As in the case of the Type I initiatives analysed earlier, some sort of information on outcomes was made publicly available by almost all monitoring initiatives falling under this category (Figure 14). Again, synthesis reports (4) and selected statistics

---

13 For illustrative purposes, each of the Type II monitoring initiatives analysed for the purposes of this study is plotted against the number of websites assessed in the framework of the initiative under investigation. Each of the blue coloured rhombi represents one initiative.
(4) represented the preferred content published. Raw data were made publicly available only by one initiative.

**Figure 14 - Overview of Type II initiatives according types of publicly available feedback**

![Bar chart showing the distribution of feedback types for Type II initiatives.](chart)

**Type III - Initiatives taken under the initiative of non-governmental actors including organisations of persons with disabilities and private sector's companies**

In total, 15 web accessibility monitoring initiatives were identified that were initiated by non-government organisations or commercial parties (Type III). As shown by Figure 15, almost three-quarters of these were conducted more than once (11), although repetition frequently occurred at irregular time intervals (7). One initiative (other) has been regularly conducted in conjunction with the national presidential elections.
Figure 15 – Overview of Type III initiatives according to repetition intervals

As shown by Figure 16, public websites at the regional and national governance levels (3 each) were slightly less frequently covered when compared with commercial websites (6), non-governmental public websites (5) and local government websites (4). In relation to websites or online surveys operated by businesses it is worth to be noted that particular business sectors are often covered – e.g. the largest banks or railway carriers – rather than commercial websites in general.

Figure 16 – Overview of Type III initiatives according to types of websites covered in relation to website ownership
Two in three monitoring initiatives (10) relied upon automatic testing, whereby half of these augmented automatic testing procedures with manual expert testing (5). In contrast, two initiatives relied upon manual testing only, either by experts or by users (Figure 17). Other approaches applied include manual assessment by instructed students (1) and a manual assessment jointly conducted by experts and users (1).

For the assessment procedure, more than half of the Type III initiatives that could be identified relied upon WCAG 2.0 (8) as a benchmark against which websites are assessed (Figure 18). National guidelines – either mandatory ones (3) or voluntary ones (3) – were applied in most other cases, whereby these national guidelines are mostly variations derived from WCAG 2.0. One initiative applied a set of 15 selected “basic” criteria for the purposes of the assessment.
All Type III monitoring initiatives relied upon third party assessment. The numbers of websites included in the assessment procedure range from 10 to 1000 per initiative. On average, 205 websites were assessed across all monitoring initiatives. However, in almost two-thirds of the initiatives the number of websites sampled was well below 200 (Figure 19 overleaf).
Information on outcomes was made publicly available by all monitoring initiatives identified under this category (Type III). Synthesis reports (9) and selected statistics (5) represent the preferred content published (Figure 20). Raw data was not made available by any of the Type III initiatives. In one case outcomes were published only in an article, but no dedicated report was made available.

Figure 20 - Overview of Type III initiatives according types of publicly available feedback

For illustrative purposes, each of the Type III monitoring initiatives analysed for the purposes of this study is plotted against the number of websites assessed in the framework of the initiative under investigation. Each of the blue coloured rhombi represents one initiative.
Again, few monitoring initiatives seem to have provided dedicated feedback to the owner organisations of the websites that were assessed, with a particular view to helping these in addressing potentially identified shortcomings in that respect. In one case, dedicated outcome reports were made available to website owner organisations on request. In another case, results were presented in dedicated meetings with umbrella organisations representing a particular industry sector covered by the monitoring effort in questions.

**Main trends detected in methodological approaches**

Based on the hitherto discussed evidence, Table 2 overleaf shows the “average” methodology derived from the available evidence base, in terms of the most frequently observed methodological pattern identified for each of the three generic types of web accessibility monitoring measures defined by the European Commission for the purposes of the current study:

- **Type I**: Initiatives initiated in order to verify compliance with national provisions in the field of web-accessibility, such as regulation or guidelines;
- **Type II**: Initiatives taken under the initiative of public administrations even if there are no legal obligations in the field;
- **Type III**: Initiatives taken under the initiative of non-governmental actors including organisations of persons with disabilities and private sector companies.

It is, however, worth noting that individual initiatives deviate considerably from the “average” patterns derived from the available data in relation to the three types of initiatives. All in all, a picture emerges where web accessibility monitoring, which is repeated at regular intervals, seems to have remained an exception to date. There is also some variation when it comes to types of websites covered. While Type I initiatives most frequently cover public website at the national, regional and local government levels, most Type II and Type III initiatives tend to focus on one type of websites only. WCAG 2.0 represents the most commonly applied standard, albeit national standards may to some extent deviate from WCAG 2.0 in one or another respect.

The number of websites assessed per initiative varies quite a lot within each individual type of monitoring initiatives, although the statistical average appears to be somewhat higher for Type I initiatives when compared with Type II and Type III initiatives. Across all three types, combined automatic and manual assessment is most frequently applied. Synthesis reports and selected statistics are the two outcome presentation formats most frequently observed across all three types of web accessibility monitoring initiatives. Apart from these, most initiatives do not give any specifically tailored feedback to the owners of those websites which have been assessed in the framework of the monitoring exercise.
### Table 2 – “Average” methodological approach observable according to type of initiative

<table>
<thead>
<tr>
<th>Methodological dimension</th>
<th>Type I *</th>
<th>Type II**</th>
<th>Type III***</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency of monitoring</strong></td>
<td>One-off monitoring</td>
<td>Irregular monitoring intervals</td>
<td>Irregular monitoring intervals</td>
</tr>
<tr>
<td><strong>Types of websites covered</strong></td>
<td>National government websites, regional government websites, local government websites</td>
<td>Regional government websites</td>
<td>Commercial websites</td>
</tr>
<tr>
<td><strong>E-accessibility guidelines/standards applied</strong></td>
<td>Mandatory standard / guidelines (partly based on WCAG 2.0)</td>
<td>Voluntary national standards/guidelines (partly based on WCAG 2.0)</td>
<td>WCAG 2.0</td>
</tr>
<tr>
<td><strong>Average no. of websites assessed</strong></td>
<td>488</td>
<td>151</td>
<td>205</td>
</tr>
<tr>
<td><strong>Assessment approach</strong></td>
<td>Third party assessment</td>
<td>Third party assessment</td>
<td>Third party assessment</td>
</tr>
</tbody>
</table>
   b) Automatic assessment only |
| **Means of public outcome reporting**                  | Synthesis report and selected statistics | Synthesis report and selected statistics | Synthesis report and selected statistics |
| **Dedicated feedback to website owner organisations beyond public reporting** | No | No | No |

*\(n=25\), **\(n=6\), ***\(n=15\)
2.2. Further insights from selected monitoring initiatives

2.2.1. Initial considerations

Further evidence was collated from 12 monitoring initiatives selected across Europe to enable a deeper understanding of the various monitoring approaches applied. The selection is based on a number of considerations:

- When it comes to the three generic types of initiatives defined by the European Commission for the purposes of the current study, the selection of instances for further investigation reflects the relative distribution of all initiatives identified through the initial country screening across the three generic types. Most of the instances selected for in-depth investigation belong to Type I respectively, although Type II and Type III initiatives are represented in the sample as well.

- The sample also reflects current practices in web accessibility monitoring as regards coverage of different types of websites (e.g. websites operated by the government, public websites operated by non-governmental bodies, commercial websites).

- When it comes to volumes of URLs monitored per initiative, the sample ranges from comparatively small scale initiatives assessing only few pages per website (below 10) to large scale initiatives assessing up to 6000 pages per website.

- Finally, the sample reflects different evaluation approaches (e.g. automatic testing and manual testing).

2.2.2. Web accessibility monitoring initiatives selected

Table 3 overleaf provides an overview of the 12 monitoring initiatives that were finally selected according to key features of the monitoring approach pursued in each case. As can be seen from the table, some monitoring initiatives have been conducted for some years already while others were implemented quite recently. It is worth noting that the monitoring initiatives under investigation are not always explicitly referenced in web accessibility related legislation or policy documents. There are also initiatives which seem to have emerged without any formalised regulatory stimulus. No evidence could be found suggesting that outcomes of web accessibility monitoring have yet been used for legally enforcing compliance with established web accessibility requirements or standards by individual website owners.

---

## Table 3 – Web accessibility monitoring initiatives selected for further investigation

<table>
<thead>
<tr>
<th>No. / Country</th>
<th>Title</th>
<th>Implementation 1st time</th>
<th>Methodological changes over time</th>
<th>Websites covered type</th>
<th>Volume16 websites / pages</th>
<th>Assessment procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BE</td>
<td>Accessibility Monitor (Toegankelijkheidsmonitor / Moniteur de l'accessibilité)</td>
<td>2007, 2009, 2010, 2011, 2012, 2013, 2015</td>
<td>type(s) of websites assessed; number of websites assessed, alignment with WCAG 2.0 requirements; other accessibility requirements</td>
<td>-public, government -public, non-government -commercial</td>
<td>- national - regional</td>
<td>209 ~6 -involves students as reviewers -no end users involved -no assistive technology utilities</td>
</tr>
<tr>
<td>3 EE</td>
<td>The conformance of public sector websites to WCAG 2.0 requirements in 2013.</td>
<td>2010, 2014, 2015</td>
<td>none</td>
<td>-public, government -national -regional -local</td>
<td>280 ≥318 -involves trained/ instructed reviewers -no end users involved -no assistive technology utilised</td>
<td></td>
</tr>
<tr>
<td>4 ES</td>
<td>Web Accessibility Observatory</td>
<td>2010, 2011, 2012, 2013, 2014, 2015</td>
<td>number of websites assessed; alignment with WCAG 2.0 requirements; software tools used for the assessment</td>
<td>-public, government -national -regional -local</td>
<td>710 17 -fully automated testing -no end users involved -no assistive technology utilised</td>
<td></td>
</tr>
</tbody>
</table>

16 The number of websites assessed and the number of webpages per website relate to the latest year of implementation.

17 No information on the average number of pages assessed per website is available. To ensure that each of the 38 stipulated success criteria is actually covered a sufficient number of webpages is to be selected per URL.

18 At minimum three pages were assessed (home page, search page, contacts). If these complied with WCAG 2.0 A level requirements additional pages were assessed.

19 Monitoring has been repeated at irregular time intervals until now, i.e. seven times for national-level websites, five times for regional-level websites and three times for local-level websites. From 2015 onwards, it is planned to conduct two monitoring cycles per year.
## Study on monitoring methodologies for web accessibility in the European Union

<table>
<thead>
<tr>
<th>No. / Country</th>
<th>Title</th>
<th>Implementation</th>
<th>Methodological changes over time</th>
<th>Websites covered type level</th>
<th>Websites covered</th>
<th>Volume</th>
<th>Assessment procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 FR</td>
<td>Ce que les sites Web publics nous disent de leur accessibilité</td>
<td>2014 none</td>
<td>none</td>
<td>-public, government</td>
<td>602 1&lt;sup&gt;20&lt;/sup&gt;</td>
<td>602</td>
<td>- involves trained/ instructed reviewers - no end users involved - no assistive technology utilised</td>
</tr>
<tr>
<td>6 NL</td>
<td>Simply Accessible (Gewoon toegankelijk)</td>
<td>2014 2015&lt;sup&gt;21&lt;/sup&gt; none</td>
<td>-public, government - national</td>
<td>611 6000&lt;sup&gt;22&lt;/sup&gt;</td>
<td>6000</td>
<td>- fully automated testing - involved crowd-sourced reviewers - involves trained reviewers - end user involvement possible - no assistive technology utilised</td>
<td></td>
</tr>
<tr>
<td>7 NO</td>
<td>State of Affairs in the Kingdom (Status i riket)</td>
<td>2014&lt;sup&gt;23&lt;/sup&gt; none</td>
<td>none</td>
<td>-public, government - public, non-government - commercial websites - national - regional - local</td>
<td>304</td>
<td>no data</td>
<td>- involves trained/ instructed reviewers - no end users involved - no assistive technology utilised</td>
</tr>
</tbody>
</table>

---

<sup>20</sup> As a first step, 602 websites were reviewed for a statement of a dedicated accessibility policy. In a second step, those 22 websites that indeed displayed such a statement were further assessed.

<sup>21</sup> The basis of Gewoon Toegankelijk is a continuous automatic monitoring approach, repeating tests of Dutch government at regular intervals (3 months).

<sup>22</sup> All available URLs within the selected domains are scraped and stored by crawler facilities, thereby following common crawling ethics. If a government organisation excludes a crawler from crawling its domain than that domain is excluded from testing. Currently 20 domains are excluded. Up to 6,000 URLs are randomly selected per website. Currently, 98% of the selected websites are tested fully because they have less than 6,000 pages.

<sup>23</sup> The monitoring is divided into two subsequent stages. The first stage focuses gathering data from a large number of websites at once. This is followed by a second stage including subsequent in-depth monitoring of individual websites and publishing of detailed results respectively. At the time this report was compiled the in-depth monitoring had been completed for two websites.
<table>
<thead>
<tr>
<th>No. / Country</th>
<th>Title</th>
<th>Implementation</th>
<th>Methodological changes over time</th>
<th>Websites covered</th>
<th>Volume</th>
<th>Assessment procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Accessibility of Higher Education (dostępności stron uczelni wyższych)</td>
<td>2011, 2012, 2013, 2014</td>
<td>number of websites assessed; software tools used for the assessment; presentation of results</td>
<td>-public, non-government -national</td>
<td>106 50</td>
<td>- fully automated testing - no end-users involved - no assistive technology utilised</td>
</tr>
<tr>
<td>9</td>
<td>Acessibilidade dos Conteúdos Web na Administração Pública Central (Web Content Accessibility on the Central Public Administration)</td>
<td>2008, 2010, 2011, 2012, 2013, 2014</td>
<td>number of websites assessed; no. of pages assessed per website, alignment with WCAG 2.0 requirements; verification method applied; software tools used for the assessment</td>
<td>-public, government -national</td>
<td>201 ~6014</td>
<td>- involves trained/ instructed reviewers - no end-users involved - assistive technology utilised</td>
</tr>
<tr>
<td>10</td>
<td>What is the State of Affairs? (Hur är läget?)</td>
<td>2012, 2015</td>
<td>none</td>
<td>-public, government -national -local</td>
<td>20125 ~6326</td>
<td>- self assessment by website owners - no end-users involved - no assistive technology utilised</td>
</tr>
<tr>
<td>11</td>
<td>Accessibility of webpages in Slovak republic</td>
<td>2005, 2007, 2008, 2009, 2010, 2011, 2013,</td>
<td>type(s) of websites assessed; number of websites assessed; no. of pages assessed per website, verification method applied</td>
<td>-public, government -national -regional</td>
<td>150 no data</td>
<td>- involves trained/ instructed reviewers - no end-users involved - no assistive technology utilised</td>
</tr>
</tbody>
</table>

24 The methodology applied did not rely on a fixed number of pages per websites. To ensure that pre-defined checklist is applied a sufficient number of pages was selected per website.
25 Public sector bodies were surveyed (480 central government bodies and 290 local governments). The response rate was below 100% whereby the quality of the answers received varied. In fact, a certain proportion of the responses received could not be considered for monitoring purposes.
26 The methodology applied did not rely on a fixed number of pages per websites. To ensure that the pre-defined checklist is applied, a sufficient number of pages was selected per website.
27 Changes to the current verification method are at the planning stage at this moment.
The Government Digital Service (GDS) has been compiling annual progress reports on open central government websites for more than five years. Since 2012, web accessibility, based on WCAG 2.0 has been included as a dedicated theme in the overall monitoring approach.
2.2.3. In-depth analysis of the monitoring initiatives

Over time, some of the investigated initiatives have undergone certain changes concerning particular methodological aspects, whereby the emergence of WCAG 2.0 seems to have acted as a stimulus in this respect. The particular methodological approaches adopted for the purposes of web accessibility monitoring vary in different respects across the analysed initiatives. In general, assessment of a sample of websites by trained or instructed experts represents a frequently adopted approach, whereas end users do not seem to be directly involved in current monitoring practices.

A slightly different approach has, however, been adopted by a monitoring initiative identified in Belgium (Toegankelijkheidsmonitor). University students from particular educational fields (for example, web design, ICT engineering) are recruited for reviewing a sample of websites. Each student receives a two-hour training on web accessibility. This is followed by a multiple choice test to facilitate coherent interpretation of web accessibility related requirements across all students. Each student also receives written instructions for utilisation during the assessment of a particular website assigned to him/her, including a selection of examples of good and bad practice. An average of 75 minutes per student are spent assessing a particular website. In the case where a particular check point is not successfully passed by the website under investigation, the student has to provide a written reasoning. Each website is evaluated by several students. To arrive at a consolidated result, the following rule is applied: if three out of five of the students evaluating a given website assess the website as failing a particular criterion, then the website is reported to fail that criterion. To date, more than 1400 students have been involved in the monitoring initiative through this approach.

Across the selected monitoring initiatives, diverse software tools are used by trained reviewers to support their assessment such as WAVE29, Accessibility Valet30 and the HTML code validator developed by W3C, the World Wide Web Consortium31. However, utilisation of assistive technology in particular seems to have remained an exception until now. Only in the framework of the monitoring initiative selected from Portugal (Web Content Accessibility in the Central Public Administration) screen reader software is applied, in particular JAWS and VoiceOver. There is also no evidence that more than one type of end user device has yet been utilised for accessibility testing purposes by the selected monitoring initiatives.

29 http://wave.webaim.org

30 http://valet.webthing.com/access/url.html

31 http://validator.w3.org
A fully automated assessment procedure has been adopted by three of the selected monitoring initiatives, namely in Poland (Study on the Accessibility of Websites of Universities), Spain (Web Accessibility Observatory) and The Netherlands (Accessibility Monitor). The approach pursued by the Polish initiative relies upon utilisation of an automatic validation software tool developed by a social enterprise, which also offers web accessibility related services on a commercial basis.\(^{32}\) Up to 100 sub-pages per selected website are tested with the help of this tool. The tool enables an automatic ranking of websites according to achieved levels of accessibility as well as dedicated reporting of the particular accessibility barriers identified at a given website. The Spanish Web Accessibility Monitor relies upon a dedicated open source tool developed by Information and Communications Technology Unit of the Central Government (DTIC) back in 2010. The development of this tool enabled for the first time an automatic assessment of a given website not only against those requirements which had traditionally been tested automatically but also against some requirements which had usually been assessed manually. All in all, 16 sub-pages from up to four levels of the navigation structure of a given website are automatically selected in addition to the main page. The tool also enables automated outcome reporting.\(^{33}\)

A similar approach has been adopted by the so-called Simply Accessible initiative in the Netherlands, albeit with the focus on assessing considerably larger volumes of sub-pages per website. A dedicated software tool was developed for this purpose by the ICTU Foundation, based on various open source components. The web domains of the organisations/bodies to be included in the monitoring exercise are automatically extracted from the Dutch Public Almanak and the State Web Register. A sample of websites is being tested on a continuous basis (at trimonthly intervals). Up to 6000 sub-pages are crawled according to international crawling ethics. Aggregated outcome data is made publicly available through the web.\(^{34}\) Detailed data is only available to website owner organisations following login. While this initiative focuses on fully automated monitoring of large volumes of webpages and automatic reporting of outcomes, interfacing with external assessment procedures is enabled\(^{35}\), including:

a) Crowd source testing: results of so called “cantTell” criteria can be presented in terms of easy to understand questions to online panels which will provide a final answer.

---

\(^{32}\) http://www.utilitia.com/

\(^{33}\) http://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Accesibilidad/pae_Observatorio_de_Accesibilidad.html

\(^{34}\) http://www.gewoontoegankelijk.nl

\(^{35}\) There are success criteria which cannot be fully tested automatically which will result in so-called “cantTell” responses. Also, certain criteria remain untested since machines are not yet capable of testing these.
b) Manual inspection: qualified inspection bodies can be provided with raw data and focus further testing on particular criteria which cannot be tested automatically.

Further to this, a user testing procedure has been recently enabled in cooperation with a spin-off of the Belgium University of Leuven. This procedure enables asking panels of users with disabilities whether they encounter any problems with particular features, content and functionalities of a given website. Current experience with the monitoring and the WCAG 2.0 guidelines suggest that not all reported failures in websites were indeed experienced by people with disabilities.

In contrast to the hitherto described monitoring initiatives, a self-reporting approach has been adopted for the purposes of a Swedish monitoring initiative entitled “What is the State of Affairs?” Here, 670 government bodies are surveyed in relation to the accessibility of the websites they operate. However, responses seem to have varied quite a lot in terms of quality.

When it comes to the particular accessibility requirements considered in the framework of the selected monitoring initiatives, a fair amount of diversity can be observed as well. Table 4 overleaf maps the requirements addressed in the framework of the selected monitoring initiatives against WCAG 2.0 AA requirements. It can be seen from the table that only few initiatives cover almost all requirements listed in the table. There is also diversity across the selected monitoring initiatives when it comes to how exactly a given website was assessed against individual requirements as discussed above. The validation exercise conducted in relation to the particular assessment methodologies employed by five of the selected monitoring initiatives sheds further light on this issue (see section on the validation exercise and specific validation reports).

36 http://www.w3.org/TR/WCAG20/
Table 4 - WCAG 2.0 requirements applied during the assessment procedure

<table>
<thead>
<tr>
<th>Requirement</th>
<th>BE</th>
<th>DK</th>
<th>EE</th>
<th>ES</th>
<th>FR</th>
<th>NL</th>
<th>NO</th>
<th>PT</th>
<th>PL</th>
<th>SE</th>
<th>SK</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1.1 Text Alternatives</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.1a Pre-recorded audio-only</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.1b Pre-recorded video-only</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.2 Captions (Pre-recorded)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td>n.a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.3 Audio Description or Media Alternative (Pre-recorded)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.4 Captions (Live)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2.5 Audio Description (Pre-recorded)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.2 Time-based Media</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3 Info and Relationship(s)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3.2 Meaningful Sequence</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.3.3 Sensory Characteristics</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1.3 Adaptable</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.1 Use of Colour</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.2 Audio Control</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4.3 Contrast (Minimum)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>n.a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 4 - WCAG 2.0 requirements applied during the assessment procedure (continuation)

<table>
<thead>
<tr>
<th>Requirement</th>
<th>BE</th>
<th>DK</th>
<th>EE</th>
<th>ES</th>
<th>FR</th>
<th>NL</th>
<th>NO</th>
<th>PT</th>
<th>PL</th>
<th>SE</th>
<th>SK</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2.1 Keyboard Accessibility</strong></td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
</tr>
<tr>
<td>2.1.1 Keyboard</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>✔️</td>
<td>n.a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1.2 No Keyboard Trap</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2.2 Enough Time</strong></td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
</tr>
<tr>
<td>2.2.1 Timing Adjustable</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.2a Pause, Stop, Hide (moving, blinking, scrolling)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.2.2b Pause, Stop, Hide (auto-updating)</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>n.a.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2.3 Seizures</strong></td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
</tr>
<tr>
<td>2.3.1 Three Flashes or Below Threshold</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>✔️</td>
<td>n.a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>2.4 Navigability</strong></td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
<td>![Accessibility Icon]</td>
</tr>
<tr>
<td>2.4.1 Bypass Blocks</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>✔️</td>
<td>n.a.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4.2 Page Titled</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>n.a.</td>
<td>✔️</td>
<td>n.a.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 2.4.3 Focus Order

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 2.4.4 Link Purpose (In Context)

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 2.4.5 Multiple Ways

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 2.4.6 Headings and Labels

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 2.4.7 Focus Visible

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 2.4.8 Multiple Ways

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 3.1 Readable

#### 3.1.1 Language of Page

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

#### 3.1.2 Language of Parts

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 3.2 Predictable

#### 3.2.1 On Focus

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

#### 3.2.2 On Input

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

#### 3.2.3 Consistent Navigation

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

**Assessment procedure involves trained/instructed reviewers and/or end user**

**Automated assessment procedure without involvement of trained/instructed reviewers and/or end users**

**Survey of website owner organisations (self-assessment)**

**Table 4 - WCAG 2.0 requirements applied during the assessment procedure (continuation)**

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Bulgaria (BG)</th>
<th>Czech Republic (CZ)</th>
<th>Denmark (DK)</th>
<th>Estonia (EE)</th>
<th>Spain (ES)</th>
<th>France (FR)</th>
<th>Ireland (IE)</th>
<th>Netherlands (NL)</th>
<th>Norway (NO)</th>
<th>Portugal (PT)</th>
<th>Slovakia (SK)</th>
<th>Sweden (SE)</th>
<th>United Kingdom (UK)</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

---

June 2016 - 39
### 3.2.4 Consistent Identification

| 3.2.4 Consistent Identification | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

### 3.3 Input Assistance

#### 3.3.1 Error Identification

| 3.3.1 Error Identification | ✓ | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

#### 3.3.2 Labels or Instructions

| 3.3.2 Labels or Instructions | ✓ | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

#### 3.3.3 Error Suggestion

| 3.3.3 Error Suggestion | ✓ | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

#### 3.3.4 Error Prevention (Legal, Financial, Data)

| 3.3.4 Error Prevention (Legal, Financial, Data) | ✓ | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

### A4.1 Compatibility

#### 4.1.1 Parsing

| 4.1.1 Parsing | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

#### 4.1.2 Name, Role, Value

| 4.1.2 Name, Role, Value | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | n.a | ✓ | n.a |

---

Assessment procedure involves trained/instructed reviewers and/or end user
Automated assessment procedure without involvement of trained/instructed reviewers and/or end users
Survey of website owner organisations (self-assessment)

### 2.3. Financial implications

The information gathering exercise conducted in the framework of the current study has shown that validated economic data is rarely available in relation to the selected monitoring initiatives. In particular, information on particular cost items such as method development, method implementation and reporting of outcomes are frequently not documented in a comprehensive manner. Cost estimates that were provided to the study team in relation to some of the investigated initiatives range from below one Euro up to almost 200 Euro per tested website. Albeit such estimates may provide at least a very rough indication of the order of magnitude at which costs may have accrued per tested website in the framework of the selected monitoring initiatives, they should be treated with some caution. In fact, a detailed analysis of the cost structure of the individual monitoring initiatives investigated in the framework of the current study was not possible on the basis of information available to the study team.

Some research on costs involved at the part of the website owners in making their websites accessible can be identified from the literature. Generally, the evidence base available published sources of information suggests that the effort required by public online services to comply with WCAG 2.0 requirements tends to vary a lot, whereby key factors seem to concern the age of the website and the particular
web technologies applied in a given instance\(^\text{37}\). The level of costs involved in developing or re-designing a given website so that it complies with WCAG 2.0 requirements thus seems to correlate not at least with the degree of sophistication of the online presence in question and the particular functionalities provided to the user (e.g. in case of a fully transactional e-government services in comparison to a public website designed for presenting static content only). When it comes to the costs involved in validating compliance of a given website with WCAG 2.0 requirements, it seems logic to assume that the required effort will also vary accordingly. Although no generalizable evidence has become available from the literature yet, such an assessment is generally supported by the outcomes of the validation exercise conducted in the framework of the current study as will be presented later in Chapter 3.

As robust evidence is neither available from the monitoring initiatives identified in the framework of the current study nor from the literature, Table 5 below presents an estimation of the average effort required for validating a given website against WCAG AA. All estimated figures were derived from experiences gained during validation exercise (Chapter 3) and earlier work of key knowledge holders within the study consortium, in particular Funka and W3C. As can be seen from, Table 5 different effort dimensions are considered which are described in more detail further below.

**Table 5 - Average effort required to validate a given website against WCAG 2.0 AA**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Average time spent per website (in hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual expert evaluation to verify compliance with WCAG 2.0 level AA according to the comprehensive WCAG-EM methodology(^\text{38})</td>
<td>40-60</td>
</tr>
<tr>
<td>Manual expert evaluation against WCAG 2.0 AA according to a set of pre-defined tests to identify common failures</td>
<td>7-9</td>
</tr>
<tr>
<td>Set up &amp; utilisation of evaluation tools by expert evaluator</td>
<td>0.5-1</td>
</tr>
<tr>
<td>Documentation of outcomes</td>
<td>0.5-2</td>
</tr>
<tr>
<td>End user involvement (using approximately 8-10)</td>
<td>30-50</td>
</tr>
</tbody>
</table>


\(^{38}\) Website Accessibility Conformance Evaluation Methodology (WCAG-EM) provides guidance on evaluating how well websites conform to WCAG 2.0. See [https://www.w3.org/TR/WCAG-EM/](https://www.w3.org/TR/WCAG-EM/) (last access 04/04/2016)
Activity | Average time spent per website (in hours)
---|---
users with different abilities) |  

Source: All estimated figures were derived from the validation exercise conducted in the framework of the current study and experiences of consortium members Funka and W3C.

**Manual expert evaluation against WCAG 2.0 AA**

In the framework of the validation exercise (Chapter 3), among other activities, a set of selected websites was assessed in relation to their compliance with WCAG AA criteria according to the Website Accessibility Conformance Evaluation Methodology (WCAG-EM) developed by the W3C consortium. This approach involved manual expert assessment supported by the utilisation of automatic tools available on the market for testing particular requirements (e.g. correct HTML coding). There are various software tools enabling evaluation of particular web accessibility requirements. However, none of the tools currently available enables the validation of all WCAG 2.0 requirements criteria as presented earlier in Table 4.

**Set up & utilisation of supportive validation tools by experts**

Potentially, a range of web accessibility evaluation software programs or online services are available to be used by human evaluators. These can be helpful in determine if web content meets particular accessibility requirements, albeit none of these enables a fully automated assessment across all WCAG AA criteria. They provide features which can assist during evaluation review. However, it needs to be borne in mind that many accessibility checks require human judgement and must be evaluated manually using different techniques. Also, evaluation tools can be prone to producing false or misleading results. For instance, there are automatic tools which indicate whether an image has an alternative text, (WCAG 2.0 AA guideline 1.1.1) but they are unable to indicate whether the alternative text correctly describes the image. Applying such tools without additional human assessment thus bears the risk of generating so called “false positives. Therefore, W3C’s Web Accessibility Initiative recommends that the results from evaluation tools should not be used to determine conformance levels unless they are operated by experienced evaluators who understand the capabilities and limitations of the tools in order to achieve accurate results. In other words, web accessibility evaluation tools are not suitable to determine the accessibility of a given website; they can only assist in doing so. For the purpose of the current study various tools were applied by the expert evaluators such as:

---

39 A list of available web accessibility evaluation tools is e.g. provided by W3C: [https://www.w3.org/WAI/ER/tools/](https://www.w3.org/WAI/ER/tools/) (last access: 04/04/2016)

40 See e.g. [https://www.w3.org/WAI/eval/selectingtools.html](https://www.w3.org/WAI/eval/selectingtools.html) (last access: 04/04/2016)

41 Ibidem.
Study on monitoring methodologies for web accessibility in the European Union

- The W3C Mark-up Validation Service (https://validator.w3.org/) was used to verify if the HTML code did follow an HTML standard correctly.
- Web Accessibility Toolbar (https://www.paciellogroup.com/resources/wat/) was used to inspect how different objects where implemented on single webpages, e.g. if something that looks like a heading is created with HTML elements h1 to h6.
- Colour Contrast Analyser (https://www.paciellogroup.com/resources/contrastanalyser/) was used to verify if text colours had enough contrast to the background.
- The “Developer Tools” in Internet Explorer (https://msdn.microsoft.com/library/hh968260(v=vs.85).aspx) was used to inspect and find specific code occurrences, e.g. if a form object is correctly associated with its label text.

Documentation of outcomes

Consistent documentation of the evaluation outcome of a given WCAG 2.0 requirement is required to enable replicability and comparability of achieved results across all websites under investigation. This is essential to ensure transparency of the evaluation process and justification for any statements to be made based on this evaluation.42

End user involvement

Pursuing a collaborative evaluation process involving the skills and perspectives of multiple evaluators can enhance the value of web accessibility monitoring when it comes to understanding how the accessibility – or non-accessibility – of a given website is experienced by the end users.43 Thus it can be helpful in better understanding accessibility issues that were identified on a given website. Different aspects deserve attention when considering the effective involvement of users in web accessibility evaluations. For instance, it needs to be ensured that diversity in disabilities will be adequately represented, that different types of assistive technology will be used, and different levels of experience with the Web will be represented as well. This will usually require a certain amount of preparatory effort being pursues in advance to the actual evaluation exercise.

As presented in Table 5, the effort required per effort dimension has been estimated in terms of person hours for the purposes of the current study. Due to

---

42 Website Accessibility Conformance Evaluation Methodology (WCAG-EM) also provides guidance on how to document evaluation outcomes. See https://www.w3.org/TR/WCAG-EM/#step5 (last access 04/04/2016)

43 See e.g. https://www.w3.org/WAI/eval/reviewteams.html (last access: 04/04/2016)
the reasons discussed in the beginning of this section, in reality the actual effort can of cause differ considerably from this estimated average figure in a given instance. Also, the actual costs accruing in a given country may vary considerably according to the hourly rates at which the required expertise will be available.

When it comes to fully automated web accessibility monitoring approaches investigated in the framework of the current study, the available evidence base on costs involved is as weekly developed as in the case of monitoring approaches involving expert evaluators. Figures that have been made available to the study team by a Dutch monitoring initiative suggest however that annual costs for maintaining a fully automated monitoring engine range at a level of 170.000 Euro, whereby 6000 web pages are automatically evaluated on a permanent basis. The caveat which needs to be made in this respect, however, concerns the fact the capabilities provided by this approach are far away from covering the full range of WCAG AA criteria. And even for those requirements that are addressed in principle it is not claimed to measure compliance but rather to identify some common problems.

Like all fully automated web accessibility monitoring procedures currently available, it is thus not comparable with monitoring approaches involving expert evaluators when it comes to the outputs ultimately generated.

2.4. Policy implications

The following is a brief summary of the aspects of the Directive most relevant in the context of this study. It focuses on scoping, standards and monitoring and reporting issues. These are not legal interpretations of the dispositions contained in the Directive but highlight those aspects of the Directive that are covered within this study.

Background to the Directive and its negotiation

In December 2012, the European Commission presented a proposal for a Directive on the accessibility of public sector bodies’ websites44, as planned for by action 64 of the Digital Agenda for Europe 2010-202045. With this proposal, the Commission aimed at improving the low levels of web accessibility of public entities (only one-third is accessible46), as well as harmonising the approach to the web accessibility guidelines (WCAG), on which all national legislations are based but with variances between Member States.


The proposal was submitted to the European Parliament and the Council of the EU. The Parliament reached a huge consensus among the different political groups (593 votes in favour, 40 negative, 13 abstentions) and set up their position in February 2014. On the other hand, Member States continued discussing the provisions of the proposal to find a common agreement while the negotiations (the so-called “trialogues”) with the other two institutions were finalised under the auspices of the Dutch Presidency early in 2016.


The Directive contains very detailed provisions on the scope of the websites (and mobile applications) to be covered. These are now considered in details.

Range of public sector bodies to be covered.

Article 3 on ‘Definitions’ refers to Article 2(1) of Directive 2014/24/EU to provide a definition for the public sector bodies covered by the Directive. Article 3 also includes associations formed by public bodies or governed by public law within the scope of the Directive.

Article 1 on “Subject matter and scope” excludes the following public sector bodies:

- Public broadcasters (Article 1(3.a));
- NGOs that do not provide services “essential to the public” or address the needs for persons with disabilities (Article 1(3.b)); and
- Schools, kindergartens and nurseries except for content relating to “essential online administrative functions” (Article 1(5)).


The Recitals to the Directive encourage Member States to extend the application of this Directive to other types of websites and mobile applications, in particular, those “used by a limited number of persons in the workplace or in education, and to maintain or introduce measures in conformity with Union law which go beyond the minimum requirements for accessibility of websites and mobile applications” or which “offer facilities and services which are open or provided to the public, including in the healthcare, childcare, social inclusion and social security areas, as well as in the transport sector and the electricity, gas, heat, water, electronic communication and postal services”. The Recitals also encourage “Union institutions” to comply with the accessibility requirements.

Range of contents covered and excluded.

Article 1(4) contains a very detailed list of content that is either completely exempted, or has a longer lead-in time for conformity i.e. content which is published before a specific amount of time after the date of entry into force of the Directive has passed.

The following contents are completely exempted:

- Live time-based media (Article 1(4.c));
- Online maps and mapping services – maps used for navigation purposes must have an accessible alternative (Article 1(4.d));
- Third-party content that is not funded, developed or under the control of the public sector body (Article 1(4.e));
- Reproductions of items in heritage collections that cannot be made accessible due to:
  - The original item does not lend itself to be made accessible – the example provided for clarification is in the case where colour contrast of the original heritage item is insufficient; or
  - The lack of available “automated and cost-efficient solutions” to render an accessible reproduction of the item (Article 1(4.f)).

Contents with longer lead-in time for conformity are:

- Websites published more than 21 months after the date of entry into force of the Directive, applicability starting 45 months after the date of entry into force of the Directive (Article 12(3.a));
- Office file formats, clarified in the recital to include formats such as Microsoft Word and Adobe PDF documents, published before 21 months after the date of entry into force of the Directive (Article 1(4.a));
- Pre-recorded time-based media published before 45 months after the date of entry into force of the Directive (Article 1(4.b));

50 Recital (34)
51 Recital (35)
• Extranets and intranets whose content is available only to a “closed group of people and not to the general public” published before **33 months** after entry into force of this Directive (Article 1(4.g)); and
• Archives – content that is “neither needed for active administration processes not updated or edited after” **33 months** after the date of entry into force of the Directive (Article 1(4.h)).

The current scope in the draft Directive of the websites and their contents does not have wide ranging implications for this study. At a practical level, the research conducted on the tests and monitoring initiatives identified in this study were more focused on issues such as the validity of the tests, ranges of sampling to provide valid and authoritative results and how the tests could be conducted. Similarly, the recommendations produced were more concerned with issues of how to monitor static versus dynamic content found on fully automated transactional websites as well as the different levels of governance of various public sector websites.

One implication of note is the complete exclusion of “live time-based media” (Article 1(4.c)) from the scope. This practically means that the recommended monitoring methodology will need to exclude from the range of tests to be developed those clauses from EN 301 549 related to live time-based media.

Therefore, this study and its recommendations in Chapter 4 remain relevant to the altered scope of the websites and contents covered by the final Directive.

**Standards**

Article 4 sets out the essential requirements for making websites and mobile applications accessible. This requires Member States to make their websites and mobile applications “perceivable, operable, understandable and robust”.

Article 6 “Presumption of conformity with the accessibility requirements” sets out the standards and technical specifications against which presumption of conformity to the accessibility requirements set out in Article 4 is to be met. It does this by providing a range of alternatives for references to these standards.

This study considers only the ones for public sector websites.

Article 6(1) sets out that, in the first instance, the accessibility requirements to be used as the basis for a presumption of conformity are those contained in a harmonised standard, reference to which is published by the Commission in the *Official Journal of the European Union*.

Article 6(3) sets out that where no reference to harmonised standards has been published, the relevant requirements in EN 301 549 are to be used.
As mentioned previously, the web accessibility requirements in EN 301 549 are identical to all level A and Level AA Success Criteria contained in WCAG 2.0. For ease of readability, and as a practical matter in comparing the current monitoring initiatives of interest, this study refers to WCAG 2.0 when referencing accessibility requirements. In this context, reference to WCAG 2.0 in this study should be taken to mean the identical set of accessibility requirements set out in the clauses of EN 301 549.

**Monitoring**

Article 8 "Monitoring and reporting" sets out that Member States shall periodically report on the compliance of websites (and mobile applications) for public sector bodies.

The Article sets out the legislative process and date by which the monitoring methodology is to be adopted as follows:

- Article 8(2) sets out that the Commission shall establish, through implementing acts, a methodology for monitoring the conformity of websites (and mobile applications)
- The methodology shall be adopted through implementing acts by 2 years after the date of entry into force or the Directive (Article 8(2)).

The Article also sets out that Member States shall designate a body that shall conduct the monitoring and that Member States shall inform the Commission who this body is 21 months after entry into force of the Directive (Article 8(7)).

The Article further states that the monitoring methodology shall be “transparent, transferable, comparable, reproducible and easy to use” (Article 8(3)) and that it may take into account expert analysis (Article 8(3.a)).

Further aspects the monitoring shall take into account in relation to websites are:

- The periodicity of the monitoring;
- Sampling of websites;
- A description of the websites compliance with the relevant standard; and
- A mechanism to provide usable information to public sector bodies on aspects of their websites that fail to comply with the standard.

The first reports on the outcomes of monitoring must be submitted by Member States to the European Commission 5 years after entry into force of the Directive. Member States must subsequently submit a report to the Commission every 3 years thereafter (Article 8(4)).

Article 8(5) sets out some specific aspects these report must include, such as descriptions of:

- How Member States have consulted with relevant stakeholders
- The procedures undertaken to make public any developments in accessibility policy;
• The experiences of Member States in implementing accessibility requirements; and
• Information on awareness raising and training.

This study is cognisant of all the aspects for the monitoring methodology covered above. It emphasises the need for the methodology to be “transparent, transferable, comparable, reproducible and easy to use”. It considers carefully the sampling sizes of cohorts of websites covered by the Directive and the time taken for and the timings of monitoring to be conducted. While this study does not prescribe what is to be contained in the report to be submitted by Member States to the Commission, the recommendations do consider issues such as how to gather and the formatting for reporting measurement data and how to provide this in a format that will be useful to public sector bodies.

**Other articles of interest**

Other articles of interest and relevance to this report include Article 5 on “Disproportionate Burden”. This study and its recommendations take careful consideration, based on the evidence available, of the costs and effort required to implement the recommended monitoring methodology.

Article 7 “Additional measures” deals in detail with the provision of an accessibility statement. While it is outside the scope of this study to provide explicit recommendations on such statements, the recommendations for a monitoring methodology and the recommendations for the production of measurement data will be useful in the formulation of such accessibility statements.
Chapter 3. Validation exercise of a set of monitoring methodologies

3.1 Purpose of the validation exercise

The overall picture that emerges from the countries covered by the current study in relation to current practices in web accessibility monitoring is one of great diversity. An “average” approach in terms of the most frequently observed methodological features can be described by combining manual expert assessment with automatic assessment procedures.

In fact, there are a number of automatic tools available to measure different web accessibility factors, although they cannot give an accurate or full picture. For example, an automatic tool can evaluate if an image has an alternative text, but it cannot tell if the alternative text correctly describes the image.

On the other hand, manual testing requires a lot of resources. There are several hundred factors that need to be tested to give a complete picture of the accessibility level of a website. It is not feasible to do a full check of all European websites manually.

A dedicated strand of work within this study therefore focused on validating commonly applied methodological features (the “average” methodology) against the Website Accessibility Conformance Evaluation Methodology (WCAG-EM) developed by the W3C consortium. The aim of this exercise was to obtain some indications on how commonly applied methodological features perform in comparison with WCAG-EM, with a view to identifying potential structural differences in the achieved outcome.

When measuring web accessibility, experience shows that a lot of manual work is needed to provide an accurate result. Although there are a number of automatic tools available to measure different web accessibility factors, they cannot give either an accurate or a full picture of compliance. Until now, no automatic tool has been able to test all criteria. And the quality of the results from automatic testing is often questionable. As per the example used previously, an automatic tool can evaluate if an image has an alternative text, (WCAG 2.0 AA guideline 1.1) but it cannot tell if the alternative text correctly describes the image. That means that the automatic testing risks showing a false positive: yes, there is an alternative text description (potentially complies with the regulation) but in reality, the alternative text is nonsense and therefore no accessibility is achieved for the end-user (does not comply with the regulation).

The goal of this exercise has been to develop a picture of how well different types of methodologies actually measure web accessibility, how reliable the methods are
and what resources are needed. All these factors are crucial to provide recommendations for European monitoring.\textsuperscript{52}

### 3.2. Selected monitoring methods

Following the in-depth analysis of the initiatives mentioned above, five initiatives were selected as follows and which we call Methods 1, 2, 3, 4 and 5:

- “State of Affairs in the Kingdom (Status i riket)”, Norway (Method 1)
- “Mapping of web-accessibility on public web-sites”, Denmark (Method 2)
- “Acessibilidade dos Conteúdos Web na Administração Pública Central (Web Content Accessibility on the Central Public Administration)”, Portugal (Method 3)
- “Simply Accessible (Gewoon toegankelijk)”, The Netherlands (Method 4)
- “Web Accessibility Observatory”, Spain (Method 5)

These methods were selected to ensure that they cover the most used approaches (i.e. manual testing, automatic testing or combinations of both) and different kind of tools and exercises. The goal was to evaluate as many different approaches as possible. The selection was also dependent on how transparent the methodologies were. In many cases it was hard or impossible to get information on exactly what had been done to end up with the results that were reported in different methodologies. For a methodology to qualify as a candidate for the practical evaluation exercise, the study group needed to obtain the complete details of the methodology to be able to use it. If parts of the methodology were secret or unknown, they could not be used in the context of this study.

### 3.3. Description of the validated “average” methodology

The following table summarises commonly applied methodological features of the five monitoring initiatives, or “Methods” identified earlier.

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Overall assessment procedure</th>
<th>Balances automatic and manual testing</th>
<th>Tools applied</th>
<th>Sampling method</th>
<th>Level of websites covered</th>
<th>Volume: Number of websites</th>
<th>Volume: Pages / website</th>
</tr>
</thead>
</table>

52 The operational conduction of the validation exercise took more time than originally expected and the current deliverable therefore presents a first description of the outcomes which will be further digested for the preparation of the final report.
<table>
<thead>
<tr>
<th>Methodology</th>
<th>Overall assessment procedure</th>
<th>Balances automatic and manual testing</th>
<th>Tools applied</th>
<th>Sampling method</th>
<th>Level of websites covered</th>
<th>Volume: Number of websites</th>
<th>Volume: Pages/website</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method 1: State of Affairs in the Kingdom (Status i riket)</strong></td>
<td>- involves trained/instructed reviewers - no end users involved - no assistive technology utilised</td>
<td>All tests are conducted by humans but with the aid of some automatic tools</td>
<td>Complete data not available, but known tools utilised are: - The W3C Markup Validation Service 53 - Web Accessibility Toolbar 54 - Colour Contrast Analyser 55</td>
<td>Pages are selected manually based on a list of types of pages to select.</td>
<td>- national - regional - local</td>
<td>304</td>
<td>No data</td>
</tr>
<tr>
<td><strong>Method 2: Mapping of web accessibility on public web sites</strong></td>
<td>- involves trained/instructed reviewers - One end user involved - no assistive technology utilised 56</td>
<td>All tests are made by humans but they use some tools to aid the work</td>
<td>No data</td>
<td>Manual, based on the WCAG-EM</td>
<td>- national - regional - local</td>
<td>204</td>
<td>No data</td>
</tr>
<tr>
<td><strong>Method 3: Acessibilidade dos Conteúdos Web na Administração Pública Central (Web Content Accessibility on)</strong></td>
<td>- involves trained/instructed reviewers - no end users involved - assistive</td>
<td>Fully automated testing but the sampling is done by a human.</td>
<td>- Opera web browser - AccessMonitor 58</td>
<td>Sampling is done by a human using the Opera web browser to gather</td>
<td>- national</td>
<td>201</td>
<td>~60</td>
</tr>
</tbody>
</table>

53 https://validator.w3.org/

54 https://www.paciellogroup.com/resources/wat/

55 https://www.paciellogroup.com/resources/contrastanalyser/

56 The assessment methodology is based on the WCAG-EM methodology. The assessment is conducted by one trained expert. In parallel to this, one user with assistive technology does a simple assessment and each website is given a general result (e.g. big problems / some problems / only minor problems).
## Methodology

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Overall assessment procedure</th>
<th>Balances automatic and manual testing</th>
<th>Tools applied</th>
<th>Sampling method</th>
<th>Level of websites covered</th>
<th>Volume: Number of websites</th>
<th>Volume: Pages/website</th>
</tr>
</thead>
<tbody>
<tr>
<td>the Central Public Administration)</td>
<td>technology utilised</td>
<td>The best and worst websites identified through the automated testing are subject to a manual review conducted by trained reviewers who also test with assistive technology</td>
<td>links from the start page (Menu &gt; Tools &gt; Links). The start page itself and pages linked to the start page are used.</td>
<td>Fully automated testing</td>
<td>- Gewoon toegankelijk(^\text{59})</td>
<td>-nationale</td>
<td>611</td>
</tr>
<tr>
<td>Method 4: Simply Accessible (Gewoon toegankelijk)</td>
<td>- fully automated testing no end users involved</td>
<td>Fully automated testing but with the possibility to make manual analyses with trained reviewers and/or user testing if a website owner orders that</td>
<td>Fully automated</td>
<td>- Gewoon toegankelijk(^\text{59})</td>
<td>-nationale</td>
<td>611</td>
<td>≤6000</td>
</tr>
<tr>
<td>Method 5: Web Accessibility Observatory</td>
<td>- fully automated testing -no end</td>
<td>Fully automated testing</td>
<td>- Self-developed automatic tool</td>
<td>Fully automated</td>
<td>-nationale -regionale</td>
<td>710</td>
<td>17</td>
</tr>
</tbody>
</table>

\(^\text{58}\) A server installation of this tool is used when monitoring. The public version of AccessMonitor is available on [http://www.acessibilidade.gov.pt/accessmonitor/](http://www.acessibilidade.gov.pt/accessmonitor/).


\(^\text{59}\) [https://www.gewoontoegankelijk.nl/en/content/monitor](https://www.gewoontoegankelijk.nl/en/content/monitor)
The 5 methods cover the requirements in WCAG to different degrees. The table below shows which WCAG 2.0 level A and AA success criteria are covered in each methodology. Some methodologies include tests measuring other accessibility related aspect, not included in WCAG 2.0 level AA. These tests were not part of the validation exercise.

---

60 [http://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Accesibilidad/pae_Observatorio_de_Accesibilidad.html#Vv4GK3Rjm00](http://administracionelectronica.gob.es/pae_Home/pae_Estrategias/pae_Accesibilidad/pae_Observatorio_de_Accesibilidad.html#Vv4GK3Rjm00)
Table 7 – WCAG 2.0 Success Criteria covered

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Theoretical number of success criteria covered (there is a total of 38 success criteria on level A and AA in WCAG 2.0)</th>
<th>Success criteria covered</th>
<th>Average number of success criteria covered in the validation exercise(^{61})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method 1: State of Affairs in the Kingdom (Status i riket)</td>
<td>35</td>
<td>All except for 1.2.3, 1.2.4, 1.2.5.</td>
<td>35(^{62})</td>
</tr>
<tr>
<td>Method 2: Mapping of web-accessibility on public web-sites</td>
<td>38</td>
<td>All</td>
<td>38</td>
</tr>
<tr>
<td>Method 3: Acessibilidade dos Conteúdos Web na Administração Pública Central (Web Content Accessibility on the Central Public Administration)(^{63})</td>
<td>21</td>
<td>1.1.1, 1.2.1, 1.3.1, 1.3.2, 1.4.3, 1.4.4, 1.4.5, 2.1.1, 2.2.1, 2.2.2, 2.4.1, 2.4.2, 2.4.4, 2.4.5, 2.4.6, 3.1.1, 3.2.1, 3.2.2, 3.3.2, 4.1.1, 4.1.2</td>
<td>11.3</td>
</tr>
<tr>
<td>Method 4: Simply Accessible (Gewoon toegankelijk)</td>
<td>17</td>
<td>1.1.1, 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.2.5, 1.3.1, 1.4.1, 1.4.2, 1.4.3, 2.4.4, 3.1.1, 3.3.1, 3.3.3, 3.3.4, 4.1.1, 4.1.2</td>
<td>8</td>
</tr>
<tr>
<td>Method 5: Web Accessibility Observatory(^{64})</td>
<td>19</td>
<td>1.1.1, 1.3.1, 1.4.3, 2.1.1, 2.2.1, 2.3.1, 2.4.1, 2.4.2, 2.4.3, 2.4.4, 2.4.5, 2.4.7, 3.1.1, 3.1.2, 3.2.1, 3.2.2, 3.2.3, 4.1.1, 4.1.2</td>
<td>19</td>
</tr>
</tbody>
</table>

\(^{61}\) The fact that a success criteria is covered by the methodology doesn’t mean that this success criteria is fully covered, it only means that the method do assess one aspect that is relevant to this success criteria in WCAG.

\(^{62}\) The validation exercise did not complete the tests on all 35 success criteria covered because of time limit.


This methodology include some tests that are relevant to success criterias on level AAA in WCAG, these references is not included in the table.

\(^{64}\) This methodology include one test that is outside WCAG. This test was not included in the evaluation of the methodology.
These five methodologies were selected to ensure that they cover the most frequently applied methodological approaches (i.e. manual testing, automatic testing or combinations) and different kinds of tools and exercises. In practical terms, the selection was also dependent on how transparent the methodologies were. In many cases it was hard or impossible to get information on exactly what had been done to end up with the results that were reported in different methodologies. For a methodology to qualify as a candidate for the practical evaluation exercise, the project needed to obtain the complete methodology to be able to use it. If parts of the methodology were secret or unknown, they could not be used in the context of this study.

3.4. Design of the validation exercise

One way to test how well different methodologies measure accessibility is to actually use them on a number of websites. By measuring the same websites with different methods and comparing the results, it is possible to uncover any variations between the different methodologies (scope).

Another important aspect is to let several web accessibility experts use the same methodology on the same website. This empirical approach illustrates how likely it is that different experts will achieve the same results using one specific method (reliability).

A third important aspect is to compare the results against a “key” to determine whether the method measures accessibility in an accurate way (validity). To achieve this baseline, it is essential to measure the websites in the most authoritative way possible. This has been ensured by using the well-known W3C evaluation methodology.

Finally, all the above testing will in turn be measured and documented in terms of how much time is needed to do the measuring of a number of websites using different methods. This part is needed to provide the cost and efficiency analysis that will become an important part of the recommendations from this study. The above results in a number of dimensions would be as follows: $N_1 \text{ methods} \times N_2 \text{ websites} \times N_3 \text{ experts} \times N_4 \text{ thorough analysis}$.

To perform this level of validations of methods even on a few websites using only a few methods would require months of work and a whole team of experts. Although this would result in a lot of interesting data to work with, it would be neither cost effective nor feasible to accomplish within the framework of this study. Furthermore, it is not necessary in order to propose a framework for a methodology.
The validation exercise has therefore been designed to obtain a reasonably accurate picture of as many dimensions as possible. The number of methods used was limited to the five most interesting ones found during the collection of methodologies from the Member States. Five methods suffice to cover different types of methods and give a broad perspective. Comparing different but similar methods would focus on small details but since the goal of this project is not to give a detailed description of a finalised method, but to develop recommendations for a framework of a method, the validation exercise has focused on comparing and validating different types of methods.

The number of websites used for the validation exercise was initially limited to five. According to the expertise of Funka and W3C, five websites include a sufficient variety of different types of websites to qualitatively assess how reliable the methods are when applied to different contexts. It is important to use the same cohort of websites when testing all methods. This ensures that the study team had sufficient time to evaluate the accessibility levels of each website. It also makes it possible to compare results, both between the different methods and also between each method and the baseline established using the W3C evaluation methodology.

For those methods that contain an element of manual assessment, two consultants independently tested each method on all five websites. This results in ten reviews for each method. This is sufficient to facilitate a comparison of the concurrence in the results when performed by different experts. It is not sufficient to provide a detailed picture of the exact reliability of each test performed, but it can show if there are large or small variations. In the end, this is also a matter of detail in the final method. The idea was oriented to analyse if one method tends to give many or larger variations while another one tends to give fewer, or smaller variations.

3.5. Analysis of the extent of methods

An important aspect in the evaluation of different methods is the extent to which they cover WCAG 2.0. A method that only covers a few criteria of WCAG 2.0 will, of course, not measure as much accessibility as a method that covers more of them. In the description of the tests and the range of each method it has been documented how many tests have been executed and how well they meet the criteria.

It is important to understand that the methods cover the success criteria in WCAG 2.0 in different ways. It is possible to measure one aspect of all criteria, or ten aspects of some criteria. This means that one method can provide a complete picture of one criterion, while another provides a few aspects of many criteria. In theory there is nearly an infinite number of aspects per criterion possible to be measured.
This fact has to be taken into consideration in the analysis of the results. However, it is not something that can be included in the validation exercise of this study. To make that kind of analysis, a much larger project would be required, since it would be needed to apply the complete methods on a larger number of websites and to compare the results with a complete key (meaning one would have to review all pages of the selected websites in order to find all accessibility issues). This is not possible within the framework of this study. Furthermore, it is not necessary to be able to analyse the important factors to be compared in each method (reliability, validity and actual accessibility).
Chapter 4. Recommendations for the development of a European web accessibility monitoring methodology

4.1. Introduction

This study aims to make recommendation on a web accessibility methodology with reference to the European Commission’s proposal for a Directive on the accessibility of public sector bodies’ websites. The finalised Directive contains a number of essential methodological aspects, including:

a. the number of websites to be monitored in each Member State;
b. the frequency these are to be assessed for compliance with WCAG 2.0 AA requirements;
c. the type of data to be generated by each Member State according to a common assessment procedure;
d. the format in which the data generated by the Member States are to be reported to the Commission.

In the following sections, the available evidence is discussed according to these aspects and methodological recommendations made respectively. Economic aspects related to methodological considerations are also discussed.

As previously stated in this report, there is no commonly accepted definition of what comprises a web accessibility monitoring methodology. For the purpose of the current study, web accessibility monitoring is understood as a dedicated measure directed towards understanding to what extent the totality of websites in a given domain or geographic region - or a sample thereof - complies with defined accessibility requirements. This must not be confused with checking the actual degree of accessibility for persons with disabilities of specific websites, since any given method can only measure parts of the accessibility requirements needed to meet all end user needs. The aim is to provide a set of feasible recommendations on how to monitor web accessibility in a meaningful and reasonable way. By monitoring a reasonable number of websites and using a flexible method, which takes into account the fast evolving technological environment and thus leaves room for development, the monitoring will continue to be relevant. By combining the manual expert monitoring with self-declaration, training and end user involvement, a high level of quality and credibility can be ensured, as well as providing a common understanding of web accessibility. By constantly developing the tests and changing the cut-off dates, the monitoring will serve as a driving force to accomplish web accessibility.
Following the overall approach of the study to gather detailed evidence of existing monitoring initiatives in Europe, and based on the results of the validation exercise which compared 5 methods currently being used in Europe with the baseline WCAG-EM method, we have concluded that none of the existing methods can be considered complete enough and suitable for a European-wide monitoring scheme.

Furthermore, none of the methods have been formally vetted by a standardisation body to ensure correct interpretation of the standard, nor have they been subject to a consensus process involving a broad participation of stakeholders from industry, disability, organizations and research.

Nevertheless, based on the evaluation of the methods carried out, we conclude that Method 1 evaluated above and used by Difi in Norway is the most comprehensive in terms of quantity and quality.\(^\text{65}\) Therefore, the study team’s recommendation is to adopt the Norwegian method\(^\text{66}\) as main reference method for a European wide monitoring methodology in the first instance, until a formally vetted and recognised methodology becomes available. Nevertheless, it must be taken in consideration that Method 1 currently contains no tests to validate three of the WCAG 2.0 level AA success criteria, namely success criteria 1.2.3, 1.2.4 and 1.2.5. Therefore, it would be need to be completed with tests procedures for these success criteria. Moreover, regarding the testing approach, Method 1 would need to be constantly developed, and the combination of existing testing approaches with self-declaration and the involvement of end users is also suggested.

### 4.2. Overall discussion and evidence-based recommendations

#### 4.2.1. Number of websites to be assessed

The number of websites to be assessed refers to how many websites are to be monitored at any given time in a given Member State.

**Discussion of the available evidence base**

Currently existing monitoring initiatives vary a lot when it comes to the amount of websites assessed, ranging from 11 to 3100. In fact, different aspects have an impact on the number of websites included in an established assessment

\(^{65}\) This method tests for the highest number of WCAG 2.0 requirements/success criteria out of the methods examined, as well as providing the most accurate test results.

\(^{66}\) See Annex 2. Description of the Norwegian methodology “Status i riket”
procedure. While there are, for instance, some monitoring initiatives which aim to generate a rather coarse snapshot of the accessibility of a website, others are directed towards enabling a more detailed understanding of the quality and/or quantity of accessibility barriers experienced by people with disabilities, sometimes with a view to actively supporting website owners to eliminate current barriers. For generating a high-level picture of the accessibility situation prevailing in a given country or service domain, a small number of strategically selected websites may suffice, when compared with a monitoring strategy that aims at enabling a more comprehensive understanding of the accessibility barriers experienced by end-users. To enable a more detailed understanding of the quality and/or quantity of accessibility barriers prevailing in a given country or service domain, a reasonably representative sample of websites should to be assessed.

Another factor typically impacting on the number of websites assessed in the framework of a given web accessibility monitoring methodology concerns the approach adopted on assessing individual websites. Here different approaches can be observed. For instance, there are many monitoring initiatives relying upon manual tests performed by accessibility experts or people who have received more or less intensive accessibility related training. Frequently, these utilise different types supportive tools which have become available over recent years for validating particular requirements stipulated by WCAG 2.0, e.g. proper HTML coding. Such tools are, however, not available for testing every single WCAG 2.0 requirement, and there are many requirements which have to be assessed by the evaluators without having any tool available for this purpose.

In contrast, there are some monitoring initiatives relying on fully automated testing procedures. In such cases typically, a dedicated software solution has been programmed to validate a selection of WCAG requirements without involving any human evaluator. These tests are therefore not identical with supportive tools typically used by expert evaluators. Fully automated testing solutions enable a larger number of websites to be tested in a timeframe that is comparatively shot when compared with the time required to carry out tests based on manual testing. However, none of the solutions currently available is capable of validating the full range of requirements stipulated by WCAG. Although they enable validating a certain proportion of WCAG requirements in a comparatively short time span, there is a certain risk of creating so-called “false positives”. For example, as previously explained, a fully automated monitoring solution can determine if an image embedded in a webpage has an alternative text, (WCAG 2.0 AA guideline 1.1) but it cannot evaluate whether or not the alternative text identified actually describes the image in question in a way that makes it accessible to the end user.

---

67 A selection of evaluation tools is available at: https://www.w3.org/WAI/ER/tools/ (last access 18.02.2016)
In summary, it can be concluded that the current web accessibility monitoring approaches tend to be directed towards balancing “breadth” versus “depth” when it comes to determining how many websites are to be assessed per round of monitoring. While fully automated assessment procedures tend to come with restrictions in relation to the level of depth at which WCAG 2.0 requirements are validated, as well as how many WCAG requirements are possible to test, manual procedures are typically restricted when it comes to the number of websites included in the assessment procedure as they tend to be rather time consuming. There is thus no commonly accepted “gold standard”, and the number of websites included in current monitoring methodologies very much depends on the strategy adopted by a given monitoring initiative towards balancing restrictions associated with the different methodological approaches respectively.

The Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0 was therefore used in this study to establish a baseline against which different methodologies were compared. WCAG-EM provides guidance on good practice in inspecting a single website and is intended for use by people experienced in evaluating accessibility using WCAG 2.0 and its supporting resources.\(^\text{68}\)

Outcomes of the validation exercise suggest that, while fully automated testing methodologies can shed light on compliance with a number of selected WCAG 2.0 requirements in a reliable manner\(^\text{69}\), they are inadequate for testing of a wide range or all success criteria from WCAG 2.0. Fully automated testing procedures may thus provide an advantage when it comes to monitoring high-level trends in web accessibility - with a particular focus on a limited number of indicators suitable for being measured automatically - due to the large amount of websites that can be assessed in a comparatively short time span or at comparatively short time intervals. In contrast, manual testing procedures - supported by available evaluation tools – tend to provide an advantage when it comes to validating a broader range of WCAG 2.0 requirements. However, very few of the manual testing methodologies included in this study’s validation exercise were designed to check every one of the success criteria from WCAG 2.0.

The number of websites typically assessed per monitoring initiative ranged from 11 websites for manual assessments up to 3100 websites for fully automated procedures. When looking across all countries with existing web accessibility regulation where monitoring initiatives have been launched – independently of whether the assessment procedure includes manual or automatic testing - an

\(^{68}\) [https://www.w3.org/TR/WCAG-EM/](https://www.w3.org/TR/WCAG-EM/)

\(^{69}\) Albeit bearing a certain risk of delivering “false positives” in relation to some WCAG 2.0 requirements as discussed earlier.
average of almost 500 websites were assessed per monitoring round (see section 2.1).

Another important aspect that has an impact on the time and effort required to assess a website against WCAG 2.0 requirements - and thus on determining the number of websites that can be assessed within the time/resources boundaries of a given monitoring initiative - concerns the approach adopted towards operationally defining what the unit of investigation should be, in other words of what a website may practically comprise of. In fact, the Web is anything else but a static technology space. Rather, Web technologies and services delivered through the Web continually change and evolve. How users of the Web experience content via connected devices – laptops, phones, tablets, wearables – also continues to undergo a dramatic change.

Across the EU, in there are different levels of complexity in relation to how online public services communicate between with citizens. These range from relatively straightforward information provision to dedicated online transaction services between the citizen and the public body.\(^7\) Beyond this, public administrations are beginning to pursue new concepts of service integration which involve the bundling of services across several administrations in order to match the service offering to citizens more closely with typical life events experienced by the citizen.\(^8\) In practice, a public organisation's ‘website’ may thus represent an entry point to a more complex process of online service delivery requiring the user to interact with different types of interfaces and perform diverse online tasks until a desired transaction can be successfully completed. There may even be instances where service integration has reached a stage where a user may need to interact with more than one website to complete a desired transactional process.

From a European perspective, the situation is particularly complex. While in some countries fully automated online services addressing particular life events (e.g. registering a car) are available to the citizens, only parts of the process can be handled online in other countries. In some countries merely service-related information may be available on a given website while the actual service process may need to be completed off-line through more traditional service channels.

All in all, this suggests that the operational definition of what should be considered a ‘website’ for the purposes of a common web accessibility monitoring methodology needs to reflect different levels of online service complexity ranging from mere provision of static information to fully automated and interactive online

\(^7\) CEC (2014): Delivering the European Advantage? ‘How European governments can and should benefit from innovative public services’. Background Report

\(^8\) Ibidem.
transactional services. Current web accessibility monitoring initiatives seem, however, not specifically geared towards assessing entire online transactional processes in relation to their accessibility for people with disabilities. Rather they tend to focus on assessing selected interactive elements provided at a given website through expert testing, if at all. Furthermore, systematic end-user involvement, for example, in terms of accessibility related ‘mystery shopping’ approach tailored towards dedicated life events (e.g. personal tax declaration), does not yet seem to have been carried out.

In contrast, this approach has recently been successfully applied for the purposes of user-centred e-government benchmarking. This approach did not, however, include a focus on web accessibility, despite the user validation of the present study showing that end users can highlight key quality aspects of a given website which were ignored by the websites owners. This user-centric methodological approach relied on a systematic collaboration of contracted consultancies with national representatives and public bodies providing their services online. This helped ensure, for example, that when deciding how a particular government service can be defined for the purposes of the testing, that national peculiarities were taken into account during the testing and reflected in the subsequent reporting.

It could be argued that omitting fully interactive transactional processes from a framework for web accessibility monitoring might ultimately disincentivize public services from taking a dedicated effort towards ensuring their accessibility. However common sampling criteria are needed to ensure the reliability, comparability and transparency of the framework for web accessibility monitoring across different countries and over time. There is also a need for some sort of a fall back system if the criteria fail for one reason or another. The sampling procedure therefore needs to be sufficiently flexible to work well in different national contexts. This requires in-depth knowledge of how services in different public sectors are organised in different Member States.

**Recommendations**

**Assess a representative sample of websites in each Member State**

The cohort of websites that come within the scope of the Directive can be expected to vary a lot across individual Member States, in terms of both their number as well as the type and design of online service processes that can indeed be accessed by the citizens through a given URL. Each Member State should clearly define the

---

cohort of websites to be monitored, based on the definition what a website entails provided by WCAG-EM. A representative sample should be drawn up in relation to the cohort defined. From a statistical point of view, a sample size comprising of on average 250 websites of the defined cohort of websites would enable achieving a reasonable level of significance.

Define the unit of investigation that enable the web accessibility assessment to be performed across different levels of online services maturity ranging from static websites to fully transactional public services

An operational definition for the basic unit of investigation should be specified for the websites to be sampled in each Member State, thereby considering essential functionalities as set out in WCAG-EM. Such a definition should enable coverage of different types of online content and interaction modes, ranging from rather simple websites designed for the purposes of mere information provision to fully automated transactional services.

4.2.2. Frequency of assessment

The frequency of assessment refers to the specific repetition intervals according to which the monitoring exercise will be carried out.

Requirements under the Directive

The Directive requires that Member States shall “periodically monitor” the compliance of websites (and mobile applications) of public sector bodies. Article 8.3(a) of the Directive states that the monitoring methodology shall include the periodicity of the monitoring.

Article 8.4 also requires that outcomes of the monitoring exercises including the measurement data shall be submitted to the European Commission “five years after the date of entry into force of the Directive” and “every three years thereafter.”

Please note this distinction between the periodicity of the monitoring exercise, which is to be defined by the monitoring methodology to be developed under implementing acts and the periodicity of the reporting of the ‘measurement data’ produced based on these exercises. The following discussion focuses on the periodicity of the monitoring exercise.

Discussion of the available evidence base

The overwhelming majority of currently existing monitoring exercises occur on a non-regular basis, i.e. either they represent a once-off monitoring activity or are repeated at irregular intervals. In one case, for example, an automated monitoring

https://www.w3.org/TR/WCAG-EM/ (last access: 18/02/2016)
tool monitors public websites on a continuous basis according to a confined number of automatically performed tests and displays statistics on outcomes on a publicly available website.

As supported by the outcomes of an expert workshop conducted as part of this study, regular monitoring, even on the basis of a confined number of automatically performed tests, can be expected to stimulate website owners to make improvements at least in relation to those requirements that can be monitored by a fully automated procedure. Outcomes of such an assessment procedure - if conducted regularly - can in used supporting website owners in achieving higher levels of accessibility in relation to all WCAG 2.0 requirements, and even beyond WCAG 2.0. At the same time, conducting automated monitoring on a regular basis bears the risk of reducing website owners' perception of web accessibility to as a tick-box exercise that covers only those aspects that can be monitored automatically.

**Recommendations**

Consider flexible cut-off dates for conducting annual monitoring

Flexible cut-off dates for annual monitoring would provide an incentive to website owners for achieving accessibility on a permanent basis rather than at a foreseeable point in time of the year. As revealed by the expert workshop conducted in the context of the current study, it can be expected that website owner organisations will also be incentivised to adequately consider web accessibility within their budget planning, cutting across the budgetary years.

4.2.3. Testing and documentation

Article 6 of the Directive references EN 301 549 V1.1.2 (2015-4), in the first instance, as the standard which contains the accessibility requirements against which all Member States shall ensure their public sector websites which come under the scope of the directive conform with. To validate if a given requirement is met by a given website, a number of tests may need to be conducted.

It is not, however, a straightforward exercise to test for these requirements. There is always an element of human subjectivity and judgement involved when interpreting results. Other conditions within a testing framework may include cultural differences, as differences in what assistive technology, browsers and devices are used in the tests. Our research has shown that all these factors may vary between countries, thus making local implementation of the requirements contained in the EN both valid and important to achieve accessibility.
The monitoring exercise must contain aspects that help to increase the level of accessibility, as well as measuring it. Research shows that intra-organisational management procedures and work flow processes are key for achieving accessibility. A combination of self-declaration with training for website owners, manual cluster sampling made by experts and end-user involvement is therefore needed.

Although the Directive does not prescribe end-user involvement during the conducting of tests as part of the monitoring methodology, Article 8.5 (a) requires that the first report shall contain a “description of the mechanisms set up by Member States for consulting with relevant stakeholders”.

**Discussion of the available evidence base**

Many of the success criteria (requirements) stipulated by WCAG 2.0 are composed of several smaller issues required to be tested if a valid assessment is to be made on whether or not a particular requirement is fully met. To give just one example here, success criteria 1.3.1 reads “Info and Relationships: Information, structure, and relationships conveyed through presentation can be programmatically determined or are available in text.” Validating this requirement in operational terms may again include testing how headings are put into a webpage, how forms and/or tables are included and many more aspects. Depending on the design of and the content provided by a given website, checking WCAG 2.0 success criterion 1.3.1 may require a combination of different tests. However, WCAG 2.0 itself does not specify what tests are to be carried out and – equally importantly – how such tests are to be carried out for validating whether or not a particular requirement is met by a given website. Currently, there is no commonly defined set of testing procedures available, the result being that each web evaluator is free to decide what test is done for validating compliance with WCAG 2.0 requirements.

In the light of this, it may not come as a surprise that different monitoring initiatives have adopted their own interpretations on how to test individual WCAG 2.0 requirements. The methodologies included in the validation exercise conducted by the current study do not represent an exception in this respect. For some of them, no documentation of the particular test performed was available at all. In fact, this renders them non-repeatable by any external party which again makes it impossible to evaluate their external validity. When it comes to their internal validity, there is a risk of generating non-valid outcomes - at least partially - if different evaluators who may be involved in the monitoring actually apply specific tests in a different way. The more room for interpretation that is left for deciding what is “good enough” for passing a particular test, the greater is the risk of compromising the internal validity of the monitoring outcomes.

[^74]: [http://www.w3.org/TR/WCAG20/#content-structure-separation](http://www.w3.org/TR/WCAG20/#content-structure-separation)
The monitoring methodologies that have been evaluated in this study, for which documentation was available, have all defined their own sets of tests for evaluating WCAG 2.0 requirements. These do not necessarily, however, cover all possible implications of the success criteria set out by WCAG 2.0. This finding supports the need for developing commonly accepted tests, for several reasons:

1) As revealed by the validation exercise, the tests documented were suitable to test only parts of the success criteria stipulated by WCAG 2.0, while important accessibility requirements remain untested.

2) Website owners are incentivised to only focus on those WCAG 2.0 requirements for which a test has been specified, they may thus believe that they comply with WCAG 2.0 while at the same leaving untested requirements unconsidered.

3) Since the web technologies in general and assistive technology as well as e-government services in particular are constantly evolving, new tests will need to be developed on an on-going basis to keep pace with the reality of end-users’ experiences in using these services.

To tackle these challenges, the monitoring initiative identified in Norway (Method 1) includes, for instance, a regular review of a set of tests defined for the purposes of validating websites against WCAG 2.0 with a view to iteratively determining whether the set of tests defined serves its purposes. Depending on ten outcomes, individual tests are going to be adapted or entirely new tests are going to be defined, e.g. due to emerging development in web technologies.

When it comes to determining what tests may be suitable for validating a given website against WCAG 2.0 requirements, evidence suggests that the focus should, however, not focus merely on the design of a given website or on the content provided. It has been shown that contextual factors exert a strong influence on whether or not accessibility standards are indeed implemented by website owners.

For instance, the formalisation of organisational policies and processes directed towards achieving web accessibility plays a crucial role in that respect. This research suggests, however, that accessibility related activities often tend to be pursued in an ad-hoc manner rather than according to systematically specified workflow processes, management procedures or documented policies. These findings may help in explaining a noticeable ‘churn’ rate observed over the time by earlier monitoring initiatives; a substantial proportion of websites that had achieved a certain level of accessibility in one year seemed not to have sustained this in subsequent years.75

---

Intra-organisational work flow processes and management procedures required to effectively ensure sustainable levels of accessibility have, however, not yet received much attention in the context of current web accessibility monitoring initiatives.\textsuperscript{76} In this context, it has been highlighted that in reality a wide variety of parties are typically involved in setting up and maintaining public online services, including technical and non-technical staff (e.g. procurement experts, web developers and editors).\textsuperscript{77} Frequently external contractors are involved as well. A variety of activities may need to be effectively coordinated if accessibility related efforts are to yield sustainable outcomes (e.g. conceptual work, development work, content generation, validation/quality control and capacity building).

Low levels of awareness of the relevance of accessibility in relation to these activities and related knowledge among the various parties involved have been identified as acting as a barrier towards achieving fully accessible online services. This not only concerns specialist expertise within dedicated technical web teams and procurement units, but more general knowledge within the overall organisation as well, for example, on the part of content creators.

Involving website owners in the accessibility related assessment of their own websites presents a way to raise the level of intra-organisational awareness of the relevance of accessibility in general and of formalised organisational accessibility policies and processes in particular. Self-assessment procedures implemented as part of a monitoring exercise can, however, rarely be found, although they can be expected to help in raising current levels of knowledge and awareness about web accessibility. As the level of reliability of monitoring data collected by means of self-assessment can be expected to vary across individual website owner organisations, a combination of self-assessment and third party assessment would

\textsuperscript{76} For instance, a new methodology currently under development by the Swedish government was presented at a workshop held in the framework of this study which, inter alia, relies on regularly surveying website owner organisations about measures taken to ensure accessibility of their websites or online services. Furthermore, a British standard on web accessibility published in November 2010 (BS 8878 Web accessibility) has been designed to introduce accessibility, usability and user experience for disabled people to non-technical professionals, some of whom may be completely new to this subject. It gives guidance on process, rather than on technical and design issues. \url{http://shop.bsigroup.com/en/browse-by-subject/accessibility/?t=1} (accessed: 14.01.2016) [check consistency of date form – 14.01.2016 or 14/01/2016 or 14 January 2016 are all acceptable but usage should preferably be consistent]

seem suitable to enable sufficient levels of data reliability. While most or even all websites falling within the scope of the directive in a given country could conceivably be monitored by means of self-assessment, a subset of websites could be evaluated by a third party, and be used to show the actual status of accessibility at a sufficient level of reliability.

There are two main reasons for involving end-users in the monitoring process: quality assurance and democracy. It is generally accepted that involving users is the only way to make sure a product or a service will work well for the users. This is the case for usability and perhaps even more so for accessibility, since persons with disabilities have the same range of varying individual usability needs as everyone else, but can also have specific needs that are very hard to detect if not tested with real users.

Involvement of end-users could be done at different levels, for example, through consultation in developing the monitoring methodology, as a means of quality assurance by some sort of users’ panel when the actual tests are being made and more directly via communications with individual users commenting on problems with a specific website.

In some monitoring initiatives examined in this study, end-users were involved, often on a semi-professional basis. These ‘semi-professionals’ are experts on their assistive technology and have a good knowledgeable and ‘lived experience’ of accessibility issues. These users were either involved in manual testing or their input regarding the accessibility was integrated as an “add on” to the monitoring exercise.

Another factor deserving attention when it comes to determining a procedure suitable for validating a given website against WCAG 2.0 requirements concerns the sampling of the content to be tested. As in the case of defining individual tests to be performed (as discussed above), approaches adopted towards defining how content which is to be tested should be sampled from a given website vary a lot across the monitoring methodologies included in the validation exercise conducted as part of this study. Individual websites tend to vary a lot in scope and design, for example, when it comes to types and volumes of content provided, web technologies applied, levels of interactive enabled etc. It is anything not a

---

78 Often the data is not really used as part of the monitoring exercise but rather to obtain an extra dimension. For example, one initiative included an expert evaluation of all success criteria in WCAG and each website got a score showing how many success criteria were failed on level A and AA. The end-user visited all websites and his sense of how easy or difficult it was to access the information was included as part of a qualitative assessment of the website, i.e. “Relatively accessible”, “Accessibility problems” or “Large accessibility problems”.

---
straightforward exercise to ensure application of the tests defined to a representative sample of content provided by given website.

The W3C Website Accessibility Conformance Evaluation Methodology (WCAG-EM) includes a section providing support in this respect. Advice is given on how to sample static, dynamic, and mobile websites and applications, rather than merely focusing on URLs.

Only one of the methodologies included in this study's validation exercise has actually applied WCAG-EM for sampling purposes. More generally, the outcomes of the validation exercise suggest that the sampling of content from a given website is one of the key factors explaining differences in outcomes of the assessment procedures involving manual testing. Often, the URL is given as the only sampling criterion in the framework of existing monitoring initiative. This may leave dynamic content and transactional online services involving different interactional steps within the overall process of service delivery entirely unconsidered when assessing the accessibility of a given website.

When it comes to documenting the outcomes of tests applied to content sampled from a given website, the WCAG 2.0 success criteria are formulated in a way so that they can either be met or not met, in terms of a “yes” or “no” statement. However, they do not provide an indication of how well a requirement is addressed, in other words of how important the accessibility barrier experienced by the user might be in practical terms. For example, a requirement may be not met because of a minor issue that does not strongly impact end-users’ experience, or because of a more substantial issue that causes significant access barriers to the users. Even if not all WCAG 2.0 success criteria may be fully met by a given website, it may nevertheless be usable in principle despite the user possibly experiencing a certain level of inconvenience.

In the light of this, different types of metrics are typically used by current monitoring methodologies:

- **Binary:**
  Using a binary scale means that there is only one decision to be made, i.e. is this website good enough or not. This means that we only have to define the difference between pass and fail.

- **Percentages:**
  Using a continual scale requires very specific algorithms to calculate a figure. For example, what does it mean to meet a requirement at X per cent? Today there is no single widely adopted algorithm to allow such calculations that are valid.

---

79 http://www.w3.org/TR/WCAG-EM/#step3
Ordinal type:
Some initiatives use scales, such as 2 = Pass, 1 = Marginal fail, 0 = Fail. This makes it possible to create a metric that not only passes and fails but that gives a more nuanced picture of the accessibility levels. The drawback is that this requires more definitions and more decisions to be made while assessing a website, but is less difficult than a percentage scale.

Most frequently, binary metrics have been used by current monitoring initiatives in terms of a simple ‘pass’ or ‘fail’ logic. There are, however, other metrics which are more pragmatic or “generous” in terms of indicating the situation in relative terms.\(^80\) In general, the challenge associated with WCAG 2.0 level AA requirements is that each success criteria can be broken down to a large number of sub-requirements that could be met with more than one solution. This makes it very difficult to have one single metric per requirement. Many methodologies therefore do not focus on the requirement as single unit to which dedicated metrics are applied. Typically, one or more tests tend to be applied per requirement, and often binary “pass” / “fail” metrics are applied to single tests rather than a complete success criterion (i.e. the complete requirement).

In reality, these methodologies are not suitable to claim that a particular success criterion is met because all tests for that success criterion have actually been passed. There could still be accessibility problems remaining unconsidered by the test procedures applied. Instead they focus on fails. A success criterion could be failed if a test finds a problem but the existence of no problems is not proven just because the tests did not find any. This in turn makes the metric semi-binary (either “fail” or “cannot verify”).

**Recommendations**

*Combine self-declaration with expert evaluation and user involvement*

---

\(^80\) A more fine-grained assessment based on WCAG 1.0 which was conducted by previous research across several EU countries has, for instance, shown that government websites were considerably more likely to fail only marginally in comparison with non-government ‘public interest’ sites. This research suggests that up to two-thirds of the government sites in the covered countries may have had quite a degree of accessibility, even if not fully compliant, and that such progress was not being picked up in generally used simple ‘pass-or-fail’ metrics. See: K. Cullen, L. Kubitschke, T. Boussios, C. Dolphin and I. Meyer (2009): Study on “Web accessibility in European countries: level of compliance with latest international accessibility specifications, notably WCAG 2.0, and approaches or plans to implement those specifications”, on behalf of the European Commission. Available at: https://ec.europa.eu/digital-agenda/en/news/study-report-web-accessibility-european-countries-level-compliance-latest-international (latest access: 18/02/2016)
Self-declaration is useful in order to increase awareness about accessibility. To ensure reliability, it has to be combined with a defined set of tests performed by experts. The results of both should be published online for total transparency.

There should be regular contact with the Disabled People’s Organisations and individuals with disabilities in order to involve users at different stages:

1. During planning: offer the possibility to contribute to the monitoring methodology as such by proposing the websites and/or aspects to be monitored;
2. During the testing exercise: setting up a users’ panel to help with the testing;
3. After the monitoring exercise: make the results public to enable end-user organisations to challenge any website owner that claims to be accessible without actually complying.

To enable website owners to perform the tests and learn about accessibility at the same time, many good sources of information on web accessibility are available in different languages. For example, there are high quality online resources of this kind in Spain, Sweden, France, Germany, Belgium, Norway and other countries. Some of the tests in the Norwegian method (Method 1) are partly based on supportive tools, for example, the W3C Mark-up Validation Service.

**Continually evolve and expand the set of testing procedures being used**

A mechanism for continued revision – and if required further development – of a defined set of tests should be embedded in the methodology. This way the monitoring in itself will help increase accessibility of the tested websites as well as capacity among suppliers who will need to keep their knowledge-base up to date when different aspects are being tested each time. Another benefit from this method is that all success criteria in WCAG 2.0 AA can be covered without the monitoring exercise being overwhelmingly resource demanding. Yet another is that the monitoring methodology will remain relevant in a fast changing technological world.

**Base the sampling on WCAG-EM and use a definition of “sample” that includes online service delivery procedures and tasks**

When selecting a sample, it is important to include most types of content and usage situations that users typically encounter. Define a clear sampling approach relying on WCAG-EM. For example, help texts, dynamically expanded areas with options, auto complete in search fields and other complex technologies should definitely be considered in this context. A sample should therefore not only be defined in terms of a set of URLs but also a set of tasks, similar to the so called ‘mystery shopping’ technique, as illustrated in the following example: “Include the start URL:
Use ordinal type scales for metrics

Since accessibility is complex and full of nuances, many metrics will be too blunt in some cases. The least problematic way of showing results is to use a scale like 2 = Pass, 1 = Marginal fail, 0 = Fail. It is important to also provide an alternative for “Not possible to test” which is not the same as Fail.

4.2.4. Format of reporting

Article 8.5 of the directive contains some direction on the contents of the report to be submitted by Member States. It requires that the first report, to be submitted five years after the date of entry into force of this Directive, contains:

“(a) a description of the mechanisms set up by Member States for consulting with relevant stakeholders on the accessibility of websites and mobile applications;
(b) procedures to make public any developments in accessibility policy relating to websites and mobile applications;
(c) experiences and findings from the implementation of the rules on conformity with the accessibility requirements set out in Article 4; and
(d) information on training and awareness-raising activities.”

In the context of this study, and the evidence base gathered on which recommendations are to be made on the format of reporting, the following discussion and recommendations are presented.

The aggregated results across websites are probably the most interesting metric from a monitoring perspective. This is where it becomes possible to see patterns and changes on a larger scale. The result for a single website can be affected by many different factors and the result might therefore not be representative for the accessibility levels of that particular website. But in the aggregated results from a larger number of websites these variations on single websites will not affect the overall results that much.

It is possible both to look at the overall results (i.e. average of an aggregated value for individual websites) or on results for individual requirements (i.e. 80 per cent of all websites fail to meet this requirement).

Benchmarking information can also work as an incentive to those websites performing less well to make an effort directed towards catching up when compared with the ‘high-performers’.
Transparency of testing and results will be important if outcomes are to be scientifically trustworthy.

**Discussion of the available evidence base**

Very few initiatives looked upon in this study have included aggregated results over a set of websites. One example where this has been done is the initiative made in Norway (Method 1) where different sectors of websites have been compared. The reason for this has been to be able to see where to focus resources. If one sector has larger problems than the other sectors, more resources can be used to monitor that sector and aid the accessibility work there. This could be done for sectors, but also countries. If accessibility levels in different countries are compared, actions for different countries or regions can be customised.

At the workshop and Advisory Board meeting conducted as part this study, there was agreement between participants that a structured way of continuous sharing of knowledge and experience in monitoring in the way that was taking place in the study, would be very positive. For example, both Norway and Sweden have official online support material widely used by website owners and suppliers as well as a market for accessibility experts.

The review of existing monitoring initiatives revealed that often publicly available information on the methodological approach adopted by a given initiative would not enable replication of its results. This can potentially adversely influence levels of acceptance at the part of website owners and the industry as well.

Previous research has, for instance, revealed that web managers sometimes question the utility of regular monitoring as they feel that there was always a subjective element to it if it is done in sufficient depth, i.e. not only in terms of automatic testing. If fully transparent, end-user organisations can also use the results from the method as the basis for discussion when engaging with or providing feedback to website owners and policy makers.

**Recommendations**

*Use the benchmarking results to focus the expert evaluation where it is best needed*

Based on the results from the benchmarking exercise, the expert evaluations can be made to focus on specific sectors or technical aspects that need further attention.

Tracking of progress over time would be useful across different types of online services and countries as well to enable identification of any potential structural accessibility deficits.

*Adopt a strategic perspective towards Europe-wide benchmarking which is suitable to maximise impacts*
The available evidence base suggests that tangible impact of a benchmarking approach mainly directed towards incentivising compliance with web accessibility requirements through a “naming and shaming” approach can be expected to be rather low, particularly in a public sector economic environment in which resources continue to be scarce.\textsuperscript{81} There may, however, be cultural differences in how “naming and shaming” on one hand and promoting “the forerunners/the good example” is perceived. Generally, adoption of a benchmarking approach should be considered which is primarily directed towards meeting website owner organisations at their current position and helping them to understand what and where improvements are called for.

**Consider setting up a Europe-wide bench learning mechanism for website owners**

Consideration should be given to setting up a cross-national mechanism enabling individual website owner organisations to learn from one another. To this end, it should be considered to not solely rely on metrics-based benchmarking which focuses on the technical accessibility of a website, but to also focus on the comparison of organisational practices, thereby focussing on formalised accessibility policies and processes that have been put in place at the level of the website owner organisations. This seems particularly important as the available evidence suggests that often accessibility-related activities tend to be pursued at the level of website owner organisations in an ad hoc manner rather than according to systematically specified workflow processes and management procedures.

**Make both tests and results fully transparent**

A common documentation framework should be applied in all Member States and outcomes be made publicly available. This could, for example, be done by setting up a database with a web interface where it is possible to see test procedures and results, and where it is also possible to report results. W3C has a reporting tool for WCAG-EM\textsuperscript{82} which makes it possible to document findings and to export them in JSON format, which could be imported into a database, but the


\textsuperscript{82} https://www.w3.org/WAI/eval/report-tool/#/
database itself does not exist and the exact reporting format needs to be developed to align with the methodology used.

As an example, such a database could include fields for:

- Website assessed
- Website owner
- Website sector
- Scope definition of the website
- Timeframe when the tests were conducted
- Reference to who or what organisation was responsible for the testing
- Sampled pages and views/states
- Title of each test to be done
- References to relevant success criteria in WCAG for each test
- Results for individual tests
- References to what sampled pages/views/states failed a specific test
- Comment, why a test where failed

The database could also include the test procedures themselves so that it is possible to update tests, add new tests and remove tests continuously to make sure that the tests are relevant and up to date. By keeping it in the same database it is possible to view results over time to see if accessibility levels are improved, even if individual test procedures have been updated.

It could also be fruitful to include fields to store end user feedback on the websites to form a more complete picture of each websites accessibility level.

Such a database could have different web interfaces added:

- One to give national experts the possibility to report “official” findings.
- One to let website owners do self-assessments.
- One to let the public see the results and give feedback on accessibility problems perceived.
- One for the European Commission to get aggregated data on sectors, countries and for the whole EU.

4.2.5. Cost of web accessibility monitoring

As discussed earlier in this report (section 2.3), a detailed analysis of the cost structure of the individual monitoring initiatives investigated in the framework of the current study was not possible on the basis of information available to the study team. The recommendations provided in the following are therefore based

---

83 For example, see background in https://www.w3.org/TR/WCAG-EM/#step1a
on experiences gained during the validation exercise and earlier work of the key knowledge holders of the study consortium, in particular Funka and W3C.

**Discussion of the available evidence base**

The information gathering exercise conducted for the purposes of this study revealed that validated economic data is rarely available in relation to the monitoring initiatives under investigation. Usually, particular cost items such as method development, website evaluation and documentation of outcomes tend to remain undocumented or are not easily retrievable from existing records. It may therefore not come as a surprise that cost estimates provided for some of the investigated initiatives vary a lot, ranging from below one Euro up to almost 200 Euro per tested website. While such estimates may provide at least a very rough indication of the order of magnitude at which costs may have accrued per tested website in the framework of existing monitoring initiatives, they do not however enable a detailed and robust analysis of the cost structure of the individual approaches under investigation.

Available evidence on costs involved in designing and/or redesigning websites so that they comply with WCAG 2.0 requirements suggest however a correlation between the costs involved and the level of sophistication ultimately achieved by a given web presence, e.g. in terms of functionalities provided to the end user and particular web technologies utilised. As supported by the outcomes of the validation exercise conducted in the framework of the current study, it seems appropriate to assume that the effort required for monitoring the accessibility of a given website varies respectively, e.g. of a fully transactional e-government service when compared with a public website which mainly provides static content. In the absence of robust empirical data, the following recommendations are based on estimated effort figures.

**Recommendations**

**Consider the relationship between cost and benefits when specifying the scope of the expert evaluation to be carried as part of web accessibility monitoring**

In the context of the current study, it took a trained expert on average 7-9 hours of work to evaluate a given website against WCAG 2.0 AA with a view to identifying common accessibility failures according to a set of selected tests. Compared to this, it is estimated that it would take the same experts on average between 40 and 60 hours of work to perform a full manual expert evaluation of all success criteria of WCAG 2.0 AA based on WCAG-EM. It therefore seems reasonable to consider focussing the expert evaluation on identifying common accessibility failures with a view to achieving an acceptable cost/benefit ration for regular monitoring purposes.
Despite potentially lower costs involved in fully automated web accessibility monitoring approaches currently available, use expert evaluation for the purposes of regular web accessibility monitoring.

Although there is a paucity of robust empirical evidence, it seems reasonable to assume that a fully automated evaluation approach enables web accessibility monitoring at lower costs per website when compared with approaches involving human evaluators. However, the capabilities provided by fully automated solutions as they currently exist leave important accessibility requirements unconsidered. Given the current state of the art in fully automated website evaluation, comprehensive web accessibility monitoring still requires the involvement of human evaluators who possess a sufficiently detailed understanding of web technologies implemented by the website owners, evaluation tools available, barriers that people with disabilities experience, assistive technologies and approaches that people with disabilities use as well as accessibility guidelines and techniques. This will become even more important as public websites becoming increasingly more interactive in terms of transactional e-government services and new web technologies which are continuing to emerge. Such developments are likely to remain unconsidered by fully automated accessibility monitoring solutions, as they currently exist.

Donate a sufficient amount of resources to collaborative web accessibility monitoring in order to make the involvement of end users a valuable exercise.

Pursuing a collaborative evaluation process involving the skills and perspectives of people with disabilities can enhance the value of web accessibility monitoring. Different aspects deserve attention when considering the effective involvement of users in web accessibility evaluations. For instance, it needs to be ensured that diversity in disabilities will be adequately represented, that different types of assistive technology will be used, and different levels of experience with the Web will be represented as well. The study team estimates that 8 to 10 users with different abilities would on average need to spend 40 to 60 hours of work on evaluating a given web in order to enable the generation of a useful output.

4.3. Towards implementing a sustainable web accessibility monitoring system at European level

The evidence presented throughout this report shows that there is a lot of variety across the EU Member States – and sometimes even within individual countries – when it comes to the question how web accessibility monitoring has been carried out today. As reflected by the recommendations presented in the previous section, the evidence base also shows that web accessibility monitoring is more than a
mere “tick-box” exercise and that a wide range of conceptual and practical aspects deserve attention in this context. Respectively, each of the monitoring initiatives investigated for the purposes of the current study had chosen its particular focus in terms of which websites were monitored, against which requirements these were tested, how these tests were performed in practical terms and how the results were documented, and possibly further utilised.

Beyond the recommendations presented in the previous section, in the following subsection some conceptual and practical considerations are therefore presented on how to practically arrive at a common web accessibility monitoring system which is relevant across the entire EU (4.3.1). This is followed by more practical considerations on how a common monitoring methodology could be implemented in more practical terms (4.3.2).

4.3.1. Conceptual considerations

In view of the great diversity of monitoring approaches identified throughout the EU, it seems useful to avoid “reinventing the wheel” wherever possible. This relates to a number of aspects discussed throughout the previous sections. In particular, there would be merit in considering a number of preparatory steps.

Preparatory step #1: Defining a common testing procedure

Assessing a given website against WCAG compliance requires a wide range of individual tests being performed. For achieving comparability across the EU, it will be important to ensure that the same tests are carried out in exactly the same manner in all countries. This requires a carefully specified testing procedure. Clearly, there would be merit in taking advantage of the work already done in the countries analysed within the scope of the study. As a starting point, it is therefore recommended to start translating tests from the Norwegian method (Method 1) into different languages. In fact, the Norwegian methodology has proved to work well in terms of accuracy. It is also appropriate to adequately consider technology trends, use of a combination of tools and expert evaluation. These tests could form a baseline for the proposed monitoring methodology.

Preparatory step #2: Relying on training and self-declaration as a means of sustainable capacity building at the part of the website owners

There are examples of good practice, e.g. in Sweden and Poland, where a combination of training and self-declaration approaches have been successfully exploited for sustainable capacity building among website owners organisations, beyond mere accessibility monitoring. There is merit in considering a similar approach for the purposes of an EU-wide web accessibility monitoring, as will be discussed in more detail in the next section. The self-declaration utilised in Sweden
or Poland can serve as a useful starting point for developing a European method. It is therefore recommended to translate these approaches into other languages in order to make these available as the basis for the recommended monitoring methodology at EU level. Alternatively, a similar approach in New Zealand has already been published in English: https://www.ict.govt.nz/guidance-and-resources/standards-compliance/web-standards/2014-web-standards-self-assessments/2014-web-standards-self-assessment-report/

Preparatory step #3: Setting up a European network of experts.
When setting up a European-level monitoring system that is truly sustainable, it will be important to bear in mind that the field of web accessibility testing is undergoing permanent changes. In line with new web technologies and practices that are constantly emerging, new assessment/monitoring tools and techniques are likely to emerge over time. In fact, ensuring web accessibility needs to be seen as an on-going journey rather than a final state of play that is to be reached at a certain point in time. Therefore, a sustainable monitoring methodology needs to flexibly cater for on-going changes within this domain. Any defined web accessibility monitoring methodology should thus be adaptable with a view to adequately responding to new technologies and practices.

Against this background, there is merit to setting up a network of specialists at European level. This would help in further developing the initially specified monitoring methodology, thereby ensuring relevant and high quality outcomes. Such a network could also effectively interface with end user organisations as well as support knowledge sharing and benchmarking more generally. The expert group would also be able to provide support and advice to individual Member States. Such a network could further be involved in the operational management of a European monitoring system, as will be further discussed in the subsequent section, e.g. when it comes to managing an online facility where national monitoring results can be uploaded and/or presented to the wider public.

Preparatory step #4: Determining the scope of an EU-level monitoring exercise
To ensure that its outcomes are valid and practically relevant, it will be important that a European-level monitoring exercise relies on a sample of websites that reflect the real world situation in the individual Members States from the perspective of disabled users. As discussed earlier in this report, public services are organised differently in different countries. This also relates to the issue of which party is providing which services through online media, and how exactly these are being provided. When determining which websites are to be tested in each Member State, a certain level of flexibility will therefore be required. The size
of a given country and the way public services are organised will influence what will be practically feasible.

In general, it is desirable that services which are most relevant to citizens are covered. Often, users are in need of contacting services delivered by local authorities rather than national government services, e.g. when it comes to addressing day-to-day issues. On the other hand, national government services tend to be of particular importance when it comes to accessing information for political opinion forming, and they tend to serve as a “role model” for other sectors. Against this background, it is desirable to achieve a good balance between public services delivered at the national, regional and local governance levels.

Taking this into consideration, the following aspects deserve attention:

- The sample of websites to be monitored should be **relevant to users**. In fact, disabled users do not have other interests than the public in general, even if the target group sometimes have extended needs. For accessibility to be meaningful, it needs to cover what people do in their everyday lives.
- The sample should include those services which are of **relevance in a given Member State**. The sampling approach should therefore enable the flexible selection of individual websites across different levels of government. For instance, while some services may be available from a single national-level provider in one country (e.g. a National Health Service), in another country the same services may be available from a number of service providers at the local/regional level (e.g. local health centres / doctors’ offices).
- The sample should cover the **whole online service delivery chain** rather than merely an entry page to a given public service that is being delivered by means of online media. If the monitoring exercise covered only a sample of static websites or the entry pages for fully digitised online services –as in the case of fully automated monitoring approaches currently pursued in some countries – the outcomes would not be relevant when it comes to an end user’s experience when trying to utilise a fully digitised public service. In any case, end users with disabilities should be enabled to utilise interactive services like any other citizen, rather than just accessing static information. In other words, the sample to be drawn should enable performing a variety of tasks typically involved in the online delivery of public services, including, for example; finding information, self-identification of the user, filling in applications, accessing dynamic content, filling in forms, verifying content, getting feedback etc.

As a starting point for achieving a meaningful sample of websites to be tested, it is therefore recommended to analyse which public services in each of the EU
member states are of particular relevance to the citizens at each of the three governance levels mentioned above. Also, such an analysis should yield an overview of the level of interactive public services have gained in the various EU Member States by now, and it should shed light on any practical impacts this may have for assessing relevant online services respectively. The latter concerns, for instance, the question of how to enable testing of online services requiring any legally defined registration procedures, as it is, for example, in the case with online tax declaration services available to the citizens in some countries. Existing analyses and statistics from the eGovernment domain may serve as a useful source of information in this respect, such as the latest eGovernment Benchmarking Report published by the European Commission (https://ec.europa.eu/digital-single-market/en/news/eu-egovernment-report-2015-shows-online-public-services-europe-are-smart-could-be-smarter). There would also be merit in consulting user organisations on this matter.

Article 1 of the Directive also contains a number of exemptions and extended timelines for certain types of content which should be considered in the context of determining the scope an EU-level monitoring exercise.

As outlined in section 2.4, exempted content includes:
- Live time-based media;
- Online maps and mapping services – maps used for navigation purposes must be have an accessible alternative;
- Third-part content that is not funded, developed or under the control of the public sector body;
- Reproductions of items in heritage;
- Websites of public broadcasters;
- Websites of NGOs that do not provide services “essential to the public” or address the needs for persons with disabilities; and
- Schools, kindergartens and nurseries except for content relating to “essential online administrative functions”.

Contents with longer lead-in time includes:
- Websites published more than 21 months after the date of entry into force of the Directive, applicability starting 45 months after the date of entry into force of the Directive (Article 12(3.a));
- Office file formats, published before 21 months after the date of entry into force of the Directive;
- Pre-recorded time-based media published before 45 months after the date of entry into force of the Directive;
Extranets and intranets whose content is available only to a “closed group of people and not to the general public” published before **33 months** after entry into force of this Directive; and

- Archives – content that is “neither needed for active administration processes not updated or edited after” **33 months** after the date of entry into force of the Directive.

### 4.3.2. Practical considerations

In the process of developing a web accessibility monitoring methodology to be implemented across Members States, seven key steps are also considered and some practically oriented recommendations and suggestions are provided.

**Practical step #1: Sampling of websites to be included in the monitoring exercise.**

Following an analysis of which websites are of particular relevance for the end users in each country, as suggested in the previous section, a common selection procedure will be required to be applied across the EU. For each of the public service domains considered to be included in the monitoring exercise, a detailed sampling procedure will need to be defined, to include health services, employment services, educational service etc. As discussed earlier, there is great diversity across countries in relation to the type of public body which delivers a particular service to the public, at which level of government it is delivered and to what extent it has yet been digitised. Also, in many countries a number of services tend to be available through a common online portal or entry website. For instance, in many countries the national employment service runs a web portal through which a diverse range of services are offered ranging from general advice to the unemployed, to dedicated online tools for job hunting, right up to services directed to employers rather than employees.

Against this background, some generic sampling principles may be considered as practically feasible across all countries:

- If a given service is available at the national level, the national website through which it can be accessed could be sampled.
- If a given service is not available at national level but at the regional level, the website through which it can be accessed by the citizens of the largest region in the country, measured in terms of inhabitants, could be sampled.
- If the service is neither available on the national level nor on the regional level but on the municipal level, the website through which it can be accessed by the citizens of the largest municipality in the country measured in terms of inhabitants could be sampled.
• If a public service based on a life event is available through a portal managed by the authority/body providing the service in question, and if this portal offers more than this one service to the citizen, the URL of the 1st webpage on the website (or portal) from where the end user can access this service could be sampled. For instance, in the case of employment services, the 1st webpage addressing the life event service of “job seeking” from the perspective of the unemployed person or an employee could be sampled (not from the perspective of the employer).

Overall, a suitable number of websites should be sampled to adequately reflect the structure and size of the overall public sector in a given country. Also, economic considerations may need to be taken into account, as costs for monitoring are also related to the sample size. The evidence collated by the current study suggests that on average a sample size of 250 websites per country may be considered as suitable for representing the real situation in a given country, while at the same time being considered as practically feasible across all countries.

The recommendation of an average sample size of 250 websites is based on the following findings. The monitoring methods in Europe today range from 11 to 3100 tested websites, with an average of 488 websites tested in each country (figure 6). There is a correlation between the chosen method of testing, the quality of the results and the numbers of websites tested (4.2.1). This study shows that to make sure that the testing is valid for the needs of persons with disabilities, correct and reliable, it needs to be done by experts. This is even more important when it comes to the dynamic and interactive services that are over taking the static websites of former information services. Therefore, our recommendation of a monitoring methodology strongly suggests to focus on quality rather than quantity. 488 tested websites being the average of the existing monitoring methods, it seems that a number of around 250 websites per country would be feasible for all Member States, taking economic aspects into account. The mentioned 250 websites should also be enough to cover the majority of relevant public sector services that are important for end users at national, regional and local level.

Testing results from 250 websites per country should provide enough material to see trends in different sectors and be statistically significant enough to be reliable for benchmarking purposes. Of course, the exact number must be decided in each Member State depending on the structure of the public sector services in the respective country and taking into account that the Web Accessibility Directive has evolved since this study was made, so it is now also covering mobile interfaces which of course must be added to this recommendation of items to be tested.

Practical step #2: Frequency of the monitoring exercise.
It is important to consider a flexible cut-off date for conducting annual monitoring. Websites are constantly evolving, with new content added frequently. This means that monitoring must be conducted regularly to show relevant results. A yearly monitoring seems reasonable both from a resource perspective and also considering how much attention the publication of results would get. Also, most public sector authorities work with one year budgets, so an annual monitoring budget would fit into the overall process of web development, management and maintenance. By making the cut-off date flexible, the monitoring would provide an incentive to work with accessibility on a permanent basis, instead of fixing small things just before the monitoring date.

**Practical step #3: Testing approach.**

Since the monitoring is legally required to show compliance, automatic tools cannot be used as the only way of monitoring. None of the automatic tools that are in the market these days are capable to check for all success criteria in WCAG 2.0 AA (see Annex 3 for further information). In fact, they do not check the respective success criteria in full, but only parts of them. Furthermore, the automatic tools are not reliable when it comes to accuracy. Rather, automatic tools should be seen as a support for the expert doing the evaluation.

It is important to remember that the web accessibility testing field is changing rapidly. New tools can be delivered and new techniques can be invented. Any method should cater for the possibility of flexibility, which is why the study team also recommends a constant evolution of the methodology.

In the Norwegian method, which is the one that we recommend to base the European monitoring methodology on, there is a large set of tests to be performed. The tests are recently developed, and constantly evolving, in two levels:

- The monitoring body decides which tests to perform at any given date, so out of say 100 tests, they only perform 20 in one given year. This is important because it makes the monitoring less resource demanding without counteracting innovation and a broader accessibility work.
- The monitoring body is constantly looking into the development of technology and enhancing the tests and developing new ones in this fast changing world. This is important because the tests need to evolve to keep up with reality and provide possibilities for future innovation.

**Practical step #4. Tests to be carried out.**

Being flexible and changing what tests to perform every year is crucial. To be able to test for all success criteria with good quality, testing would be extremely resource demanding. That is why the study team recommends to have test
procedures for all success criteria, but only test for a sample of them each time. This has been tested in real life in Norway and it works very well. It is also easier to market, since the complex story of accessibility can be made easier to understand for media and websites owners alike.

Limiting monitoring to the same sample tests every year would counteract accessibility improvements. This has been experienced in many countries and it is clearly limiting the aim for accessibility to end with a small number of identified tests, instead of working towards compliance. The decision on what to focus on each year could be decided nationally or centrally by the network of specialists. We strongly recommend the monitoring body – central or national - to work closely with the DPOs to understand what is most pressing at any given point. The advantage of making this at national level is of course that focus can differ between countries and the closer to end users you are, the more accurate the focus will be.

If raising overall agreement turns out to be too complicated, a random sample of tests every year could be a second option. This would have the same positive push to improvement, but without the accuracy in monitoring the most prioritised topics from a user perspective.

To keep on developing the tests is absolutely crucial when it comes to the web – this is truly a moving target. The WCAG standard is technology independent, so when technology moves ahead, new testing techniques are needed. Therefore, this is an important issue justifying why the network of experts can play an important role.

Choosing which tests to run should also take into consideration the differing timelines for contents under the Directive, such as:

- office formats;
- pre-recorded media;
- extranets/ intranets; and
- archives.

Practical step #5: Testing procedures

One of the most important practical recommendations involves building a list of test procedures. This should include the following three parts:

- **Combined training and self-declaration (all websites, sample of tests).**
  It is important to raise awareness and build capacity among website owners. In this way, the monitoring will not only monitor but also increase accessibility. Website owners need to be able to declare the level of
accessibility through conducting easy tests. The combination of training and self declaration will train them and increase knowledge. The tests chosen for self declaration will be the easiest ones to perform by non-experts with the help of tools. The point is not to get 100% accurate results, but to progressively train the website owners to measure accessibility. If for example year one has a focus on links, there could be made 5-10 easy tests for website owners that also work as a manual for how to publish links in an accessible way. Instead of filling in a form saying “my site is accessible (or not)”, the website owners will be led through a series of testing procedures where they learn how to make an accessible link. So at the same time as they declare the accessibility of the website, they get the chance to improve this part of accessibility. There is already self declaration going on in Sweden, Poland and New Zealand with interesting and positive results.

- **Expert evaluation (sample of websites, all tests).** This is made after the website owner has uploaded its self-declaration result. Expert evaluation is carried out to ensure a valid result of accessibility progress, and to provide benchmarking statistics for compliance. Please note that the recommendation is that this testing is made for all success criteria but only on 50 websites per Member State. The idea is to base the test on the Norwegian method, and to use the evaluation method of WCAG EM (See detailed explanations in Annex 2). If each country has 250 self declaration documents and 50 expert evaluations for compliance, it will be possible to draw conclusions on the accuracy of all websites in the country based on these statistics.

- **End user involvement (selection of tests to focus on, handling of complaint mechanism).** This approach is helpful in at least two ways: after the website owner has uploaded its self declaration result, end user organisations and individuals with disabilities can react on lacking accessibility by using the feedback mechanism. This will in itself help website owners to understand and make things right. The dialogue around issues to focus on is another important aspect of end user involvement.

**Practical step #6: Time line for the monitoring cycle**

The testing exercise, as recommended in Practical step #3, should imply a flexible cut-off date for conducting annual monitoring. Below, an example on how it could work with timings is provided.

Example:
January: The monitoring body presents this year deadline for self declaration.  
February: Conference with the DPOs to choose the focus of the year.  
March: Translation and communication on the upcoming testing.  
April: Preparation of tests and sending out a link to the self declaration tool.  
May: Deadline for upload. Reminder is being sent out automatically by the system.  
June-August: Expert evaluations are carried out on a small sample of websites. This can be done randomly and/or on websites whose self declaration results seem unlikely.  
September: Publication of results.  
October: Dialogue with website owners on actions to improve accessibility.  
November: The monitoring body presents next year deadline for self declaration.  
December: Conference with the DPOs to choose the focus of the year...

The final Directive contains detailed timelines on the implementation of the monitoring. This practical step does not contradict or prejudice these timelines.

**Practical step #7: Reporting of results**

Based on the specific examples of good practice found in the countries analysed, it is important to ensure total transparency of the monitoring methodology, its implementation and its reporting. Taking into consideration that the monitoring results shall be used by different target groups, it is desirable to publish results in different formats and level of details, as follows:

- Raw data for benchmarking between countries or sectors and to measure improvements.
- Specific data for each website owner (including detailed test results) from manual expert evaluation to encourage further development.
- Overall results for the media, general public and DPOs.

There is also be merit to considering making outcomes available in commonly accepted Open Data formats, e.g. through the European Union Open Data Portal (https://data.europa.eu/euodp/en/data). This would enable easy access to a growing body of monitoring data through a single point of access. This would enable easy processing of the monitoring data by third parties, e.g. for research purposes. Also, such an approach would help fostering the transparency and the accountability of a European web accessibility monitoring system.
Annex 1. List of stakeholders consulted for the purposes of the study

Organisations consulted for the in-depth analysis of 12 selected monitoring initiatives

<table>
<thead>
<tr>
<th>No. / Country</th>
<th>Title</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Accessibility Monitor (Toegankelijkheidsmonitor / Moniteur de l'accessibilité)</td>
<td>Blindenzorg Licht &amp; Liefde</td>
</tr>
<tr>
<td>2</td>
<td>Mapping of web-accessibility on public websites</td>
<td>Danish Agency for Digitisation</td>
</tr>
<tr>
<td>3</td>
<td>The conformance of public sector websites to WCAG 2.0 requirements in 2013.</td>
<td>Ministry of Economic Affairs and Communications</td>
</tr>
<tr>
<td>4</td>
<td>Web Accessibility Observatory</td>
<td>ICT Directorate at Ministry of Finance and Public Administrations</td>
</tr>
<tr>
<td>5</td>
<td>Ce que les sites Web publics nous disent de leur accessibilité</td>
<td>Braillenet</td>
</tr>
<tr>
<td>6</td>
<td>Simply Accessible (Gewoon toegankelijk)</td>
<td>Quality Institute of the Dutch Municipalities (KING); The Association of Dutch Municipalities (VNG); The Ministry of the Interior and Kingdom Relations (BZK)</td>
</tr>
<tr>
<td>7</td>
<td>State of Affairs in the Kingdom (Status i riket)</td>
<td>Agency for Public Management and eGovernment (Difi)</td>
</tr>
<tr>
<td>8</td>
<td>Accessibility of Higher Education (dostępności stron uczelni wyższych)</td>
<td>Foundation Institute for Regional Development</td>
</tr>
<tr>
<td>9</td>
<td>Acessibilidade dos Conteúdos Web na Administração Pública Central (Web Content Accessibility on the Central Public Administration)</td>
<td>FCT Fundação para a Ciência e a Tecnologia / Unidade de Acesso</td>
</tr>
<tr>
<td>10</td>
<td>What is the State of Affairs? (Hur är läget?)</td>
<td>Swedish Agency for Participation</td>
</tr>
<tr>
<td>11</td>
<td>Accessibility of webpages in Slovak republic</td>
<td>Slovak Blind and Partially Sighted Union</td>
</tr>
<tr>
<td>12</td>
<td>Gov.uk Reporting on progress: central government websites</td>
<td>The Government Digital Service (GDS)</td>
</tr>
</tbody>
</table>
Organisations consulted for the validation exercise

<table>
<thead>
<tr>
<th>No. / Country</th>
<th>Title</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mapping of web-accessibility on public websites</td>
<td>Danish Agency for Digitisation</td>
</tr>
<tr>
<td>2</td>
<td>Web Accessibility Observatory</td>
<td>ICT Directorate at Ministry of Finance and Public Administrations</td>
</tr>
<tr>
<td>3</td>
<td>Simply Accessible (Gewoon toegankelijk)</td>
<td>Quality Institute of the Dutch Municipalities (KING); The Association of Dutch Municipalities (VNG); The Ministry of the Interior and Kingdom Relations (BZK)</td>
</tr>
<tr>
<td>4</td>
<td>State of Affairs in the Kingdom (Status i riket)</td>
<td>Agency for Public Management and eGovernment (Difi)</td>
</tr>
<tr>
<td>5</td>
<td>Acessibilidade dos Conteúdos Web na Administração Pública Central (Web Content Accessibility on the Central Public Administration)</td>
<td>FCT Fundação para a Ciência e a Tecnologia / Unidade de Acesso</td>
</tr>
</tbody>
</table>

Organisations consulted for the user validation

<table>
<thead>
<tr>
<th>No. / Country</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>European Dyslexia Association (EDA)</td>
</tr>
<tr>
<td>2</td>
<td>Brain Injured &amp; Families - European Confederation (BIF-EC)</td>
</tr>
<tr>
<td>3</td>
<td>European Network of Independent Living (ENIL)</td>
</tr>
<tr>
<td>4</td>
<td>CBM Cognitive</td>
</tr>
<tr>
<td>5</td>
<td>Inclusion Europe</td>
</tr>
<tr>
<td>6</td>
<td>Autism Europe</td>
</tr>
<tr>
<td>7</td>
<td>European Blind Union (EBU)</td>
</tr>
<tr>
<td>8</td>
<td>European Deafblind Network (EDbN)</td>
</tr>
<tr>
<td>9</td>
<td>European Federation of Hard of Hearing (EFHOH)</td>
</tr>
</tbody>
</table>
Annex 2. Description of the ‘Method 1’ Norwegian methodology “Status i riket”

The methodology is designed to evaluate individual websites to find violations to the requirements in WCAG 2.0 on level AA with 3 exceptions (success criteria 1.2.3, 1.2.4 and 1.2.5).

The approach is to use a set of well-defined manual tests that include the use of some automatic tools to perform the tests. There is at least one, but often several tests for each success criteria in WCAG 2.0. Nowadays, there are over 50 different tests in the methodology, but these are constantly refined and developed. The only way to actually decide if a website do comply with WCAG 2.0 level AA is to actually do a full manual investigation based on the WCAG-EM. But this would require at least 50-60 hours per website if the result is to be documented in a way that makes it transparent and replicable. By applying all tests in the Norwegian methodology you get a pretty good picture of the accessibility level of the website in a shorter timeframe while maintaining transparency and replicability.

In Norway the methodology is not used to determine if a website complies with WCAG 2.0 level AA. Instead it is used to identify larger trends, such as if a specific sector does have bigger accessibility problems, or if there is a principle or success criteria in WCAG that websites violate more often than others. This data is then used to help organisations increase the accessibility of their websites. Because of this, the tests are used differently depending on the purpose of the exercise. To get a picture of the accessibility levels of a larger set of websites, a subset of tests is selected and used. This matches what was done in MEAC III, but on a larger scale (more tests). When looking at a smaller amount of websites all tests can be used to get a more detailed picture of the accessibility levels.

The method does not define neither the frequency of testing nor which websites are being tested.

The testing

The tests are based on manual testing, but in many tests automatic tools are used as support. Some examples of tools are:

- The W3C Mark-up Validation Service (https://validator.w3.org/)
- Web Accessibility Toolbar (https://www.paciellogroup.com/resources/wat/)
- Colour Contrast Analyser (https://www.paciellogroup.com/resources/contrastanalyser/)
The proposed tools can in many cases be replaced by other similar tools. New tools can also be developed to minimize the manual part of the testing.

**Step 1**

The scope is decided depending on the actual need at the moment. The scope includes which websites to test, and which tests are to be performed. This is not defined in the method itself.

**Step 2**

On each website a number of pages is selected. The selection is made manually to ensure that the pages selected are relevant and representative. The list of pages to select include among others:

- Start page
- 3 different article pages of the website.
- One page with a form including at least five different form objects (input fields, select lists, buttons and other form controls)

The method does not include instructions on how exactly the pages should be selected. This means that two different persons doing the testing will end up with a slightly different sample which in turn could lead to a result that might vary slightly. This was also shown in the validation exercise, but even so the validity of the result was better than for the other methods being used.

If this methodology is adopted for use in the EU, the selection of content to be assessed will need to be reworked. To include processes rather than pages, and to minimize the risk of different persons ending up with different content to test. When performed by one organisation in one country the selection process can be harmonised on the go, but if it is to be used in several countries this needs to be better defined.

**Step 3**

Each test is defined in detail by step–by-step instructions. The exact procedure depends on what is to be tested, but it roughly looks like this:

1. Selection of pages
   - What subset of pages from the selection is to be assessed on this test?

2. Check the number of occurrences
   - If the test is to determine if images do have text alternatives, this step would include several sub-tasks including, categorising images depending on whether they are purely decorative, are complex, are linked and so on, and counting the number of occurrences of images in each category.
3. Check for violations
   E.g. how many of the images that should have alternative text lack an alternative text that meets the requirements.

4. Calculate the result
   This could also be a set of tasks. If a complex image does lack a text alternative this could make the whole test fail, but if less important images lack an alternative text it might not be enough to fail the test.

Each test has different possible outcomes:
- 2 points: No problems
- 1 point: Only minor problems
- 0 points: Fail
- X: Not applicable or not possible to test.

Since the tests are very detailed the persons doing the actual testing do not need to be accessibility experts. Instead they only need a shorter training and understanding of the goal of the testing and support in the few occasions where it might be difficult to make a decision (e.g. is this an image of type A or type B).

**Using the results**

The methodology does not define how the results from each test are being used. It could be used to calculate a score, or a percentage. One example would be to look at the proportion of tests resulting in 0 points. If 0 % of the tests performed do result in 0 points, this indicates that the website might comply with WCAG 2.0 level AA. But if the result is 40 % it indicates that the website has large accessibility problems since 40 % of all tests are failed.

The overall result from each website can then be used to aggregate other data:
- Comparing websites in different sectors
- Comparing websites in the same sector but on only one or few success criteria
- Comparing websites on one specific principle in WCAG

An important aspect of the methodology is the constant updating of the tests. Since each test is described in such detail, the tests constantly need to be reviewed so that they include accessibility supported solutions. This is extra important on an EU level since it might differ between countries. A specific solution might be suitable in one European country but not appropriate in another country.
List of success criteria and tests covered

The data in the table below is based on how tests were designed in 2014. Since there has been changes so there might be differences in individual number of tests and tools being used. Also note that the tools included below can be substituted for other tools offering the same functionality.

<table>
<thead>
<tr>
<th>Success Criteria</th>
<th>Approximate number of tests</th>
<th>Automatic vs Manual</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>3</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>1.2.1</td>
<td>2</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>1.2.2</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>1.2.3</td>
<td>0</td>
<td>-</td>
<td>Not included in this methodology, this needs to be addressed on an EU level</td>
</tr>
<tr>
<td>1.2.4</td>
<td>0</td>
<td>-</td>
<td>Not included in this methodology, this needs to be addressed on an EU level</td>
</tr>
<tr>
<td>1.2.5</td>
<td>0</td>
<td>-</td>
<td>Not included in this methodology, this needs to be addressed on an EU level</td>
</tr>
<tr>
<td>1.3.1</td>
<td>9</td>
<td>Manual testing with some use of WAT and the developer tool in Internet Explorer</td>
<td></td>
</tr>
<tr>
<td>1.3.2</td>
<td>1</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>1.3.3</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>1.4.1</td>
<td>2</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>1.4.2</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>1.4.3</td>
<td>1</td>
<td>Manual testing with some use of WAT, Colour Contrast Analyser and A Ruler</td>
<td></td>
</tr>
<tr>
<td>1.4.4</td>
<td>1</td>
<td>Manual testing with some use of Developers tool in Windows</td>
<td></td>
</tr>
</tbody>
</table>

84 Web Accessibility Toolbar (https://www.paciellogroup.com/resources/wat/)

85 The “Developer Tools" in Internet Explorer (https://msdn.microsoft.com/library/hh968260(v=vs.85).aspx)

86 Colour Contrast Analyser (https://www.paciellogroup.com/resources/contrastanalyser/)

87 A Ruler for Windows (http://www.arulerforwindows.com/)
<table>
<thead>
<tr>
<th>Success Criteria</th>
<th>Approximate number of tests</th>
<th>Automatic vs Manual</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.5</td>
<td>1</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>2.1.1</td>
<td>3</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>2.1.2</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>2.2.1</td>
<td>1</td>
<td>Manual testing with some use of Developers tool in Windows</td>
<td></td>
</tr>
<tr>
<td>2.2.2</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>2.3.1</td>
<td>1</td>
<td>Manual testing with some help of Snagit</td>
<td></td>
</tr>
<tr>
<td>2.4.1</td>
<td>2</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>2.4.2</td>
<td>1</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>2.4.3</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>2.4.4</td>
<td>1</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>2.4.5</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>2.4.6</td>
<td>2</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>2.4.7</td>
<td>2</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>3.1.1</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>3.1.2</td>
<td>1</td>
<td>Manual testing with some use of Developers tool in Windows</td>
<td></td>
</tr>
<tr>
<td>3.2.1</td>
<td>2</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>3.2.2</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>3.2.3</td>
<td>3</td>
<td>Manual testing with some use of WAT</td>
<td></td>
</tr>
<tr>
<td>3.2.4</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>3.3.1</td>
<td>1</td>
<td>Manual testing with some use of Developers tool in Windows</td>
<td></td>
</tr>
<tr>
<td>3.3.2</td>
<td>2</td>
<td>Manual testing with some use of Developers tool in Windows</td>
<td></td>
</tr>
<tr>
<td>3.3.3</td>
<td>2</td>
<td>Manual testing</td>
<td></td>
</tr>
<tr>
<td>3.3.4</td>
<td>1</td>
<td>Manual testing</td>
<td></td>
</tr>
</tbody>
</table>

Snagit screen capturing tool ([https://www.techsmith.com/snagit.html](https://www.techsmith.com/snagit.html))
## Study on monitoring methodologies for web accessibility in the European Union

### Table

<table>
<thead>
<tr>
<th>Success Criteria</th>
<th>Approximate number of tests</th>
<th>Automatic vs Manual</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.1</td>
<td>2</td>
<td>Manual testing with some use of W3C Mark-up Validation Service[^89]</td>
<td></td>
</tr>
<tr>
<td>4.1.2</td>
<td>1</td>
<td>Manual testing with some use of Developers tool in Windows</td>
<td></td>
</tr>
</tbody>
</table>

[^89]: The W3C Markup Validation Service ([https://validator.w3.org/](https://validator.w3.org/))
Annex 3. Success criteria testing approach today and foreseeable in the future

This annex provides a table where the testing approach today and what testing approach may be foreseeable in the (near) future is specified for each success criterion.

The goal with this table is to show the complexity of testing compliance and the possibilities of different testing procedures per success criterion.

It is important to remember that every success criterion cover many different aspects and techniques. It is not possible to decide on one test that can measure all relevant aspects of each success criterion, since it takes several different tests to monitor compliance. For example, the first and probably best known success criterion in WCAG, 1.1.1 “Provide text-alternatives for any non-text content …” has more than 60 techniques that might be relevant to test in different situations. And these are just recommended techniques, there can be a wide range of solutions that is not included in that list, and furthermore, it is constantly changing.

The European monitoring methodology is going to test for compliance, which makes automatic tools impossible to use as the sole method. None of the automatic tools on the market are capable of checking for all success criteria in WCAG 2.0 AA. In fact, they do not check the whole success criterion, but only parts of them. Furthermore, the automatic tools are not reliable when it comes to accuracy. Rather, automatic tools should be seen as a support for the expert doing the evaluation.

This table goes beyond the validation exercise in this study and includes also open source and commercial tools on the market. Therefore, it is important to remember that this whole topic is changing rapidly. New tools can be developed and new techniques can be invented. Any monitoring methodology should cater for the possibility of flexibility, which is why the study team also recommend a constant evolution of the methodology.

As a way to structure the analysis, the testing approaches have been organised into five categories:

A: Can be fully automated
B: Mostly automated with some human support/decision making
C: It can, at some extend, be automatically tested, but it needs human evaluation
D: Human evaluation with some help of automatic tools
E: Can only be evaluated by humans

Notably, in a situation where most content on a webpage is dynamically introduced and altered, no success criteria can be automatically tested until we have solved how to do this without installing testing software on the server running the actual website. Therefore, the following table reflects the situation only for static content in a web environment.

Also note that many success criteria need to be tested in different devices, screen sizes and window sizes to determine how the content works in different situations. This is also hard and in some extent impossible to do automatically without human support, or done by a human with support of good automatic tools.

<table>
<thead>
<tr>
<th>Success Criteria</th>
<th>Testing today</th>
<th>What might be possible in the foreseeable future</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.1</td>
<td>D</td>
<td>C</td>
<td>To determine what non-text content is important and what text description serves the same purpose will not be possible to do fully automatically in the foreseeable future. Even though some development is being done in automatic captioning images, automatic tools cannot determine if an alternative text serves the exact same purpose as the non-text content.</td>
</tr>
<tr>
<td>1.2.1</td>
<td>E</td>
<td>D</td>
<td>Even though it is a straightforward success criterion, it takes a human to determine if the alternative is equivalent to the media content. With the development of AI (Artificial Intelligence) automatic tools might be able to identify instances of time-based media, and to do some automatic captioning and analysis of what is spoken, but it will not be possible to determine if a text serves the same purpose, or if the visual information in a video is covered in the text or captioning.</td>
</tr>
<tr>
<td>1.2.2</td>
<td>D</td>
<td>D</td>
<td>Even though it is a straightforward success criterion, it takes a human to determine if the alternative is equivalent to the media content. With the development of AI (Artificial Intelligence) automatic tools might be able to identify instances of time-based media, and to do some automatic captioning and analysis of what is spoken, but it will not be possible to determine if a text serves the same purpose, or if the visual information in a video is covered in the text or captioning.</td>
</tr>
<tr>
<td>Success Criteria</td>
<td>Testing today</td>
<td>What might be possible in the foreseeable future</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>1.2.3</td>
<td>E</td>
<td>D</td>
<td>Even though it is a straightforward success criterion, it takes a human to determine if the alternative is equivalent to the media content. With the development of AI (Artificial Intelligence) automatic tools might be able to identify instances of time-based media, and to do some auto captioning and analysis of what is spoken, but it will not be possible to determine if a text serves the same purpose, or if the visual information in a video is covered in the text or captioning.</td>
</tr>
<tr>
<td>1.2.4</td>
<td>D</td>
<td>C</td>
<td>Even though it is a straightforward success criterion, it takes a human to determine if the alternative is equivalent. In this case, since it is live media, the captioning in almost all cases would be closed making it easier to evaluate with an automatic tool. Still the quality must be evaluated by humans.</td>
</tr>
<tr>
<td>1.2.5</td>
<td>E</td>
<td>D</td>
<td>Even though it is a straightforward success criterion, it takes a human to determine if the alternative is equivalent. In this case the test must determine if the spoken text includes all visual information that is relevant, that will not be possible to do with an automatic tool in the foreseeable future.</td>
</tr>
<tr>
<td>1.3.1</td>
<td>C</td>
<td>C</td>
<td>This is a very complex success criterion with a lot of applications. Some of which can be automatically determined, but far from all. The success criterion includes that text that serves as headings visually also need to be identified with text or programmatically determined, that tables do have correct headings, that form fields and labels are associated and much more. Automatic tools can identify many different errors, but not verify compliance here.</td>
</tr>
<tr>
<td>1.3.2</td>
<td>D</td>
<td>D</td>
<td>What is to be considered a meaningful sequence can only be determined by a human.</td>
</tr>
<tr>
<td>1.3.3</td>
<td>D</td>
<td>C</td>
<td>There can be a wide range of types of instructions that fail this success criteria, and the technological development may lead to new types of possible problems that an automatic tool might miss.</td>
</tr>
<tr>
<td>1.4.1</td>
<td>D</td>
<td>C</td>
<td>Some issues can be found with automatic tools, and more can be done, but it is not feasible to rely only on automatic tools to identify all possible problems that for instance occur in diagrams and complex illustrations</td>
</tr>
<tr>
<td>1.4.2</td>
<td>D</td>
<td>B</td>
<td>This success criterion is not that complex, but it is difficult to analyse the controls on the page to ensure that they do control the audio and that they not only state that audio could be controlled with the control, but that it also actually works.</td>
</tr>
</tbody>
</table>
### Study on monitoring methodologies for web accessibility in the European Union

<table>
<thead>
<tr>
<th>Success Criteria</th>
<th>Testing today</th>
<th>What might be possible in the foreseeable future</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.4.3</td>
<td>D</td>
<td>B</td>
<td>Contrast can be measured automatically but contrast of text with image backgrounds is difficult, especially if the text is part of the image or the placement of the text on the background image is shifting in different screen sizes. We think that more can be done to test this automatically but that it won’t be possible to do entirely automated.</td>
</tr>
<tr>
<td>1.4.4</td>
<td>D</td>
<td>D</td>
<td>Even though this to some extent can be automatically tested, it is impossible to test 100 percent that no problems occur when users zoom and enlarge text on different types of devices. Note that this means testing how the interface behaves in different devices in different user situations, some examples are desktop, mobile, smart watches, with maximised or not maximised applications windows.</td>
</tr>
<tr>
<td>1.4.5</td>
<td>D</td>
<td>B</td>
<td>Automatic tools can identify possible situations where text might be part of an image, but it is not probable that they can determine if it is essential or not. Today we see that this is done quite seldom but it wouldn’t be impossible to improve the automatic testing quite significantly with AI.</td>
</tr>
<tr>
<td>2.1.1</td>
<td>D</td>
<td>C</td>
<td>A tool might identify suspects, situations that might be violations, but only a human can determine if they are situations that create problems for users. More can be done automatically than is being done today. The code could be analysed to find typical problems but in the end it takes a human to determine if everything can be done with a keyboard.</td>
</tr>
<tr>
<td>2.1.2</td>
<td>E</td>
<td>C</td>
<td>Tabindex loops could be found with an automatic tool but script generated problems will probably not be possible to identify without human testing.</td>
</tr>
<tr>
<td>2.2.1</td>
<td>E</td>
<td>C</td>
<td>Time limits often occur on the server and to test this with an automatic tool is difficult. Furthermore, it takes a human to determine if it is an essential exception. It would be possible to construct a tool that would log in (if a log in is required and the log in details are known) and then observe changes on the screen waiting for a change (log out) to occur. But then a human need to evaluate the results.</td>
</tr>
<tr>
<td>2.2.2</td>
<td>D</td>
<td>C</td>
<td>It is not possible to identify all such situations with an automatic tool and an automatic tool will have problems verifying that if the users pushes a “stop”/“pause” button, that it actually do stop the moving/blinking/scrolling/auto-updated content.</td>
</tr>
<tr>
<td>Success Criteria</td>
<td>Testing today</td>
<td>What might be possible in the foreseeable future</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>2.3.1</td>
<td>D</td>
<td>C</td>
<td>This success criterion is quite straight forward and there are already tools that can measure for flashes in videos. But it is hard to analyse flashing animations and background images with an automatic tool and decide if the problem is big enough in different screen sizes. Also remember that this needs to be measured in different screen sizes since the portion of the screen affected and the viewing distance is relevant for the outcome of the testing.</td>
</tr>
<tr>
<td>2.4.1</td>
<td>D</td>
<td>C</td>
<td>For ordinary webpages this might be solved automatically but it takes a human to identify the need for these types of mechanisms in e-services or websites not following the most common concepts</td>
</tr>
<tr>
<td>2.4.2</td>
<td>C</td>
<td>B</td>
<td>It possible that automatic tools in a future can do an educated guess whether a page title is relevant or not, but the ultimate decision needs to be done by a human. With AI more can be done automatically than is done today.</td>
</tr>
<tr>
<td>2.4.3</td>
<td>E</td>
<td>D</td>
<td>It is difficult to test this automatically since meaning, and what is logical can be very different in different situations. But it might be possible to find some suspected errors with automatic tools in the future, but it will still need human evaluation.</td>
</tr>
<tr>
<td>2.4.4</td>
<td>D</td>
<td>C</td>
<td>This success criterion is hard to evaluate automatically, it is easy to test for common failures, but it takes a human to actually verify that a links purpose is clear from the link text in its context. With AI more can be done, but this needs to be verified by humans.</td>
</tr>
<tr>
<td>2.4.5</td>
<td>E</td>
<td>B</td>
<td>This success criteria could be met in several different ways, but it is often fulfilled by providing a search function and a menu. This could be verified automatically but when one of these two things is missing, it takes a human to verify that there is another solution that serves the same purpose.</td>
</tr>
<tr>
<td>2.4.6</td>
<td>E</td>
<td>D</td>
<td>To determine if a heading or label exists you can use automatic tools, but there is no possibility to automatically determine if that heading or label actually do describe the topic or purpose. With AI more can be done, but this needs to be verified by humans.</td>
</tr>
<tr>
<td>2.4.7</td>
<td>E</td>
<td>C</td>
<td>An automatic tool could probably identify if focusable objects have some kind of focus highlighting, but will have trouble if it is done with script, and will have trouble to determine if it is easy for the user to see.</td>
</tr>
<tr>
<td>3.1.1</td>
<td>B</td>
<td>A</td>
<td>This would be possible with a good automatic tool. This requires a tool not only to look for language codes, but also to verify that the language actually is the one specified.</td>
</tr>
<tr>
<td>Success Criteria</td>
<td>Testing today</td>
<td>What might be possible in the foreseeable future</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>3.1.2 D A</td>
<td>This would be possible with a good automatic tool. This requires a tool that can find shorter phrases on other languages than the main language, identifying the actual language used and verifying that the code reflects this.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2.1 D C</td>
<td>It is easy to find suspected problems with automatic tools but it is hard to find every type of possible problem, and it is impossible to automatically decide if a potential problem is a real problem for users. More can probably be done automatically than is being done today, but far from everything.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2.2 D C</td>
<td>It is easy to find suspected problems with automatic tools but it is hard to find every type of possible problem, and it is impossible to automatically decide if a potential problem is a real problem for users. More can probably be done automatically than is being done today, but far from everything.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2.3 E B</td>
<td>This can be done in a large extent with automatic tools, but it is hard to identify if a user interaction that leads to a change is logical, so there will always be a need for some human involvement. This is a success criterion where we haven’t seen any automatic tools being used, there might be, but it seems that this is an area that seems to be open for possible improvement.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.2.4 D C</td>
<td>This can be done automatically to some extent. One example would be a Print link that should be consistent identified on different pages. But an automatic tool cannot always understand if the same function in different contexts should be considered to have the exact same functionality. A user might not perceive it as the same thing so a human must decide if it is appropriate with the same identification or not.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3.1 E E</td>
<td>Since automatic tools cannot test forms, posting them with all possible different error types and evaluate the response without installations on the servers, this success criteria cannot be tested automatic for monitoring purposes.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3.2 E D</td>
<td>To determine if a label exist you can use automatic tools, but there is no possibility to automatically determine if that label actually do describe what the user needs to do. More can be done with AI, but only as support for human evaluations.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.3.3 E E</td>
<td>Since automatic tools cannot test forms, posting them with all possible different error types and evaluate the response without installations on the servers, this success criteria cannot be tested automatic for monitoring purposes.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Success Criteria</td>
<td>Testing today</td>
<td>What might be possible in the foreseeable future</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>-------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>3.3.4</td>
<td>E</td>
<td>E</td>
<td>Since automatic tools cannot post forms and go through the entire process this will not be possible without installing a tool on the server, which is not possible when monitoring websites.</td>
</tr>
<tr>
<td>4.1.1</td>
<td>D</td>
<td>C</td>
<td>Even though most code can be automatically tested, it is not possible to design a tool that can extract all possible code variations from a webpage when content is dynamically introduced, hidden and removed.</td>
</tr>
<tr>
<td>4.1.2</td>
<td>D</td>
<td>C</td>
<td>This is a very complex success criterion that covers a lot of different situations depending on the solution and choice of technologies. Some issues and possible problems can be automatically identified but to decide if the interface is good enough for the user requires manual testing. More can be done with AI, but a human needs to be involved to verify compliance.</td>
</tr>
</tbody>
</table>