

**Specification and Implementation of the experimental
environment**

Deliverable Number: 4.1.1



Version: FINAL

Date: 2006-02-08

Authors: Jenny Craven, Peter Brophy: Sections 1, 4 and 5

Boleslaw Szomanski: Sections 2 and 3

Dissemination Level: Internal

Status: FINAL

This document consists of 100 pages plus this cover

Version Control

<i>Version</i>	<i>Status</i>	<i>Date</i>	<i>Change</i>	<i>Author</i>
0.1	DRAFT	2005-06-06	First draft	Jenny Craven
	DRAFT		Comments and input to Sections 3, 4 and 5.	Pat Bertini
0.2	DRAFT	2005-06-08	Second draft	Jenny Craven
		2005-06-08/09/10	Input to examples in Sections 3, 4, and 5	Pat Bertini
		2005-06-10	Input to examples in Sections 3,4, and 5	Jenny Craven
		2005-06-14	FTB comments	Olaf Perlick, Helmut Heck
		2005-06-29	Examples of Polish websites	Jacek Chmielewski
0.3	DRAFT	2005-07-14	Amendments relating to comments	Jenny Craven
0.4	DRAFT	2005-10-04	Major amendments following the User Testing Plan discussion document	Jenny Craven, Pat Bertini, Helmut Heck, Mikael Snaprud, Justina Kieliszczyk, Boleslaw Szomanski.
0.5	DRAFT	2005-10-10	Combining of User Testing Plan to D4.1.1	Jenny Craven
0.5		2005-10-14	Amended according to updated Checkpoints and to updated UWEM	Jenny Craven, using data provided by Olaf Perlick, and the UWEM 0.5 published on the Cluster web site.
0.6	DRAFT	2005-11-07	Input to section 2 and 3 from TUW	Boleslaw Szomanski
	DRAFT	2005-11-07	Proof reading and preparation for release of Draft 6 for comments	Jenny Craven
0.7	DRAFT	2005-11-18	Correction to English language in sections 2 and 3; other minor amendments	Peter Brophy
0,75	DRAFT	2005-11-20	Correction after review	Boleslaw Szomanski, Magda Mazur
0,8	DRAFT	2005-11-23	Structure correction	Boleslaw Szomanski
0.8	DRAFT	2005-11-24	Proof reading of sections 2 & 3. Updating of sections 5, 6, & 7.	Jenny Craven and Peter Brophy
0,82	Draft	2005-11-27	Changing section 3	Boleslaw

				Szomanski, Lukasz Forytek
0,85	Draft	2005-11-28	Adding changes after teleconference	Boleslaw Szomanski
1.9	DRAFT	2005-11-29	Final editing before release as final draft for comment to the EIAO consortium	Peter Brophy
		2005-12-20	Amendments to sections 4 and 5 according to inspection comments.	Jenny Craven
1.9	DRAFT	2006-01-05	Revised versions 2 and 3	Boleslaw Szomanski
FINAL		2006-02-08		Jenny Craven

1. Table of Contents

1	Introduction.....	5
1.1	Brief project description.....	5
1.2	Scope of WP4.....	5
1.3	Scope of D4.1.1.....	6
1.4	User testing undertaken by Cluster projects.....	9
1.4.1	Support EAM:.....	9
1.4.2	Ben-To-Web:.....	9
1.5	How this work relates to EIAO and the Cluster.....	10
2	The background of UWEM Testing	11
2.1	Introduction	11
2.2	Background.....	11
2.3	People involved in evaluation process	12
2.4	Process of evaluation	13
3	UWEM Testing: specification and implementation.....	20
3.1	Introduction.....	20
3.2	Initial evaluation of the methodologies (Stage 1).....	20
3.2.1	Criteria for the initial evaluation and selection.....	20
3.2.2	Collection of the methodologies for preliminary evaluation	22
3.2.3	Evaluation of collected methodologies.....	28
3.2.4	Selection of methodologies	34
3.3	Testing of the methodologies for the evaluation of the web accessibility (stage 2)...	38
3.3.1	Introduction.....	38
3.3.2	Preparing criteria and documentation for testing.....	38
3.3.3	Evaluation criteria	39
3.3.4	Test result form.....	44
3.3.5	Method for the methodologies comparison.....	51
3.3.6	Preparation of the user testing	52
3.3.7	Performing the user testing	53
3.3.8	Evaluation of the tested methodologies.....	54
3.3.9	Final results of the methodologies evaluation.....	55
3.4	Pilot testing.....	56
3.4.1	Aim of the pilot testing.....	56
3.4.2	Planing pilot testing.....	56
3.4.3	Running pilot testing.....	56
3.4.4	Check pilot testing results.....	56
3.4.5	Improvement of test forms and criteria weighs.....	56
4	User testing of WAM relevance: background.....	57
4.1	Introduction.....	57
4.2	Cluster involvement.....	58
4.3	Background to methods for testing.....	58
4.3.1	Test Suites.....	59
4.3.2	User Test Site.....	60
4.3.3	User testing protocols.....	61
5	Testing of WAM relevance: specification and implementation.....	63
5.1	Purpose of the user testing.....	63
5.2	Development of tasks and questions.....	67
5.2.1	Tasks.....	67

- 5.2.2 Questions.....68
- 5.3 Enlisting of participants..... 69
- 5.4 Pilot testing..... 70
- 5.5 Running the tests.....71
- 5.6 Data analysis and report writing..... 73
- 5.7 Result of initial pilot testing..... 75
- 6 Appendix One: Websites for testing of WAM Relevance..... 77
- 7 Appendix Two: Tasks for Testing of WAM Relevance.....82
 - 7.1 Checkpoint 1.1.....82
 - 7.2 Checkpoint 3.4.....84
 - 7.3 Checkpoint 3.5.....85
 - 7.4 Checkpoint 5.5.....87
 - 7.5 Checkpoint 5.6.....89
 - 7.6 Checkpoint 9.4.....91
 - 7.7 Checkpoint 12.1.....93
 - 7.8 Checkpoint 12.4.....95
 - 7.9 Checkpoint 13.1.....97

1 Introduction

1.1 Brief project description

The goal of the European Internet Accessibility Observatory project is to contribute to better e-accessibility for all citizens and to increase the use of standards for online resources.

The project will establish the technical basis for a possible European Internet Accessibility Observatory (EIAO) consisting of:

- A set of web accessibility metrics.
- An Internet robot for automatically and frequent collecting data on web accessibility and deviations from web standards (the WAI guidelines).
- A data warehouse providing online access to collected accessibility data.

The collection of web accessibility metrics and the tools for automated data collection and dissemination will be continuously improved by feedback from end users and user testing to sharpen the relevance of the automatically collected data.

This deliverable will describe methods for conducting user testing of the Unified Web Evaluation Methodology (UWEM) and user testing to help verify the Web Accessibility Metrics (WAMs). The deliverable will address Task 4.2 of WP4 (it should be noted that Task 4.1 has now been merged with Task 4.2): Implementation of the experiment environment, and will be used to inform Task 4.3: Run the experiments, evaluate and analyse the outcomes (Deliverable 4.3.1: Experimental results, evaluation and analysis, is due project month 18).

The EIAO project is being carried out as part of the Web Accessibility Benchmarking cluster together with two other EC funded projects: Support-EAM and BenToWeb,¹ and in co-ordination with the World Wide Web Consortium Web Accessibility Initiative (W3C/WAI) to develop a Unified Web Evaluation Methodology (UWEM)².

1.2 Scope of WP4

Four main areas of user testing for the EIAO project have been identified:

¹ Details of the Cluster projects are available at: www.wabcluster.org/

² <http://www.wabcluster.org/uwem05/>

1. User testing of the UWEM methods.
2. User testing of WAM relevance.
3. Usability testing of the Datawarehouse.
4. User testing of the Observatory content (this comes under WP2).

The main objective of WP4 is to perform user testing to help verify the Web Accessibility Metrics (WAM) model and to test the Observatory Datawarehouse user interface. The specification and implementation of the first two test areas above are covered in this document (see below). The usability testing of the Datawarehouse will be covered in the next iteration of this document (due Month 22) and the user testing of the Observatory content will be covered in the second iteration of D2.1.2 (due Month 20).

WP4 comprises the following tasks:

- Implementation of the experimental environment
- Run experiments, evaluate and analyse the outcomes.
- Validation of UWEM Web Accessibility Metrics
- Test of data warehouse user interface.

The work is divided into three deliverables:

- D4.1.1: Specification of the experimental environment.
- D4.3.1 Experimental results, evaluation and analysis (two iterations).
- D4.3.2 Final report on the findings and scientific results (co-ordinated or merged with D4.3.1).

Work is conducted by MMU, TUW, AUC, NK, FTB, FBL, AAU.

The user testing experiments are conducted by MMU and TUW.

1.3 Scope of D4.1.1

Deliverable D4.1.1 reports on the specification and implementation of the experimental environment for WP4.

The main tasks for D4.1.1 are:

- Design and implement the test environment.
- Setup and test the software and test environment.
- Undertake expert walkthroughs of the setup environment.
- Undertake pilot tests with users; revise any features as necessary.
- Distribute test environment.
- Hold meetings with partners who will be running the tests.
- Writing of D4.1.1.

Generic methods for user testing should include the following steps:

- Specification of the experimental environment.
- Implementation of the experimental environment.
- Enlist participants (ongoing from start of process).
- Pilot testing of the experimental environment (adjustments made where necessary).
- Run sessions - either remotely or face-to-face depending on circumstances and resources.
- Data analysis and report writing.

Whilst each area may require the development of a separate set of tests/scenarios and users, questions should be developed in a way to allow for cross comparison.

Classification of people with access problems has been described in WP2: D2.1.3 and also in the UWEM document (Section 7.3). Using these documents and other relevant work identified³ (see also D2.1.3), the following people who may experience access problems have been identified as including:

- Functional blindness: People who are blind (either totally blind or no useful sight) who need to use screen reading technology or refreshable Braille to access the Web.

³For examples see: <http://www.w3.org/WAI/EO/Drafts/PWD-Use-Web/#diff>
Centre for HCI Design. 2003. *Usability Studies – JISC Services and Information Environment*, May.
Disability Rights Commission (DRC). 2004 *The Web: Access and Inclusion for Disabled people. A formal Investigation conducted by the Disability Rights Commission*. London: DRC.
Forrester Research. 2003 *The wide range of abilities and its impact on computer technology. A research study commissioned by Microsoft Corporation and conducted by Forrester Research Inc.*

- Partial sight: People with a visual impairment who need to use screen magnification or screen enlargements/adjustments to access the Web.
- Dyslexia or literacy problems: People with a learning difficulty such as dyslexia who need to adjust the screen or text, or who use screen reading aids to be able access the Web.
- Physical disabilities: People who have a physical impairment which does not allow them to use the mouse, or who need to use assistive technologies such as joysticks or speech input to access the Web.
- Deafness or hard of hearing: People who have a hearing impairment and need to have any audio or sound captioned or described in text.

Participants will be selected on a basis of the following:

- Disability.
- Use of assistive technologies.
- IT experience.

Taking the above categories into account and to be in accordance with the UWEM (7.3), at least 3 users per disability group should be represented with each group comprising a mix of 'semi-novice' and 'expert' users. Disability groups selected to undertake testing should also reflect the aims of the user testing. A control group of non-disabled users should also be included.

Mobile phone users (and users of other devices with limited display capabilities) will not be included in the first tests as this is out of the scope of the project. However, technically it could be an important part of what we can achieve with the Observatory (so may be considered in later tests).

User testing will be undertaken using a combination of remote and face-to-face testing methods depending on participants' circumstances and resources available within the project (both time and budget related).

1.4 User testing undertaken by Cluster projects

1.4.1 Support EAM:

S-EAM have developed a questionnaire to be used to evaluate the UWEM. It is currently available on the WAB Cluster website at: <http://www.wabcluster.org/uwem05/>

Other methods for UWEM feedback include a discussion list and a FAQ page.

1.4.2 Ben-To-Web:

Ben-to-web have developed an eight-step framework for user-based evaluations which EIAO could draw on, also using the UWEM and other related documents.

Ben-to-Web will be using a number of User Panels for the project work. According to the Ben-To-Web Deliverable D3.1 (Public) the following User Panels will be required:

- Main user panel (disabled people).
- Older User Panel.
- Colour Vision Deficiencies Panel.
- Web developers.
- Web commissioners/owners.

A recruitment questionnaire has been developed to be used by partner organizations who have links with potential participants to directly recruit people (ISdAC is one example – users are paid for their participation). The project will create links to other User Panels to recruit sufficient people.

The framework for the user-based evaluations will be mostly face-to-face involving disabled participants and web developers. In some cases remote evaluations will be conducted. The framework will follow eight steps:

- Identify the purposes of the evaluation.
- Identify tasks and build scenarios.
- Define what needs to be measured.
- Develop the overall protocol for the evaluation sessions.
- Pilot the protocol.

- Recruit participants.
- Run sessions.
- Analyse data.

(see D3.1 available at www.bentoweb.org)

1.5 How this work relates to EIAO and the Cluster

The work of WP4 will be in co-operation with WP3 to verify the models in WP3 and the user testing protocol methods of the UWEM. Results from WP2 will also be used to inform the selection of checkpoints for testing. This document is related to the following documents:

- D-WAB2 Unified Web Evaluation Methodology (UWEM 0.5).
- D2.1.4 Report on primary user requirements.
- D3.1.1 First Version of ROBACC WAMS.
- D4.3.1 Experimental results, evaluation and analysis.

Methods describe in Sections 2, 3, 4 and 5 of this document, and the results from D4.3.1 can be used by the Cluster to inform Section 7: User Testing Protocol of the UWEM and to some extent inform Section 6.4: the UWEM user centric accessibility barrier model of the UWEM.

The following sections report on the specification and implementation of two areas of user testing:

- Sections 2 and 3: User Testing of UWEM Methodology. TUW reports on the specification and implementation and will be undertaking and reporting the user testing results of this area in D4.3.1.
- Sections 4 and 5: User testing of WAM relevance. MMU reports on the on the specification and implementation and will be undertaking and reporting the user testing results of this area in D4.3.1.

2 The background of UWEM Testing .

2.1 Introduction

This section describes the scope of the evaluation of methodologies for user testing of the accessibility of web pages with disabled people.

Section 2.2. introduces proposed primary foundations of the methodology evaluation and the resources used. Section 2.3. defines people that are included in UWEM evaluation. Section 2.4 illustrates a proposed 2 tier approach. In section 2.5 describes responsibilities of people that are included in this process.

2.2 Background

All methodologies for testing the accessibility of web pages with disabled people are potentially in scope for this work. However, the primary purpose of the testing is to evaluate the UWEM methodology against other methodologies. Evaluation will consider only user testing, and no other components of methodologies. It will involve the preparation of test scenarios, procedures and sets of criteria, test execution, the evaluation of tested web pages using the selected methodologies and finally an evaluation of the applied methodologies.

Selection of the evaluation methodologies to test against UWEM and testing itself will be a two-stage process:

Stage 1: The collection of all methodologies known by EIAO and Cluster partners. Those methodologies will be evaluated with regard to their use of user-based testing, the costs of using them, and their requirements for sample groups, etc. Evaluation will be carried out by expert testers based on agreed selection rules and criteria (see 3.2 below). During this phase all relevant methodologies are compared analytically, using expert judgement and a set of relevant criteria (the criteria for this analysis is defined and documented in the criteria subsection, 3.2.1). As a result we identify 1 or 2 methodologies beyond UWEM that will be tested practically. Chosen methodologies will be introduced in detail at point 3.2.4. The process run of the initial evaluation is introduced in Tab.1 in section 2.4.

Stage 2: Testing the UWEM methodology against the other selected methodologies. The testing procedure is presented in section 3 and consists of the:

- Phase 1: preparation of the criteria phase (3.3.2). In this phase the criteria for methodology evaluation (3.3.3.), the form of collected data (3.3.4.) and method for comparing tested methodologies (3.3.5.) were elaborated and introduced.
- Phase 2: preparation phase. This phase precedes a main process of testing and it includes organizational and technical activities related with testing preparation, activities such as selection of people for test execution and qualification of web pages for testing (3.3.6).
- Phase 3: a practical phase. UWEM and the 2 other methodologies are tested practically – they are actually used by test engineers to design, run and evaluate end-user test results (sections 3.3.7 and 3.3.8).
- Phase 4: consists of final evaluation by expert testers, using materials from testing and comments from user-testers and test-engineers (section 3.3.9).

The process is graphically shown in section 2.4, and clearly described in section 3.3. Before starting the main process of testing it is advisable to run the pilot testing during which most of the documents and forms used for testing will be checked and problems necessitating more careful preparation will be indicated. The assumptions of the pilot testing will be introduced in section 3.4.

2.3 People involved in evaluation process

The following people will take part in the tests: expert testers, test engineers, user-testers. Expert -testers are people having required knowledge and experience in evaluating the methodology. They have experience in organizing the testing process, and conducting testing of many applications and IT systems . They are acquainted with quality issues, experienced with web page designing or evaluation methodologies and they own PhD degree or ISEB certificates. Expert-testers' responsibilities in the project are supervision over the methodologies testing process and methodologies final evaluation.

At least four expert testers will be responsible for: selection of the methodology to test (procedure step 1), training of test engineers, evaluation of test scenarios prepared by test engineer and their possible improvement, selection of web pages to test, comparison of test results and making the final conclusions.

The test engineer is a person responsible for the preparation, supervision and analysis of results of the software/web pages testing. Test engineers' objectives in the project are leading the test, to familiarize themselves with the particular methodologies, prepare test scenarios, test plans and if necessary questionnaires. The test engineer will also assign the execution of the test to users, the collection of results and drawing up conclusions regarding web page evaluation and the evaluation of the proposed methodologies.

The user-tester is a person who actually tests the web page i.e. carries out the test tasks. The term 'user' means a disabled person identified in the groups defined in section 1.3. The user tester will take part in the testing and evaluation from the perspective of an end-user. Users will be chosen from a group of 45 people (the number of the users depends on methodology).

- The Polish Association of the Blind. Adult blind people including 2 blind computer scientists engaged in designing web pages. Group of 10 participants, some of whom can do the test at home (people from outside Warsaw); for others testers will attend their home/place of work for the test.
- School for partially sighted. Warsaw Koźmińska street. Group of 10 participants with poor vision with age 15 – 18 years old, plus 2 computer science teachers. The school has developed a detailed classification of eye illnesses that will enable characterization of test participants in the next stage. Testing will be performed in the school building.
- Center for the blind and partially sighted in Laski. Group of 10 participants (blind and with poor vision) in age range 15 – 20 years. Testing will be performed in a computer room with the assistance of 2 teachers.
- Foundation for disabled computer scientists. 15 participants with poor vision, deaf and physically disabled. Testing will be performed at home; test engineers will attend for the tests. This group includes dyslexic students from among Warsaw University of Technology students and other people.

2.4 Process of evaluation

As it was mentioned in section 2.2. the evaluation process is a complex organizational process that includes two steps:

Step 1 – the process that includes initial evaluation and selection for further testing is described in Table 1. Beside the particular action stages, reference to sections are included where details are discussed relating to particular activities, people taking part in the evaluation and initial documents.

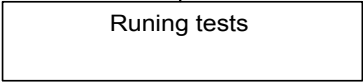
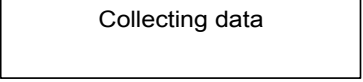
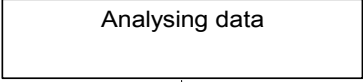
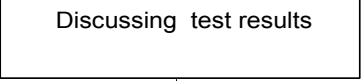
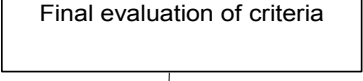
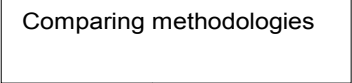
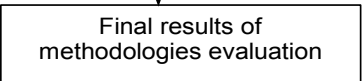
Step 2 – the execution of user testing and final evaluation was introduced in Table 2. Similar as in Tab.1 beside the particular action stages, people taking part in testing and evaluation, and initial documents are introduced.

Table 1. Process of initial evaluation and selection

Stage	Step	Participants	Link	Output
1	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Developing initial evaluation criteria</div> <div style="text-align: center; margin-top: 10px;">↓</div>	Expert testers	Section 3.2.1	Initial Evaluation criteria
1	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Collecting all possible methodologies</div> <div style="text-align: center; margin-top: 10px;">↓</div>	All participant in the web cluster and expert testers	Section 3.2.2	Short description off collected methodologies
1	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Evaluating collected methodologies</div> <div style="text-align: center; margin-top: 10px;">↓</div>	Expert testers	Section 3.2.3	Report of evaluation methodologies
1	<div style="border: 1px solid black; padding: 5px; width: fit-content; margin: 0 auto;">Selection of methodologies</div>	Expert testers	Section 3.2.4	Detail description of selected methodologies

Table 2. Testing of the methodologies for the evaluation of the web accessibility

Stage	Step	Participants	Link	Output
2	<div style="border: 1px solid black; padding: 5px; text-align: center;">Preparing evaluation criteria</div>	Expert testers	Section 3.3.3	Evaluation criteria
2	<div style="border: 1px solid black; padding: 5px; text-align: center;">Preparing test result form</div>	Expert testers	Section 3.3.4	Test results form
2	<div style="border: 1px solid black; padding: 5px; text-align: center;">Preparing method for the methodologies comparison</div>	Expert testers	Section 3.3.5	Method for the methodologies comparison
2	<div style="border: 1px solid black; padding: 5px; text-align: center;">Collecting user testers</div>	Expert testers and User testers	Section 3.3.6	User testers
2	<div style="border: 1px solid black; padding: 5px; text-align: center;">Selecting test engineers</div>	Expert testers and test engineers	Section 3.3.6	Test engineer
2	<div style="border: 1px solid black; padding: 5px; text-align: center;">Training test engineers</div>	Expert testers and test engineers	Section 3.3.6	Report of training

2		Test engineers and user testers	Section 3.3.7	User opinion of test, record of test
2		Test engineers	Section 3.3.7	Test form from test
2		Expert testers	Section 3.3.8	Draft result of evaluation for each methodologies
2		Expert testers and test engineers	Section 3.3.8	Protocol of discussion, opinion of test engineers
2		Expert testers	Section 3.3.8	Final evaluation for criteria
2		Expert testers	Section 3.3.8	Results of comparing methodologies
2		Expert testers	Section 3.3.9	Report of evaluation methodologies

Responsibilities of test participants

Responsibilities for particular phases of testing in the second stage are summarized in the following table 3.

Table 3. Responsibilities for each action in user testing

Stage	Expert-testers	Test engineers	User-tester
Stage 1	<ul style="list-style-type: none"> • Preparing criteria for initial evaluation and selection of the methodologies • The methodologies collecting • Initial analysis of the methodology • Definition of precise domain of use of the methodology • Initial evaluation of methodologies • Selection of methodologies 		
Stage 2 preparin g criteria phase	<ul style="list-style-type: none"> • Preparing criteria, form and method for methodologies evaluations 		
Stage 2 preparin g test phase	<ul style="list-style-type: none"> • Selection of a user-testers group • Selection and training of test-engineers to undertake evaluation • Selection of web pages 		

Stage	Expert-testers	Test engineers	User-tester
Stage 2 test phase		<ul style="list-style-type: none"> • Designing test cases and test scenario to perform the end-user test of a web page. • Management of the user-testers group • Collection of test results • Documentation of test results (test form introduced in section 3.3.4) • Addition of comments referring the methodology, questionnaires etc. 	<ul style="list-style-type: none"> • Performing the test according to the plan developed by the test engineer • Addition of comments referring the web pages • Addition of comments referring to the methodology used
Stage 2 evaluation Phase	<p>Comparison of test reports from different user-testers and different methodologies.</p> <ul style="list-style-type: none"> • Evaluation of each methodology according to the criteria defined in section 3.3.3. • Analysis of comments from test-engineers and user-testers • Preparation of the draft report and its presentation at the experts' meeting • Correction of the report after experts' meeting • Preparation of summarized comments from user-testers and test engineers • Preparation of formal evaluation of methodologies with usage of the method described in section 3.3.5 	<ul style="list-style-type: none"> • Participation in the meeting with expert testers • Addition of comments to the draft report • Preparation of tested web pages list and its source codes 	

Stage	Expert-testers	Test engineers	User-tester
	<ul style="list-style-type: none"> • Addition of remarks showing desirable UWEM methodology improvements • Preparation of final version of evaluation report • Sending final report to WP4 leader (MMU) 		

3 UWEM Testing: specification and implementation

3.1 Introduction

Section 3.2. introduces the first stage of the evaluation process, that is the initial evaluation of the methodologies for evaluating the accessibility of web pages accessibility for disabled people. First in section 3.2.1. criteria is presented for initial evaluation and selection of methodologies. In section 3.2.2. a short description of the candidate methodologies is presented. In section 3.2.3. results of the evaluation according to the criteria defined in section 3.2.1 are presented. In section 3.2.4. selected methodologies are presented. This completes procedure stage 1 as described in table 1.

Section 3.3. - process stage 2 shows how evaluation of selected methodologies in stage 1 will be performed. First step prepares the criteria and the form for the evaluation (section 3.3.2) Section 3.3.3. presents the criteria for the evaluation of the selected methodologies by expert testers. Section 3.3.4 provides the form to be used by the test engineers for the evaluation of each methodology. Section 3.3.5. introduces the formal model for methodologies comparison. Section 3.3.6. describes the preparation phase of the user testing. Section 3.3.7. provides a detailed procedural description of the methodology for the user testing. Section 3.3.8. describes how evaluation of methodologies are performed. How the final report will be prepared is presented in section 3.3.9. Procedural approach of the stage 2 is presented in table 2.

Section 3.4. describes the idea of pilot testing needed to improve the proposal documents and procedures.

3.2. Initial evaluation of the methodologies (Stage 1)

3.2.1 Criteria for the initial evaluation and selection

The following criteria will be applied for methodology evaluation:

- availability of a published description of the methodologies or possibility of contact with their authors to obtain full details
- usage of user testing in the methodology
- use or partial use of the methodology for usability testing by disabled people

- no requirement for special software usage connected with this methodology or incurring license fees for its usage
- estimated costs and work consumption of user testing in this methodology: excessive costs disqualify the methodology
- significantly different approaches to user testing compared with the UWEM methodology. Greatly different methodologies will be taken into consideration for the purpose of their possible advantages in the next version of UWEM.

Methodologies relying on verification exclusively by software will not be taken into consideration for the next stage. Methodologies based on special-purpose browsers for blind people will not be taken into consideration in the next stage. Methodologies mostly intended for all-purpose web pages evaluation and only in a minor part for disabled people will also not be taken under consideration in the next stage.

A detailed description of initial evaluation criteria is presented in table 4.

Table 4 Criteria for initial evaluation

Criterion	Description	Input	Output
Access	<i>published description of the methodologies or possibility of contact with their authors</i>	Text of methodologies or contact with author	Yes if access No if no access or methodology need to be translated
User testing	<i>usage of user testing in the methodology</i>	Use user testing	Yes if uses No if not
For people with disabilities	<i>use or partial use of the methodology for usability testing by disabled people</i>	Part of methodology for usability people	All – specially for disability people More – most part Little – only small participant No – any part
Not require special software or purchase license	<i>no requirement for special software usage connected with this methodology or incurring license fees for its usage</i>	No need Special software or fees for license	Ok – if no need Require – if needed
estimated costs	<i>estimated costs of user testing in this methodology</i>	Cost for user testing	Small Medium High

Criterion	Description	Input	Output
Work consumption	<i>estimated work consumption of user testing in this methodology</i>	Work consumption for user testing	2-3 people 4-6 people 7- 10 people more than 10 people
Need for modification	Need for modification to user testing, preparing forms, questionnaires etc.	Time needed for modification	No time little time much time

After completing all criteria expert testers will decide if the methodology is useful for the user testing stated on 1-7 scale:

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

3.2.2 Collection of the methodologies for preliminary evaluation

Methodologies have been collected based on suggestions of EIAO project members and cluster partners. A short description and analysis of the collected methodologies is presented in table 5.

Table 5. Short description of selected methodologies

Name of methodology	"See it right" accessible website audits
Author of methodology	RNIB
Link	http://www.rnib.org.uk/xpedio/groups/public/documents/publicwebsite/public_seeitrightcheckpoints.doc
Description	This document describes how to prepare a proper accessibility audit. It has full information about accessibility standards, audit process and detailed audit criteria. RNIB is a member of W3C and most of their methods are based on w3w standards (http://www.w3.org/TR/WAI-WEBCONTENT/). They are using triple priority grade system (A,AA,AAA). More detailed about methodology on http://www.w3.org/TR/WCAG20/

Name of methodology	<i>"See it right" accessible website audits</i>
Target group	For web page designers and source code creators
Commentary	Needs to be adapted (only the method description provided, needs to be prepared as for our test specification) After a few modifications could be usable. Compatible with W3C standards. Resources: needs time and resources to prepare for our purposes

Name of methodology	<i>AIR (Accessibility Internet Rally) judging forms from knowbility.org</i>
Author of methodology	knowbility.org
Link	http://www.knowbility.org/content/common/AIRJudgingForm.xls
Description	This is a research form used for statistic purposes. It has a very precise description of what a page needs to be accessible. The test is based on negative points, which means that every error takes some points from total number of points. At the end there is a bonus area where a page can get some extra points for its design and whole effect. The test is divided in parts: basic accessibility, techniques used for universal accessibility, barriers to defeat, usability and bonus points.
Target group	For web page designers and source code creators
Commentary	Needs to be adapted (only the questionnaires provided) Very specialized, for advanced users only.

Name of methodology	<i>Barrierefinder Accessibility test</i>
Author of methodology	
Link	http://www.barrierefinder.de/start.asp
Description	It is simple accessibility test. There are strictly defined barriers. They are grouped in phases. Phases are set in special order – from basic ones to complex. Every phase has a five scale degree. Low degree shows the weak point of a site. On that basis it answers about whole accessibility of web page.
Target group	For web page designers and source code creators, for users
Commentary	

Name of methodology	<i>Barrierefinder Accessibility test</i>
	<p>Needs to be adapted (prepared form is not suitable for our user testers needs)</p> <p>Only German version available.</p> <p>Unknown standards.</p> <p>Does not ensure comprehensive web pages evaluation.</p>

Name of methodology	<i>Biene Award criteria</i>
Author of methodology	Lack of data
Link	http://www.einfach-fuer-alle.de/award2005/kriterien/
Description	Criteria of accessibility. They need a special jury that is checking if the criteria are fulfilled. After that they give answers about its functionality.
Target group	For unknown group of users
Commentary	<p>Needs to be adapted (only the criteria provided)</p> <p>Very specified, for advanced users only.</p> <p>German version only</p>

Name of methodology	<i>BIK BITV short test</i>
Author of methodology	Lack of data
Link	http://www.bik-online.info/verfahren/kurztest/index.php
Description	Test is very similar to UWEM and is based on methodology of w3c. It uses page http://www.w3.org/WAI/eval/#eval
Target group	For unknown group of users, web page designers and others
Commentary	<p>Needs to be adapted (but uses the W3C methods)</p> <p>German version only</p>

Name of methodology	<i>DIN CERTCO certification</i>
Author of	Lack of data

Name of methodology	<i>DIN CERTCO certification</i>
methodology	
Link	http://217.110.109.153/de/produkte_und_leistungen/produkte/barrierefreie_planungen_bauten_und_produkte/barrierefreie_website_internetpraesenz_.php
Description	Mostly uses the basics of other testing standards, especially w3c.
Target group	For unknown group of users
Commentary	Needs to be adapted. Most part is confident. German version only

Name of methodology	<i>WAES (Web Accessibility Evaluation System)</i>
Author of methodology	The Cyberspace Policy Research Group (CyPRG), at the Univ. of Arizona, Tucson and George Mason Univ
Link	http://www.cyprg.arizona.edu/index.html
Description	This methodology is based on criteria of proper web page design. It uses 0 or 1 indicator system. Each criteria can get 0 or 1 point. At the end there is a summary of points. WAES give answer to the page designer about accessibility to all data important for user. It is only useful for already created pages.
Target group	For web users or code correctors
Commentary	Adaptable (only question are provided) Usable for user testing but not usable for disabled people. Almost compatible with WCAG standards (attachments)

Name of methodology	<i>PubbliAccesso Guidelines for testing</i>
Author of methodology	Lack of data
Link	http://www.pubbliaccesso.it/biblioteca/documentazione/guidelines_study/index.htm
Description	Methodology based on w3c standards. Simplified w3c methodology
Target group	For web users and source code creators
Commentary	Needs to be adapted.

Name of methodology	<i>PubbliAccesso Guidelines for testing</i>
	Not very useful. English version available

Name of methodology	<i>Programa Acesso in Portugal</i>
Author of methodology	Lack of data
Link	http://www.acesso.unic.pcm.gov.pt/
Description	Lack of data
Target group	Lack of data
Commentary	Portuguese version only.

Name of methodology	<i>The JISC methodology</i>
Author of methodology	Lack of data
Link	http://www.jisc.ac.uk/uploaded_documents/Usabilitymethods.doc
Description	The JISC methodology is used to test web-based services in respect of their usability and accessibility. Tests aim to define usability and accessibility of services for users that have different experience and different knowledge levels of information technology.
Target group	For web page designers, source code creators and users
Commentary	Adaptable for our needs (needs questionnaire) Reliable – thanks to a different level of users Compatible with w3c standards The method can be prepared in many ways, that ensures proper evaluation of user tests. Needs further description to prepare proper version for disabled user testers Test is performed during the design stage, prototype realization and application implementation. Test is repeatable in many phases – time and cost consuming

Name of methodology	<i>The PZDN methodology</i>
Author of methodology	Authors of the PZDN methodology worked for The Polish Blind Association
Link	http://www.pzn.org.pl
Description	The methodology is based on source code analysis done by experienced programmers They search in page for parts that make the whole site inaccessible. A basic assumptions of the methodology is that user testing is done by blind users with qualifications in this computer science.
Target group	For web page users and site overseers
Commentary	Adaptable for our needs (development of result forms, needs detailed description) Usable for user testing. Compatible with standards (use w3c checkpoints). Reliable – can give true result of accessibility by disabled people. Suitable for our needs but needs disabled people with computer science knowledge.

Name of methodology	Unified Web Evaluation Methodology (UWEM)
Author of methodology	24 European organizations in three European projects combined in a cluster, [BentoWeb][EIAO][S-EAM] information http://www.wabcluster.org/
Link	http://www.wabcluster.org/
Description	UWEM 0.5 provides an evaluation procedure consisting of a system of principles and practices for manual and automatic evaluation of Web accessibility for humans and machine interfaces. The methodology aims to fully conform with WCAG 1.0 guidelines. The methodology covers evaluations of one web page, an entire site (irrespective of size), or multiple sites, a method for sampling, clarifications of the checkpoints, user testing protocols and information necessary for interpretation and integration/aggregation of results.

Name of methodology	Unified Web Evaluation Methodology (UWEM)
Target group	For web users, page designers and source code creators
Commentary	Adaptable (almost everything is prepared) Usable for user testing. Compatible with WCAG standards. Reliable – can give true result of accessibility by disabled people. Best for user testing needs.

3.2.3. Evaluation of collected methodologies

The result of the evaluation of the collected methodologies is presented in Table 6.

Table 6. Evaluation of collected methodologies

Name of methodology	"See it right" accessible website audits	
Criterion	Input	Output
Access	Text of methodology available Contact with author available	Yes
User testing	Undefined part for user testing	No
For people with disabilities	Part of methodology for usability people with disabilities	More part
Not require special software or purchase license	Standard software requirements, no license, only written permission need	Ok
estimated costs	Cost for user testing	Medium
Work consumption	Work consumption for user testing is undefined	4-6 people
Need for modification	Time needed for modification	much time
Final satisfaction		5. Quite satisfied

Name of methodology	AIR (Accessibility Internet Rally) judging forms from knowbility.org	
Criterion	Input	Output
Access	Text of methodology unavailable Contact with author available	No/yes One of conditions is fulfilled
User testing	Designed for user testing, form provided	Yes
For people with	Part of methodology for usability	All

Name of methodology	AIR (Accessibility Internet Rally) judging forms from knowbility.org	
disabilities	people with disabilities, but programming specialists	
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	High
Work consumption	Work consumption for user testing is undefined	7-10 people
Need for modification	Time needed for modification	much time
Final satisfaction		4. Neither dissatisfied nor satisfied

Name of methodology	Barrierefinder Accessibility test	
Criterion	Input	Output
Access	Text of methodology available only in German version Contact with author available	No/Yes
User testing	Designed for user testing, form unprovided	Yes
For people with disabilities	Methodology based on user possibilities	All
Not require special software or purchase license	Standard software requirements, no license, open source	Ok
estimated costs	Cost for user testing	Medium
Work consumption	Work consumption for user testing is undefined	2-3 people
Need for modification	Time needed for modification	Little time
Final satisfaction		5. Quite satisfied

Name of methodology	BIK BITV short test	
Criterion	Input	Output
Access	Text of methodology available only	No/Yes

Name of methodology	BIK BITV short test	
	in German version Contact with author available	
User testing	Designed also for user testing, form provided form w3w site	Yes
For people with disabilities	Part of methodology for people with disabilities.	All
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	High
Work consumption	Work consumption for user testing is undefined	4-6 people
Need for modification	Time needed for modification	much time
Final satisfaction		4. Neither dissatisfied nor satisfied

Name of methodology	Biene Award criteria	
Criterion	Input	Output
Access	Text of methodology available only in German version Contact with author available	No/Yes
User testing	Each criteria checked by special jury	No
For people with disabilities	Part of methodology for programming specialists with disabilities	Little
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	High
Work consumption	Work consumption for user testing is undefined	more than 10 people
Need for modification	Time needed for modification	Much time
Final satisfaction		3. Not very satisfied

Name of methodology	DIN CERTCO certification	
Criterion	Input	Output
Access	Text of methodology unavailable, only German version	No
	Contact with author available	Yes
User testing	Unknown	No
For people with disabilities	Part of methodology for people with disabilities.	All
Not require special software or purchase license	Special software requirements, license	Require
estimated costs	Cost for user testing unknown	Medium
Work consumption	Work consumption for user testing is undefined	Over 10
Need for modification	Time needed for modification	Much time
Final satisfaction		2. Not satisfied

Name of methodology	WAES (Web Accessibility Evaluation System)	
Criterion	Input	Output
Access	Text of methodology available	Yes
	Contact with author available	
User testing	Not designed for user testing,	No
For people with disabilities	Not for disabled people	No
Not require special software or purchase license	Standard software requirements, license	ok
estimated costs	Cost for user testing	medium
Work consumption	Work consumption for user testing is undefined	2-3 people
Need for modification	Time needed for modification	much time
Final satisfaction		1. Not at all satisfied

Name of methodology	PubbliAccesso Guidelines for testing	
Criterion	Input	Output
Access	Text of methodology available, full text in Italian, methodology in English Contact with author available	Yes
User testing	Involves user testing	Yes
For people with disabilities	Partially might be used by disabled people	Little – only small participant
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	High
Work consumption	Work consumption for user testing is undefined	4-6 people
Need for modification	Time needed for modification	much time
Final satisfaction		2. Not satisfied

Name of methodology	The JISC methodology	
Criterion	Input	Output
Access	Text of methodology available Contact with author available	Yes
User testing	Designed for user testing and has many other properties	Yes
For people with disabilities	Methodology designed also for use of disabled people	Most
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	Small
Work consumption	Work consumption for user testing is undefined	4-7 people
Need for modification	Time needed for modification	Little time
Final satisfaction		6. Satisfied

Name of methodology	The PZDN methodology	
Criterion	Input	Output
Access	Text of methodology available Contact with author available	Yes
User testing	Designed for user testing with computer specialists	Yes
For people with disabilities	Methodology designed for disabled people	All
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	small
Work consumption	Work consumption for user testing is undefined	2-3 people
Need for modification	Time needed for modification	Little time
Final satisfaction		6. Satisfied

Name of methodology	Unified Web Evaluation Methodology (UWEM)	
Criterion	Input	Output
Access	Text of methodology available Contact with author available	Yes
User testing	Designed for user testing, form provided	Yes
For people with disabilities	Designed for disabled people use	All – specially for disabled people
Not require special software or purchase license	Standard software requirements, no license	Ok
estimated costs	Cost for user testing	Small
Work consumption	Work consumption for user testing is undefined	4-6 people
Need for modification	Time needed for modification	No time
Final satisfaction		7. Very satisfied

3.2.4. Selection of methodologies

There are only three satisfying or very satisfying methodologies, so after the Stage 1 - evaluation a decision has been made to evaluate the following methodologies:

- UWEM [<http://www.wabcluster.org/uwem05>]
- JISC [http://www.jisc.ac.uk/uploaded_documents/Usabilitymethods.doc]
- PZDN [www.pzn.org.pl]

This section gives a short description of the three methodologies that will be evaluated.

UWEM methodology

Unified Web Evaluation Methodology (UWEM) is a result of a joint effort of 24 European organizations in three European projects combined in a cluster, [BentoWeb][EIAO][S-EAM] information [<http://www.wabcluster.org/>]

UWEM 0.5 provides an evaluation procedure consisting of a system of principles and practices for manual and automatic evaluation of Web accessibility for humans and machine interfaces. The methodology aims to fully conform with WCAG 1.0 guidelines. Currently the UWEM is limited to priority 1 guidelines, but in the coming phase of the Cluster, priority 2 guidelines will be added.

The methodology covers evaluations of one web page, an entire site (irrespective of size), or multiple sites, a method for sampling, clarifications of the checkpoints, user testing protocols and information necessary for interpretation and integration/aggregation of results.

The methodology is designed to meet the following requirements:

- Technical conformance to existing Web Accessibility Initiative (WAI)
- Recommendations and Techniques documents.
- Tool and browser independence: questions and tests are given in a 'pure' form, making them as tool independent as possible.
- Unique interpretation: questions shall have only one way of being interpreted, e.g., through simple language and if necessary a specific example.
- Replicability: different web accessibility evaluators who perform the same tests on the same site should get the same results within a given tolerance.
- Translatability: the methodology will address localization issues.

- Compliance with Regulation (EC) No 808/2004 of the European Parliament and of the Council of 21 April 2004 concerning Community statistics on the information society.

The methodology includes information about:

- Statistical methods for sampling
- Critical task analysis and key use scenarios
- Computer assisted selection of content for evaluation
- Manual selection of content for evaluation
- Interpretation and integration/aggregation of results
- Information about user testing
- Information for policy makers

The UWEM methodology is a fundamental methodology to evaluate user testing in the EIAO project and provides a basis for comparison with other methodologies. The UWEM methodology contains detailed characteristics of user tests (Chapter 7) and related forms (Appendix D).

JISC methodology

The JISC methodology is used to test web-based services in respect of their usability and accessibility. Tests aim to define usability and accessibility of services for users that have different experience and different knowledge levels of information technology. Web services/applications have to be friendly for recipients with vision defects, hearing defects and physical disability. The methodology assumes that additional factors influencing accessibility are: stress, tiredness, temporary disability (for example broken arm, wound, etc.) and environmental conditions (for example noise, improper lighting, power fluctuation, etc.)

In performed tests, the level of usability and accessibility is defined through usage of the following techniques or their combinations:

- opinion poll,

- focus groups,
- interview,
- World Wide Web Consortium (W3C) check lists,
- user testing (formal and retrospective),
- evaluation of preliminary version (prototype),
- heuristic evaluation,
- automated testing – Bobby (now WebXact),
- analysis of web logs.

During the usability of web services/applications testing the methodology assumes to carry out two issues: who is a receiver of the web-service, and is it attractive for this group. Generally testing of interface attractiveness and functionality is carried out among the end users through questionnaires or laboratory user testing. Expert testing and usage of heuristic methods by expert users aims to answer the following questions: is the service/application interface friendly? is it possible to learn how to operate the system easily and quickly? Interviews and focus groups aim to define users' expectations and preferences. The result of this analysis depends greatly on the experience and professionalism of person carrying out the tests.

In the case of testing of web service/application accessibility, automatic testing is mostly used. The methodology prefers for this reason software tools such as *Bobby* (now WebXact) and *LIFT*.

The JISC methodology is applicable for defining usable and accessible requirements of newly created web services/applications and for their testing according to early stage scenarios. Testing is performed during the design stage, prototype realization and application implementation. The methodology is concerned only with usability and accessibility testing, looking at operation and usage. It is not used for security evaluation and more widely defined quality assessment. It is not suitable for application testing and evaluation in the exploitation phase. However, it contains a developed user testing part that can be used to help accomplish our goals.

PZDN Methodology

The PZDN (Polskie Zasady Dostępu Niewidomych - Polish Association of the Blind) methodology was developed by blind computer scientists from the Polish Association of the Blind working on the design and evaluation of web pages under the guidance of Jacek Zadrożny.

The methodology does not have a precisely formalized form, and is used only by blind experts in e-usability evaluation. A basic assumption of the methodology is that user testing is done by blind users with qualifications in this computer science. These people have a deep knowledge of HTML. They will assess the web page not by looking at graphic elements nor at the presentation of the web page in a browser, but only through reading the HTML syntax and text using a screen reader.

These blind people do not use a mouse, using only a keyboard.

Additionally in the methodology tools are used for checking the syntax such as the following software:

Bobby - [www.bobby.watchfire.com]; now [www.webXact.com]

Cynthia Says - [www.cynthiasays.com]

check lists based on the WCAG specification.

Issues that may cause problems for disabled people in web pages accessibility in the PZDN methodology are:

1. graphics and graphic links; without suitable alts;
2. flash animations;
3. forms;
4. document structure; especially badly described headings
5. applets and scripts;
6. navigation system.

The PZDN methodology is a specific approach designed only for web page evaluation in respect of web accessibility for disabled people; it is based on user testing, where users are specialists in computer science.

A disadvantage of the methodology is the lack of formal description and limited use of it. However it has been used for the evaluation of many web pages, for example banking web pages in Poland.

As the methodology is based almost only on expert user testing it is suitable to compare with user testing methodologies used in UWEM. However a detailed development of PZDN results forms would be necessary. The PZDN methodology is also usefully for other disabilities.

3.3. Testing of the methodologies for the evaluation of the web accessibility (stage 2)

3.3.1. Introduction

Section 3.3.2. presents the criteria preparation and form documents for the methodology evaluation. Section 3.3.3. presents the criteria for the evaluation of selected methodologies by experts. Section 3.3.4 provides the form to be used by the test engineers for the evaluation of each methodology. Section 3.3.5. introduces the formal model for the methodologies comparison. Section 3.3.6. describes the preparation phase. Section 3.3.7. provides a detailed procedural description of the methodology for the user testing. Section 3.3.8. describes how the methodologies evaluation is performed. How the final report will be prepared is presented in section 3.3.9. The complete stage 2 is described in table 2.

3.3.2. Preparing criteria and documentation for testing

Before starting the testing of the chosen methodologies, it is necessary to elaborate the methodology evaluation criteria and forms for the collection of the tests results data. This must be done earlier so that the results of the methodologies testing would not affect the criteria. Developed criteria were introduced in section 3.3.3. in table 7. Methodology evaluation is more simple to estimate when the test forms are carried out by the user and cover the chosen criteria. Additionally, the forms should also contain other information not to suggest user testers and test engineers what is the most important in the evaluation. The evaluation methodology form is introduced in table 8, section 3.3.4. The methodologies comparison can be done with usage of the methodologies of many criteria, for example by using criteria defined in section 3.3.3. and comparing the criteria results against each other. Results obtained this way lead to the ambiguous interpretations. Method which is proposed will allow for the evaluation of the methodology with using one number. This number is a

part of the different criteria. The proposition of such correspondence is introduced in table 9, section 3.3.5. The numerical values assigned to the importance of particular criteria have subjective character, and that is why they should be verified by independent experts that do not take part in testing, the best after pilot testing (section 3.4.).

3.3.3. Evaluation criteria

This section provides a complete list of criteria used for evaluation of the tested methodologies by expert testers. Criteria for assessment of web accessibility evaluation methodologies is presented in table 7

Table 7. Criteria for assessment web accessibility evaluation methodologies

Criterion	ADAPTABILITY
Description	Methodology is adaptable – if user can use it without any need of additional modifications.
Measurement input	Number of modifications, such as: addition of questionnaires, data collect programme.
Measurement result	Number of modifications necessary for using the methodology and Subjective users' opinion, stated on 1-7 scale: <ol style="list-style-type: none"> 1. Not at all satisfied 2. Not satisfied 3. Not very satisfied 4. Neither dissatisfied nor satisfied 5. Quite satisfied 6. Satisfied 7. Very satisfied

Criterion	USABILITY
Description	Methodology is usable – if it qualifies for evaluation of web pages for disabled people.
Measurement input	Users' opinions (test engineers and test users)
Measurement result	Subjective users' opinion, stated on 1-7 scale: <ol style="list-style-type: none"> 1. Not at all satisfied

Criterion	USABILITY
	2. Not satisfied 3. Not very satisfied 4. Neither dissatisfied nor satisfied 5. Quite satisfied 6. Satisfied 7. Very satisfied

Criterion	COMPATIBILITY
Description	Methodology is compatible – if includes standards, such as WCAG, standards .
Measurement input	WCAG requirements must be defined.
Measurement result	Number of WCAG requirements that were not included in the methodology.

Criterion	EVALUATION COMPLETENESS
Description	Methodology ensure comprehensive web page evaluation
Measurement input	List of missing points in the methodology – Information obtained from users.
Measurement result	Number of missing elements in the methodology.

Criterion	PERSONAL DATA SECURITY AND TESTERS' SAFETY
Description	Methodology take notice on the personal data security of testers and their physical safety (e.g. does it require displacement of disabled people to special laboratories)
Measurement input	Means for security of personal details (agreement for test execution, agreement on terms and amount of payment, agreement on personal data processing). Not exposing disabled people for additional problems.
Measurement result	Verification whether the methodology fulfils requirements according to the check-list defined in the measurement input. (Yes/No response)

Criterion	STAFF RESOURCE REQUIREMENT
Description	Amount of time that must be dedicated for methodology execution.
Measurement input	<p>Define the resource requirement in terms of testers' hours and required number of testers.</p> <ul style="list-style-type: none"> • Staff resource requirement test scenarios preparations • Staff resource requirement test explanation • Staff resource requirement test execution • Staff resource requirement results recording <p>Staff resource requirement for drawing conclusions and final web page evaluation in accordance with the methodology.</p>
Measurement result	Number of hours necessary to develop above mentioned elements.

Criterion	COSTS
Description	Costs which must be incurred to undertake the methodology
Measurement input	Costs incurred for required number of users, test engineers , experts, and other people needed to use the methodology. Costs of additional equipment and software. Do not include costs of methodologies comparison.
Measurement result	Sum of the actual costs of methodology implementation and costs of one web page evaluation.

Criterion	INDEPENDENCE
Description	Independence of the methodology from equipment and software used .
Measurement input	Browsers on which tests can be carried out, required operating systems, requirements regarding additional computer hardware and software.
Measurement result	Number of browsers, participation of browsers in the market where tests can be carried out with consideration of the browser version. Percentage participation in the operating systems market where

browsers operate. Independence from the browsers version and additional plug-in.

Criterion	RELIABILITY
Description	Definition of whether the approach and results obtained give the same results in the case of test execution by different testers and different users.
Measurement input	Results of web page measurement obtained by different teams of testers.
Measurement result	Differences in web page measurements with usage of the same methodology and executed by different testers.

Criterion	TRANSLATABILITY
Description	Methodology can be used in EU countries that use different languages.
Measurement input	Languages in which the method is written and questionnaires supporting the method.
Measurement result	Number of languages the method is translated into; number of world languages in which the method is known (English, Spanish, French, Chinese, Arabic, Russian), method description accessibility and possibility of free translation (open licence).

Criterion	ATTRACTIVENESS
Description	methodology capability of being attractive for the user.
Measurement input	Users' opinions (test engineers, test users and test experts)
Measurement result	Subjective users' opinion, stated on 1-7 scale: <ul style="list-style-type: none"> 1. Not at all satisfied 2. Not satisfied 3. Not very satisfied 4. Neither dissatisfied nor satisfied 5. Quite satisfied 6. Satisfied

7. Very satisfied

-

Criterion	LEARNABILITY
Description	Capability of enabling the user to learn the methodology.
Measurement input	Users' opinions (test engineers, test users and test experts)
Measurement result	Subjective users' opinion, stated on 1-7 scale: 1. Not at all satisfied 2. Not satisfied 3. Not very satisfied 4. Neither dissatisfied nor satisfied 5. Quite satisfied 6. Satisfied 7. Very satisfied

Criterion	MAINTAINABILITY
Description	Methodology capability of being modified. Modifications may include corrections, improvements or adaptation to new browser and tools.
Measurement input	Possibility of methodology modifications without infringing copyright.
Measurement result	Number of possible corrections, number of tools that methodology can cooperate with.

-

Criterion	FLEXIBILITY
Description	Methodology capable of maintaining a specified level of performance when used under specified conditions for example lack of test users in specific disabilities.
Measurement input	Requirements regarding user testers.
Measurement result	Percentage participation of required testers for which methodology provides reliable results with reference to the obligatory number.

Criterion	FAULTLESSNESS
Description	Methodology in included any fault or problems in implementation.

Measurement input	Fault or problem in implementing methodologies
Measurement result	Number and importance of known fault or problems

Criterion	EASINESS
Description	easiness of using methodology
Measurement input	Easy to understand, implement and fill questionnaires for disable people
Measurement result	Subjective users' opinion, stated on 1-7 scale: <ul style="list-style-type: none"> 1. Not at all satisfied 2. Not satisfied 3. Not very satisfied 4. Neither dissatisfied nor satisfied 5. Quite satisfied 6. Satisfied 7. Very satisfied

3.3.4. Test result form

The following table presents the methodology evaluation form to be used by user-testers:

Table 8 Form used in testing

Methodology Testing	
Methodology name	
Methodology version	
Source address of methodology description	
Person/institution responsible for the methodology	
Test engineers	
User - Tester	Disability type

Methodology Testing			
Date and time of test beginning		Place where test was held	
Date and time of test completion			

Number	Checkpoint		Used Criterion
1	<p>Main focus of the methodology</p> <p>Describe what the main focus of the methodology (e.g. association of blind? A technical company? Engineers?...) and which aspects have been left out.</p> <p>Is methodology usable for people with disabilities? Which disabilities?</p>		usability
2	<p>What is missing in the methodology? (According to the WCAG requirements.)</p>		Evaluation completeness

Number	Checkpoint		Used Criterion
	Describe missing elements.		
3	<p>If methodology has some problems or error in implementation? Repeat for all problems encountered</p> <p>List errors that occur in the aims, definition and solution.</p>		faultlessness
4	<p>Is method easy to understand?</p> <p>Indicate elements with ambiguous interpretation.</p>		learnability
5	<p>Indicate elements that can be omitted and the grounds for omitting them.</p> <p>Indicate elements that do not follow the needs of the disabled person.</p>		flexibility
6	Indicate elements inconsistent with other elements.		reliability

Number	Checkpoint		Used Criterion
	Indicate elements internally inconsistent.		
7	Describe resources necessary to realize functions mentioned in the methodology. Define resource costs apart from standard costs of computer, operating system and browser.	People	cost
		Computer hardware	costs
		Software	cost
8	<p>If methodology needs special hardware or software</p> <p>Describe problem with obtaining necessary equipment</p>		independence
9	<p>Does the methodology contain the complete description of its implementation and not need to change?</p> <p>List elements need to change in methodology implementation.</p>		adaptability
10	Can all elements described in the		Evaluation completeness

Number	Checkpoint		Used Criterion
	<p>methodology be tested during user testing?</p> <p>List elements that impede user testing.</p>		
11	<p>Are questionnaires easy to fill in by disabled persons? Is there need for an additional person to help?</p> <p>Define problems occurring when filling in the questionnaires</p>		easiness
12	<p>Does the method guarantee protection of users' rights? . For example protection of user-testers' personal data, protection of minor users' rights, security agreements and safety of testers</p> <p>How is this done?</p>		Personal data security and tester safety
13	<p>Define methodology staff resource requirement [in hours]:</p> <ul style="list-style-type: none"> • Staff resource for 		Staff resource requirement

Number	Checkpoint		Used Criterion
	<p>test scenarios preparation</p> <ul style="list-style-type: none"> • Staff resource for test explanation • Staff resource for test execution • Staff resource for results recording • Staff resource for drawing conclusions and final web page evaluation by methodology. 		
14	Define other costs of user tests execution in this methodology.		cost
15	<p>Is the methodology consistent with WCAG requirements?</p> <p>If not which WCAG requirement not included</p>		compatibility
16	<p>Is methodology attractive for user tester?</p> <p>Why attractive or why not</p>		attractiveness
17	In which language is methodology described		translatability

Number	Checkpoint	Used Criterion
18	<p>If methodology easy to change?</p> <p>Which change you can be included without problems</p>	maintainability
	Signature of person responsible for test preparation and execution.	

The following attachments must be provided alongside the completed form:

- Shortened method description (test selection, method of test execution, reports forms , results presentation).
- Detailed testing areas (for example by WCAG criteria).
- Blank questionnaires.
- Developed and completed test scenarios.
- Method for the completion of questionnaires by users.
- Users' comments concerning testing process.
- Test engineers - comments concerning methodology.
- List of tested domains and web pages.
- Copy of tested web pages.(source code, print screen)
- description of the test environment - version of OS, name and version of browsers, name and version of software used by users, name of hardware technology used by users (braille bars, trackballs....)
- Other problems in test accomplishment and observed incidents (software crash, operating system crash, web page accessibility failure, problems with network

accessibility, prolonged reaction time , requirement of additional plug-ins, equipment breakdown, etc.).

- Recording or video documentation from performed tests.
- Filled form of numerical evaluation of method according to the proposed criteria.

3.3.5. Method for the methodologies comparison

The importance of each criterion for the methodology and numerical methods of evaluating the degree of fulfilling can be defined. The method below gives one value for each methodology; before using this approach it is necessary to agree the list of criteria and percentage score for each. It is also advisable to carry out pilot testing.

A first version of the method is given in table 9

Table 9 Method of compare methodologies.

<i>Criterion</i>	<i>Percentage Participation</i>	<i>Part of test form which used criterion</i>
ADAPTABILITY	5	9
USABILITY	10	1
COMPATIBILITY	10	15
EVALUATION COMPLETENESS	5	2 and 10
PERSONAL DATA SECURITY AND TESTERS SAFETY	5	12
STAFF RESOURCE REQUIREMENT	10	13
COSTS	10	7 and 14
INDEPENDENCE	5	8
RELIABILITY	5	6
TRANSLATABILITY	5	17

<i>Criterion</i>	<i>Percentage Participation</i>	<i>Part of test form which used criterion</i>
ATTRACTIVENESS	5	16
LEARNABILITY	5	4
MAINTAINABILITY	5	18
FLEXIBILITY	5	5
FAULTLESSNESS	5	3
EASINESS	5	11

The tables includes only the first version of method for evaluation on methodologies. The weight given to each criterion needs more discussion by expert testers. Criterion usability, compatibility, cost and needed staff is of course more important than others, but it is difficult to say whether 2 times is enough. After user testing weight of criterion can be changed.

3.3.6. Preparation of the user testing

Before starting the user testing, beside the preparation of the evaluation criteria, there are necessary such organizational actions as selection and training of user testers, selection and training of test engineers, selection of web pages for testing, and definition of which test engineer will test which methodology.

Each of the methodologies described in section 3.2.4. is assigned to at least 2 test engineers

The test engineers will include TUW students specializing in computer science, who have basic knowledge concerning usability and accessibility testing. Prior to the test preparation, students will be trained in the scope of WCAG specification, and legal regulations concerning web page evaluation. Additional training will refer to psychological aspects of testing with users, especially disabled users.

The selection of students for carrying out testing aims to construct a homogeneous research group. Additionally each methodology will be used in the training by at least two people.

That will enable additional evaluation of the degree of ease in user testing by people not initially trained in the methodology.

The aim of using people not initially trained in the methodology they are testing is to prevent testing by experts who may have a vested interest in the methodology they are using and thus the evaluation may be compromised. The introduction of two test engineers for the evaluation of each methodology will enable comparison and discussion of the results; not having prior in-depth knowledge of the methodology aims to ensure the evaluation is more objective.

Web pages for testing will be selected from a list of web pages which contain errors in relation to WCAG guidelines (3 pages) and randomly selected for set of potential web pages that disabled people in Poland are interested in (2 pages) (for example governmental web pages, companies and institutional web pages that facilitate the needs of disabled people (transport, banks, insurance, social service), web pages containing information helpful to disabled people (news, work, publications, libraries)).

Selected web pages are as follows:

<http://www.sejm.gov.pl/> Polish parliament
<http://www.teatr-rozrywki.pl/> theater
www.zus.pl Polish social security institution
www.um.warszawa.pl Warsaw town council
www.inteligo.pl/infosite/ internet bank
www.kopernik.lodz.pl hospital

3.3.7. Performing the user testing

Usability testing of the methodology will be undertaken by test execution with real users from among the identified disability groups. Repeatability of tests will be assured by testing web pages by different people. Representativeness of tests will be assured through test execution by different people with the same initial preparation and through testing of web pages by many users.

After familiarization with the methodology, the next stage will be preparation of test scenarios and test plans by test engineers.

If test questionnaires are missing in the methodology, they will be prepared. Where necessary questionnaires will be translated into Polish.

After preparation of test scenarios and test plans by test engineers, experts will evaluate their methodological correctness and compatibility with the selected methodology; any mistakes will be corrected. This verification aims to assure the suitable quality of the materials prepared before the beginning of user testing.

Users for testing will be selected from among the groups defined in section 1.3. and according to the given methodology's requirements. User testing will be performed after agreeing with users on a suitable time and place. The number of users will be chosen according to the requirements defined in each methodology. After test completion, testers will gather all materials and screen shots of web pages. If possible, test engineers will download the source code (HTML and CSS) for later analysis. Then they will perform their own evaluation of methodology.

The output will include the user test plans, follow-up reports, etc. produced by the test engineers. However, to assess the methodology TUW will also develop questionnaires, interviews, etc. through which they would find out more about how the particular methodology worked. The forms to be used by test engineers for evaluation of the methodology are provided in section 3.3.4.

General rules for the user test are:

- 1) Each test engineer uses the assigned methodology to design test plan and test case to carry out an user test of a website.
- 2) The user-tests are run according to the plan developed by test engineers.
- 3) Each user -tester uses the assigned methodology to collect, analyze and report results of tests.
- 4) Every test engineer documents test results and his/her own comments concerning the methodology.

3.3.8. Evaluation of the tested methodologies

Test results will be delivered to experts, who will analyze the results and compare them with their own evaluation of web pages, and will evaluate each methodology according to criteria defined in section 3.3.3. Evaluations of the methodology by particular test engineers will be

compared against each other and verified by experts, who will have the right to correct errors.

A meeting will then take place with expert testers and test engineers for the purpose of drawing conclusions on the evaluation of the different methodologies and summarization of the research. Conclusions and recommendations to be included in the evaluation final report will be introduced at the meeting. After the meeting of expert testers they will prepare formalized conclusions using the method shown in section 3.3.5. Results are analyzed by expert testers, who will evaluate the methodologies used in the tests in accordance with criteria agreed earlier (section 3.3.3).

Collected information determines the foundation for the preparation of the report from methodology evaluation.

3.3.9. Final results of the methodologies evaluation

Final report from the methodology evaluation will include:

- evaluation (according to criteria introduced in section 3.3.3) made by experts.
- summary evaluation of the methodologies according to the method introduced in section 3.3.5., made by experts.
- aggregated comments from user-testers about each methodology.
- aggregated evaluation reports made by test engineers recorded in the test form introduced in section 3.3.4.
- aggregated comments from test engineers about the methodology
- remarks showing any required improvements of the UWEM methodology, arising from performed tests.
- list and source code of tested web pages.

The final report will be prepared within one week of this meeting, translated into English and forwarded to the WP4 Work package leader for comments from partners and the Cluster.

3.4 Pilot testing

3.4.1. Aim of the pilot testing

Before performing the complete examination of the methodologies, the test procedures will be verified by pilot testing, which will cover verification of only one methodology (probably UWEM). Pilot testing aims to verify if the proposed procedures, including test forms, fit the requirement. In particular it will check that the organization of the procedure will assure efficient testing. Special care will also be taken to verify that test forms and tests results will be adequate to enable evaluation of the methodologies by the chosen criteria and that the proposed weights are appropriate by checking the sensitivity of the overall result to variations in criteria scores.

3.4.2. Planing pilot testing

One expert tester working with one test engineer will be responsible for he planning of the user pilot testing. Co-operation is useful because because if any problems exist, the cooperation can help to solve it. First user testing of only one selected methodology (UWEM) is run.

3.4.3. Running pilot testing

Test engineer is responsible for preparing test scenarios and test cases. Test engineer is also responsible for the execution of the user testing and collection of data.

3.4.4. Check pilot testing results

After finishing pilot user testing, test engineer sends all collected data to the expert testers. Expert testers can use collected data for the evaluation of the methodologies.

3.4.5. Improvement of test forms and criteria weighs

Most important part of the pilot testing is improvement of the prepared test form criteria and testing process. After finishing the evaluation test, experts will have a meeting with test engineers to discuss the form and criteria. User testers can also participate in this meeting. Discussion will concentrate on forms and testing process. After the discussion, changes to forms or testing process can be made.

If there are problems with pilot testing, or great changes in forms are made, expert testers can include next methodologies to the user testing.

4 User testing of WAM relevance: background

4.1 Introduction

Test Area Two will undertake user testing to validate the Web Accessibility Metrics (WAMs) developed in WP3 and integrated into the UWEM web accessibility indicator model.

Validation of the WAMs will be achieved through comparison of barriers detected by users with those detected by the Observatory.

MMU will lead on this testing area and the following responsibilities have been agreed with WP4 partners:

MMU:

- Selection of example web sites to be included in the User Test Site (input from all partners).
- Identify participants for user testing (mostly identified from the user requirements gathering in WP2 through the FAST User Forum and the RNIB Research Database).
- Identify websites to be used, based on the WAMs provided by WP3.
- Develop content for the User Test Site using the UWEM and other related work, this will include related tasks and questions (input from FBL).
- Questions for online evaluation forms to be included in the User Test Site.
- Set up, pilot test, and run remote user testing.
- Analysis and writing of results for D4.3.1/4.3.2

AUC:

- Design and implement test site for user testing and online feedback (with NK).
- Methods for identifying user's PC set-up (with NK).
- Run comparative tests using the Observatory.

NK

- Design and implement test site for user testing and online feedback (with AUC).
- Methods for identifying user's PC set-up (with AUC).

TUW:

- Input to selection of example web sites.
- Participation in discussions and comments relating all draft documents.

FBL:

- Input to selection of web sites (including comments to be used to develop scenarios and tasks).
- Development of tasks and questions.

FTB:

- Confirm proposals fit into WP3 work.
- Provide a list of the most important WAMs for testing in this first iteration.
- Input to Observatory comparison tests.

All Partners

- Input to D4.3.1.

4.2 Cluster involvement.

The User Testing Protocol section of the UWEM will be used to inform the methods developed for user testing. Work undertaken by other Cluster projects (Ben-to-Web in particular) will also be used to develop the experimental framework for WP4.

The results of the remote user testing methods can also be fed back to the next version of the UWEM and used to inform work of the other Cluster projects.

4.3 Background to methods for testing

Development of the methods for testing the relevance of the WAMs has been informed through identification and analysis of different methods for user testing. This section describes some of the background data which has been used to develop the specification and implementation of the experimental environment described in Section 5.

4.3.1 Test Suites

Test Suites are a means of testing complex objects, particularly computer software systems and hardware components, for conformance with laid down standards or protocols. The Concise Tech Encyclopedia (<http://www.tech-encyclopedia.com/>) defines the term *Test Suite* as “A collection of test scenarios and/or test cases that are related or that may cooperate with each other” (http://www.tech-encyclopedia.com/term/test_suite cited 30/06/05). A *Test Case* (synonymous with *Test Scenario*) is defined as “A set of test data and test programs (test scripts) and their expected results. A test case validates one or more system requirements and generates a pass or fail.” (http://www.answers.com/test_case cited 30/06/05). Test suites are often automated and consist of perhaps hundreds or even thousands of tests.

Test Suites are typically used in testing components or large-scale software products that have to perform very complex routines. For example they are used in the communications industry for testing the performance of equipment such as cable modems, automatic telephone exchanges, gateway switches, messaging systems, and transponders used in satellite telecommunications. For such equipment to perform very accurately at very high speeds there has to be an extremely high degree of conformance to standards protocols.

Examples of the use of test suites for software products are found for implementations of the operating system Unix; computer programming languages C, Java and Ada; Bluetooth technology; Apache web server products; aviation industry computer based training systems; and specifications such as the Extensible Mark-up Language (XML). The test suites for some of these consist in fact of a collection of several test suites.

Test suites are likely to be complex software packages themselves which will automatically test the target subject/product. The suite will be designed to carry out tests that are standard for the subject/product under scrutiny, and will automatically apply many, perhaps thousands of tests and produce ‘pass’ or ‘fail’ results for each test. Some test suite software packages - Bobby for instance - are available for on-line use, rather than the package being available locally.

Alternatively, a test suite can consist of a list of criteria to be tested, with sample tests and instructions on how to carry out the tests, for manual use by the tester. Typically this consists of a comprehensive checklist of test criteria and dependent criteria, each with a sample test routine. Each routine has a list of possible outcomes, usually 'pass' or 'fail', and might include suggested actions for any necessary repair. The WCAG Test Suite for web site accessibility falls into the latter category.

The HTML Test Suite for WCAG 2.0 W3C Working Draft 2005-08-11

(<http://www.w3.org/WAI/GL/WCAG20/tests/>) states:- "This document provides information to Web content developers who wish to make their content comply with the 'Web Content Accessibility Guidelines 2.0' It includes test cases, code examples, and references to help authors test for compliance." The test suite is a working draft and currently (23 November 2005) and incomplete, as are the WCAG 2.0 guidelines it is based on.

Further work on Test Suites has been conducted by the Ben-To-Web project and is documented in their public deliverable 4.1:

http://www.mla.gov.uk/documents/longitude_rep.doc

Although the WCAG Test Suites have many elements that would be very useful to EIAO, for the purposes of Test Area One, this may not be appropriate. It seems more appropriate to use it in conjunction with the more stable version 1.0 of the guidelines WCAG 1.0 <http://www.w3.org/TR/WCAG10/> as a guide to designing a User Test Site for EIAO which is suitable for the subject(s) of the testing. The proposed User Test Site is described in Section 4.3.2 below.

4.3.2 User Test Site

The User Test Site will be provided online with a URL for participants to access. Participants will be asked to log in to the site, this will enable identification of the type of assistive technology used (if any) and of specific disabilities.

The Test Site will provide online instructions asking participants to look at up to five archived websites, undertake up to two tasks for each, and provide feedback. This will be achieved using an online form specifically designed for the purpose. Participants should be able to complete and submit this process remotely, without a facilitator, at a time and place convenient to them.

The Test Site will be kept in on a server that is private and that can only be accessed by those involved in the project. The Test Site will be coded so that web crawlers are unable to index the site.

Pilot tests using a shorter version of the Test Site will highlight any accessibility problems with the Test Site and indicate how long each task is likely to take. Any adjustments can then be made before participants are asked to undertake the testing proper. All participants will be asked to take part in the pilot test to ensure they are able to use the final Test Site properly.

4.3.3 User testing protocols

A number of user testing protocols have been examined⁴. From these, two user testing protocols have been considered. These are remote and face-to-face user testing.

The advantages of remote testing are that users can undertake the process at a time and location convenient to them. They can also use the ICT they are most comfortable with and are not performing in a 'test' situation - thus their actions and responses are likely to be more realistic than when performing tasks in a lab situation with an observer. The disadvantages are that it is not possible to observe and understand what users are doing and therefore misunderstandings may occur in relation to a perceived barrier and an actual barrier. By asking users to complete a problem evaluation task evaluation form, describing the nature of the problem experienced and an evaluation of the task, it is hoped that this will lessen the risk of misunderstanding.

The advantages of face-to-face testing are that the data captured is often much richer than if gathered remotely. When observing a task it is possible to achieve a good understanding of the user experience and of any barriers faced. It is also possible to ask post-task questions to clarify any actions or responses that are unclear to the observer, thus minimising the risk of misunderstandings. The disadvantages of face-to-face testing are that it is a very time consuming and often expensive process, requiring specialist equipment and

⁴http://www.jisc.ac.uk/project_usability1.html

<http://www.bitvtest.de/main.php?a=dl>

Disability Rights Commission (DRC). 2004 *The Web: Access and Inclusion for Disabled people. A formal Investigation conducted by the Disability Rights Commission*. London: DRC.

<http://www.wabcluster.org/uwem05/>

a lab environment to undertake the testing. It also places the user in an unnatural setting, possibly using equipment they are unfamiliar with. This may have an effect on the way the user performs a task and thus on their response to the experience. To avoid this, every attempt needs to be made to ensure the user feels relaxed and comfortable with the equipment they are using. A strong emphasis also needs to be made on the fact that the user is not being testing, simply the process (e.g. the web site).

Pilot testing of any methods should be undertaken to ensure:

- The testing framework is usable and understandable.
- The time allocated is realistic (e.g the whole process should not take more than two hours).
- The responses will provide appropriate data for analysis and comparison.
- The data collected will be reliable and repeatable.

The testing of WAM relevance will mainly use remote user testing. The decision to use this method was influenced by the participants who were recruited through the FAST User Forum and the RNIB Research Database. These participants (the majority) have quite a wide geographical spread thus making it difficult to bring them to one location (or to visit them in their location of choice).

5 Testing of WAM relevance: specification and implementation

A framework for user testing will draw on those developed by Ben-to-Web along with the UWEM and other related documents which have been identified by the project. The user testing framework should include the following steps:

- Clarify purpose of the testing, what needs to be measured etc.
- Development of scenarios, tasks and questions.
- Enlisting of participants.
- Pilot testing of the experimental environment.
- Run user testing.
- Data analysis and report writing.

The above steps will be discussed in more detail below:

5.1 Purpose of the user testing

The purpose of the user testing is as follows:

- User testing to compare barriers identified by users (i.e. ranked by each user group as input to barrier probability estimates) with those detected by WAMs.
- Where possible, to make a differentiation between a general barrier and a barrier experienced when solving a given task (pursuing some critical path). For example: a general barrier would be present in many places of a site, e.g. a menu based on Java Script or a fixed font size (see Appendix Two Section 7.2); a barrier linked to some critical task could be part of a form to be filled out online that is not accessible to a screen reader (see Appendix Two, Section 7.8).
- Differentiation of barriers by disability. It should be possible to assess failure modes relating to the barrier and the disability. This could also be used to assess the user centric barrier probability model (see UWEM 0.5 Section 6.4.3) although further development of this model would be required before effective assessment could take place.

- Comparison of barriers detected by the observatory (WAMs) (this will not be run simultaneously with the user testing).

The complete list of B-WAMS (failure modes) according to the latest version of D3.1.1 are as follows:

- 1.1 alt missing for
- 1.1 alt empty (image map)
- 1.1 alt empty (image contained in empty link)
- 1.1 alt too long for
- 1.1 multiple instances with different descriptions
- 1.1 empty body (image map)
- 1.1 empty body (empty link)
- 1.1 multiple <object> instances with different descriptions
- 1.1 empty body (video)
- 1.1 empty body (audio)
- 1.1 alt missing for <area>
- 1.1 alt too long for <area>
- 1.1 alt empty for <area>
- 1.1 alt missing for <input>
- 1.1 alt too long for <input>
- 1.1 alt empty for <input>
- 1.1 alt missing and body empty for <applet>
- 1.1 alt empty and body empty for <applet>
- 1.1 alt too long for <applet>
- 1.1 use of frames without <noframes>
- 1.1 no link in <noframes>

- 1.5 missing corresponding text link for <area>
- 3.3 deprecated features for layout
- 3.4 absolute sizes in <frameset>
- 3.4 absolute sizes for font
- 3.5 no <h*> on page
- 3.5 header level skipped
- 3.5 higher level than first header
- 4.2 first <abbr> without title
- 4.2 first <acronym> without title
- 4.3 primary language not defined
- 5.2 <th> in nested <table>
- 5.2 id reference missing for headers
- 5.2 <colgroup> missing
- 5.2 rowgroup element missing
- 5.2 <th> without scope or id (2 level table)
- 5.2 <td> without headers and not in scope (2 level table)
- 5.2 <th> without id (3+ level table)
- 5.2 <td> without headers (3+ level table)
- 5.5 summary missing for <table>
- 5.5 summary empty for <table>
- 5.6 <th> too long without abbr
- 5.6 abbr too long for <th>
- 7.4 META refresh on page
- 9.2 onclick used
- 9.2 onmouseup without onkeyup
- 9.2 onmousedown without onkeydown

- 9.2 onclick without onkeypress
- 9.4 tabindex not a number
- 9.4 multiple definition of tabindex value
- 9.5 multiple definition of acceskey value
- 10.1 javascript with "window.open()"
- 10.1 target attribute with value "_blank"
- 10.4 Value missing for text <input>
- 10.4 value of length 0 for text <input>
- 10.4 string length 0 for <textarea>
- 11.1 doctype too old
- 11.2 use of deprecated features
- 12.1 title missing for <frame>
- 12.1 title empty for <frame>
- 12.1 title too long for <frame>
- 12.4 id missing for form element
- 12.4 corresponding <label> missing for form element with id
- 12.4 for missing for <label>
- 12.4 for attribute for <label> uses undefined id
- 13.1 different link targets
- 13. <title> missing

A total of 24 WAMs (based on WCAG 1.0) have been identified as the A-WAM tests for EIAO v1. A selection of the most appropriate will be included in this iteration of user testing:

- 1.1: missing image element.
- 3.4: absolute font sizes used.
- 5.5: summaries for tables missing or inappropriate.

- 5.6: inappropriate abbreviations for header labels.
- 12.1: titles for frames missing or inappropriate.
- 12.4: poor use of labels for form fields.
- 13.1: links for different targets given the same name or description.

Using these checkpoints as a basis, a purposive sample of sites has been identified (see Appendix One) and will be used to track whether users perceive them as a barrier. Ideally each site should include one potential error, but could include using more than one page in a site to test this (e.g. to test navigation of a site).

Having undertaken the user testing, the same websites will be checked by the Observatory to compare barriers identified automatically with barriers identified by users.

5.2 Development of tasks and questions

5.2.1 Tasks

For the testing of each checkpoint an appropriate website has been identified using Section 5 of the UWEM and WP3. Tasks and questions have been developed for each website, based on the elements which can be tested both by users and by the Observatory (e.g. missing ALT text, Null ALT text, input elements etc) so that the user tests and results can be comparable with the results of the Observatory.

Tasks have been developed to help simulate an actual searching and browsing experience. The aim of this is to achieve more realistic (and thus reliable) results from user testing by developing tasks that reflect the general experience of users. Each task will be appropriate to the checkpoint being tested. For example, to test Checkpoint 1.1 (relating to missing ALT text), a task will be developed for someone who cannot see the screen and needs to use a screen reader/refreshable Braille (although users with other disabilities will also be asked to perform the task). A website with missing ALT text will be identified and a suitable task developed with related questions.

A full list of tasks and questions are provided in Appendix Two.

5.2.2 Questions

Questions aim to gather a mixture of information on:

- The user: disability and assistive technologies used (if any)
- Rating of the task performed (satisfaction, overall experience, opinions etc).
- Identification and description of barriers to accessibility.

Identification of users' disability or disabilities and their use of assistive technologies has already been achieved via the end-user surveys and interviews conducted for WP2. This will also be verified when participants are enlisted for the user testing. Data relating to the workstation environment of each user could also be captured automatically when the user logs in.

Users will be asked a set of questions about the task they have performed (based on the UWEM User Testing Protocols, Likert scales, open responses etc). It may also be possible to ask them for the time required to perform some specific tasks, or to find a simple way of automatic logging the time spent. Comparisons could then be made of how different disabilities and technologies act differently and how much time they might require to perform a single task. Time can also be directly related to the accessibility/usability of a site, as the faster a task can be completed could be an indication of how well the site has been designed.

Users will be asked to rank the success of the task using the rating developed for the UWEM but should also include a "0: I don't agree that this is a usability problem at all" rating, as suggested in the Ben-To-Web Report on evaluation framework for project monitoring: D3.1, page 21⁵:

- Rating of the severity of the problem by the participant:
 - 4 = Catastrophe.
 - 3 = Major problem.
 - 2 = Minor problem.
 - 1 = Cosmetic problem only.

5ftp://ftp.fit.fraunhofer.de/bika/projects/bentoweb/deliverables/BenToWeb_D3.1.pdf

0 = Not a problem.

If appropriate, a more detailed explanation of the meaning for each severity rate could also be given (see 7.8.2 of UWEM); this could be done verbally for face-to-face testing. For remote testing and online forms, text could also be accompanied by an image showing a suitable facial expression for each response (e.g. a smiley face for a positive response) which would help people with dyslexia.

A full list of tasks and questions is provided in Appendix Two.

5.3 Enlisting of participants

Selection of user groups will be based on the recommendations made in the UWEM (0.5) to conduct user testing with people with each of the following disabilities: three with functional blindness, three with partial sight and three with dyslexia or literacy problems.

Other studies will also be used to inform selection. For example, the Disability Rights Commission study⁶ evaluated 100 web sites by 50 people with a wide range of disabilities (e.g. blind, partially sighted, deaf, hard of hearing, physically impaired and dyslexic) and found that all accessibility problems were uncovered by the testing with blind, partially sighted and dyslexic people. Therefore for web sites without atypical components or requirements, testing with these three participant groups is considered sufficient.

Participants will be selected on a basis of the following:

- Disability.
- Use of assistive technologies.
- IT experience.
- A control group of non-disabled people.

Potential participants have been involved in the user requirements data gathering on a voluntary basis. As a result of the end-user follow-up interviews, around 40 people have agreed to take part in the Test Area Two. From these, 25 users have been selected to take part in the first iteration of user testing according to the disability criteria described in Section 5.2.1. For this iteration of testing people who are deaf or hard of hearing have not been

⁶Disability Rights Commission (DRC). 2004 *The Web: Access and Inclusion for Disabled people. A formal Investigation conducted by the Disability Rights Commission*. London: DRC.

included unless they experience one or more of the disabilities described above. A control group of non-disabled people will be included in the 25.

Currently, the following assistive technologies will be used by participants (this list may be amended prior to the testing):

- Voice recognition, no mouse.
- Jaws 7.0.
- JAWS 6.1.
- Dragon naturally speaking.
- JAWS 7 , no mouse.
- Enlarged text size using Accessibility in Windows XP.
- Zoomtext 7.1.
- At work: Dolphin Lunar 5.21. At home: Windows 2000 professional, High contrast black.
- JAWS, adaptations to fonts and colours (large fonts, white on black text).
- SuperNova 6.5 and JAWS (on Pacmate).
- JAWS 6.1, no mouse.
- Text enlargement and colour adjustments.
- JAWS 6.20, no mouse.

To encourage participation, a nominal sum will be paid to participants for their time. Other rewards could also be considered such as a public acknowledgment or a gift or prize draw.

5.4 Pilot testing

Pilot testing must be undertaken prior to the final implementation of the experimental environment. Pilot testing will reveal important issues such as how long (on average) it takes users to complete the tasks and will allow adjustments to be made if necessary. It may be necessary to reduce the number of sites tested if the average time taken is unreasonable

(over 2 hours). It will also reveal any problems with the protocol developed and allow adjustments to be made accordingly.

5.5 Running the tests

The testing will mainly be undertaken remotely. For users of assistive technologies (ATs) such as a screen reader or refreshable Braille, there is a possible issue here about the quality of assistive technologies used, i.e. success may depend on the quality of the AT equipment rather than the accessibility of the website. To avoid this participants will be allowed to use their own ATs, or ATs which they are comfortable using, therefore reducing the effect the AT may have on the success of the task.

Each user will evaluate up to 5 websites, with up to 2 tasks per site. The total amount of time taken to complete all tasks is aimed at up to 2 hours (pilot testing will confirm this). This is considered an acceptable amount of time for a participant to spend on the testing.

If appropriate, users will be asked to evaluate the sites which may cause them the most problems, e.g. blind users to evaluate missing ALT text. However, this should be undertaken with caution as assumptions should not generally be made about which sites may cause problems for users.

The following steps should take place for the remote testing:

- Participants receive instructions for conducting the remote testing (if possible this could include a pre-meeting/training session to explain process).
- Logging of PC set up (captured when the user logs in to the test site).
- Undertaking of the tasks:
 - Task One:
 - Access Web site for testing and tasks.
 - Task Evaluation form.
 - Submission (which takes user to the next task).
 - Task Two:
 - Access Web site for testing and tasks.
 - Task Evaluation form.

- Submission (which takes user to the next task).

This continues through each task until the final submission when user has completed all tasks and they get a message thanking them for their input and advising them of how payment will be received.

The Test Site will be divided into two sections:

- Test Site One (10 users, 5 sites each): Sites and tasks relating to missing ALT text checkpoints (e.g. 1.1), and to Headings (e.g. 3.4; 3.5;) and Links (13.1) and Tables (5.5).
- Test Site Two (10 users, 5 sites each): Sites and tasks relating to Tables and labels (e.g. 5.5; 5.6), Navigation (e.g. 9.4) Frames (12.1) and Forms (12.4).

The control group of non-disabled users will undertake both Test Sites One and Two.

Suggested spread of participants are as follows:

- Test Site One Participants:

Visually impaired: Zoomtext level 2, version 7.1, reverse colours, enlargement.

Visually impaired: Lunar 5.21, display settings in high contrast black.

Visually impaired: JAWS and PC adaptations – large fonts white on black text.

Visually impaired and deaf: text enlargement and some colour adjustments.

Blind: Screen reader.

Blind: JAWS 6.20 screen reader no mouse.

Blind: JAWS 6.1 and dragon Naturally speaking.

Blind: JAWS 7 screen reader, no mouse.

Physically impaired: No assistive technology or PC adjustments.

Physically impaired: Voice output.

Dyslexia: No assistive technology or PC adjustments.

Dyslexia and visually impaired: JAWS screen reader, subtitles.

- Test Site Two Participants:

Visually impaired: Accessibility enlarged text in Windows XP.

Visually impaired: SuperNova 6.5 and JAWS.

Visually impaired: Zoomtext 7.1 screen mag.

Visually impaired: 18 pt screen mag and large monitor.

Blind: JAWS 7 screen reader, Focus 80 Braille display.

Blind: JAWS 6.1 no mouse.

Blind: JAWS, Supernova/Hal, no mouse.

Physically impaired: No assistive technology or PC adjustments.

Dyslexia: No assistive technology or PC adjustments.

Dyslexia: No assistive technology or PC adjustments.

- Test Site One and Two Participants:

Control group of up to five non-disabled people.

5.6 Data analysis and report writing.

The tests will comprise a set of pre-defined, off-line web sites containing different barriers to be checked (See Appendix One). The barriers are based on the Web Accessibility Metrics (WAMs) identified in WP3 (see 5.1) and will be evaluated by users representing the different groups identified, and then by the observatory tools. Different sets of WAMs may be used in the different releases, and these will influence the sequence of testing.

Data from all participants in user testing will be aggregated. Results will be compared with minimum thresholds for performance and acceptance, for example:

- Success rate: a web site will not be deemed accessible unless disabled participants have given their confidence in completing the task a mean rating of at least 4 (on a 7 point scale). This can be compared with their satisfaction of the accessibility of the site (again a mean rating of at least 4).
- Problems: a web site will not be deemed accessible unless disabled participants can undertake tasks without encountering any problems deemed to be “accessibility catastrophes” (“completely impedes/disrupts my progress with the task”).
- Acceptance: a web site will not be deemed accessible unless disabled participants give the site a mean rating of at least 4 (on a 7 point scale) for overall accessibility.

To verify the above, users will be asked a set of questions based on the UWEM User Testing Protocol Section 7 (Likert scales, open responses etc). Barriers detected by users

and verified through the questions can then be compared to the barriers detected by the Observatory.

- Questions of a quantitative nature (the ranked questions) will be analysed using SPSS or a Spreadsheet as appropriate.
- Questions of a qualitative nature (the open questions) are likely to be analysed using a spreadsheet and database (i.e. without the use of an automated tool such as Atlas-ti). The reason for this is because answers in the pilot test have tended to be brief and to the point. If however the length of the responses increase considerably it may be necessary to use a more powerful tool for analysis (i.e. Atlas-ti).

Data relating to the user's disability, experience, and use of assistive technology has already been logged and will be used with the above results to identify possible relationships and to assess the strengths of these (as described in 5.7).

It will be difficult to undertake any statistically significant data analysis with this area of user testing because a random sample is not being used. However, it will be possible to look at the data using cross tabulation to indicate possible relationships between variables (e.g. satisfaction and disability) and then to look further at the strengths of those relationships (correlation). For example, the relationship between:

- The ability to understand an element presented as a graphic, and the overall satisfaction with completing the task.
- The ability to use the whole page (e.g. which uses more than one frame) with the overall satisfaction with completing the task.
- The disability.

Results from user testing will be compared with results from testing the same web pages using the Observatory. This will validate the relevance of the selected WAMs to the user experience, which can then be used to inform further development for WP3. Evaluation of values for the relevance and weight of failure modes will also be explored and may be used to inform the development of the Fcui value (see UWEM 0.5 Section 6.4).

The tested methods will also be fed into the UWEM User Testing Protocol – in particular a new section on remote user testing. Findings will be used to validate and improve the UWEM by:

- Inspection of responses to ensure they can be analysed in a way that is comparable (amendments made where necessary).
- Analysis of responses to ensure success, problems and acceptance can be identified.
- Analysis of responses to ensure identification of accessibility barriers can be identified and are repeatable (amendments made where necessary).

5.7 Result of initial pilot testing

An initial pilot test was run with a blind participant using JAWS version 7.

Two tasks were piloted. Each took the participant around 20 minutes to complete (task and evaluation). This suggests that the original plan to ask users to test up to five sites with up to two tasks per checkpoint could be a little ambitious. Further piloting will confirm this, but it is likely that participants will only be asked to perform one task per checkpoint to be tested.

The testing identified no significant accessibility problems with the Test Site, however recommendations for enhancements to the site were made and acted upon.

A recommendation was made that a mixture of compliant and non-compliant pages be tested. The reason for this is that users very quickly feel de-motivated and may give up with the testing process if they feel they are failing every task. This recommendation has been taken into account.

Pilot testing of the revised version then took place with all participants. The average time to complete both tasks was 30 minutes, which confirmed the need to develop a final version which asks the participants to test five web pages, with one task per page.

The pilot testing identified some minor accessibility problems relating to the Test Site. For example, the addition of a paragraph between each of the evaluation options was requested; the shortening of questions to help improve accessibility when using screen magnification; correct labelling of radio buttons; time-out facility to be removed or set to a longer amount of time.

Other issues raised related to terminology and the need for clearer instructions. For example, an explanation of what is meant by 'satisfied', 'dissatisfied' in relation to the tasks performed.

The pilot test confirmed that the process would provide results that were relevant to the user testing of WAM relevance and would provide reliable data between user groups that could be compared and analysed. Any issues raised in the pilot feedback will be addressed before the final version of the Test Site is distributed to participants, thus ensuring the testing process will not impede the actual testing undertaken.

Thus, the next stage will be the first iteration of user testing, evaluation and analysis of WAM relevance, which will be reported in D4.3.1: experimental results, evaluation and analysis.

6 Appendix One: Websites for testing of WAM Relevance.

Checkpoint 1.1 (P1)

Provide a text equivalent for every non-text element (erg., via "alt", "longdesc", or in element content). This includes: images, graphical representations of text (including symbols), image map regions, animations (erg., animated GIFs), applets and programmatic objects, ascii art, frames, scripts, images used as list bullets, spacers, graphical buttons, sounds (played with or without user interaction), stand-alone audio files, audio tracks of video, and video.

Examples of web page/s that conform

www.rnib.co.uk

www.wired.com :Images used as buttons has ALT and title attribute used correctly.

Empty ALT are marked with "*"

Examples that do not conform

www.amazon.co.uk

www.centraltrain.co.uk : (See --> centraltrain timetable): map of trains with no ALT and neither Longdesc. Also, page has not a relevant title element.

Checkpoint 1.5 (P3)

Until user agents render text equivalents for client-side image map links, provide redundant text links for each active region of a client-side image map.

Examples of web page/s that conform

www.travelocity.co.uk

<http://www.map-of-spain.co.uk/> Maps' links are provided both via a map and via ordinary text links.

<http://www.cubamapa.com/index.htm> It allows to access maps both via graphical map and also with plain links on the page.

Examples of web page/s that do not conform

<http://www.azwardriving.com/maps/>

<http://www.metoffice.gov.uk/>

Checkpoint 3.3 (P2)

Use style sheets to control layout and presentation.

Examples of web page/s that conform

[Http://www.happycog.com](http://www.happycog.com)

<http://www.csszengarden.com/>

<http://www.zeldman.com/>

<http://www.webstandardsawards.com/>

<http://www.paumanokreview.com/>

Examples of web page/s that do not conform

<http://www.dmeurope.com/>

<http://www.links2mobile.com/> Use of BOLD attribute for presentational effects (e.g. Mobile Digital Rights Management heading).

<http://www.monster.co.uk/>
www.bbc.co.uk Use of BOLD attribute for presentational effects (erg. BBC Two link).

Checkpoint 3.4 (P2)

Use relative rather than absolute units in markup language attribute values and style sheet property values.

Examples of web page/s that conform

<http://www.webstandards.org/inc/css/wsp.css>
<http://www.capgemini.com/css/main.css>
<http://www.stuffandnonsense.co.uk/assets/css/fruit.css>

Examples of web page/s that do not conform

<http://lonelyplanet.com/>
http://lonelyplanet.com/css/lp_home.css The site has absolute values for table height and width, while the CSS uses absolute (px) values for font attributes.
<http://www.ebu.ch/styleEi.css> Uses CSS with absolute px values
<http://www.oireachtas.ie/styles/web.css> The css defines absolute font size

Checkpoint 3.5 (P2)

Use header elements to convey document structure and use them according to specification.

Examples of web page/s that conform

www.bbc.co.uk
 Proper use of headers, H1, H2, H3, H4.
<http://www.dmag.org.uk/>
 Proper use of Headers.
http://www.stroke.org.uk/accessible_site.html Uses H1 and h2 only...

Examples of web page/s that do not conform

www.stanfords.co.uk
 Does not use any headers.
<http://www.apple.com/> Doesn't use headers at all.

Checkpoint 5.2 (P1)

For data tables that have two or more logical levels of row or column headers, use markup to associate data cells and header cells.

Examples of web page/s that conforms

www.waitrose.co.uk

Examples of web page/s that do not conform

<http://www.viamichelin.co.uk/viamichelin/gbr/tpl/hme/MaHomePage.htm>

Checkpoint 5.5 (P3)

Provide summaries for tables.

For example, in HTML, use the "summary" attribute of the TABLE element.

Examples of web page/s that conform

www.mmu.ac.uk Home page uses tables for layout purposes. It states this in the Summary attribute of the TABLE element.

http://edcdaac.usgs.gov:80/glcc/globdoc2_0.asp The table 1.1 has summary attribute. Same for "file listing" table. No caption attribute.

Examples of web page/s that do not conform

<http://www.natcom.org/Instruction/pathways/Table1.htm>

<http://kropla.com/dialcode.htm>

www.bbc.co.uk BBC Home page uses tables for layout, but does not provide a summary attribute for the TABLE element.

Checkpoint 5.6 (P3)

Provide abbreviations for header labels.

For example, in HTML, use the "abbr" attribute on the TH element.

Examples of web page/s that conform

www.mmu.ac.uk

<http://validator.w3.org/> For example, see the URL and FAQ acronyms.

<http://www.gawds.org/>

Examples of web page/s that do not conform

<http://www.bayer.com/about-bayer/names-figures-facts/page2351.htm>

<http://www.novartis.com/>

<http://www.ietf.org/>

Checkpoint 9.4 (P3)

Logical tab order

- tabindex is NAN
- Multiple definition of tabindex values

Examples of web pages that conform

• <http://www.apple.com/>

• <http://www.section508.com/>

Checkpoint 11.1 (P2)

Use W3C technologies when they are available and appropriate for a task and use the latest versions when supported.

Examples of web page/s that conform

<http://www.webstandardsawards.com/>

<http://www.joshuaink.com/>

<http://www.redhat.com/>

<http://www.opera.com/>

Examples of web page/s that do not conform

<http://www.pegasus.com/>
<http://www.eudora.com/>
<http://www.monster.co.uk/>

Checkpoint 12.1 (P1)

Title each frame to facilitate frame identification and navigation.

Examples web page/s that conforms

<http://www.hq.nasa.gov/alsj/frame.html> Page has two frames which have been given Titles: Title=Navigation bar, Title=main content.

Examples of web page/s that do not conform

<http://keirsey.com/frame.html> Uses two frames, but does not give them a title. Instead the code just says name=typeframe, name=mainframe.

Checkpoint 12.4 (P2)

Associate labels explicitly with their controls.

Examples of web page/s that conform

<http://www.rnib.co.uk>
 Properly labeled forms for the Eye Test Poll.
http://www.persil.com/ask_persil/cu_email_us/
<http://www.unilever.com/resources/contactus.asp>
<http://www.opera.com/contact/>
https://www.redhat.com/apps/response/web_contact.html

Examples of web page/s that do not conform

www.pizzahut.com Labels have not been applied to the Login to Order form, so a non-visual user may not be able to identify the correct label for each control.
http://www.dove.com/your_dove/newmember.asp?cid= Labels have not been applied to the online form, so a non-visual user may not be able to identify the correct label for each control.

Checkpoint 13.1 (P2)

Clearly identify the target of each link.

Examples of web page/s that conforms

<http://www.webaim.org/>
<http://www.alistapart.com/>

Examples of web page/s that do not conform

<http://www.thefamilyhistoryproject.co.uk/index.php> Hypertext links are written as “click here”, “Go”, and “Continue”, which if read out of context (erg. using a Links List) may not make sense.

<http://www.stakes.fi/cost219/>

Checkpoint 13.2 (P2)

Provide metadata to add semantic information to pages and sites.

For example, use RDF ([RDF]) to indicate the document's author, the type of content, etc.

Note. Some HTML user agents can build navigation tools from document relations described by the HTML LINK element and "rel" or "rev" attributes (erg., rel="next", rel="previous", rel="index", etc.).

Examples of web page/s that conform

<http://www.alistapart.com/>

<http://dublincore.org/>

<http://www.semanticweb.org/>

Examples of web page/s that do not conform

<http://www.ietf.org/>

<http://www.aerlingus.com/cgi-bin/obel01im1/index.jsp>

<http://www.ebu.ch/en/index.php>

<http://www.coke.com/flashIndex1.html>

7 Appendix Two: Tasks for Testing of WAM Relevance

The following tasks and questions have been developed for each WAM/Checkpoint identified for possible inclusion in the first iteration of the Observatory.

Where possible, the websites have been selected from those identified in Appendix One.

Descriptions provided in quotes for each checkpoint are taken from WCAG 1.0 (<http://www.w3.org/TR/WCAG10/>). The remaining descriptions for the checkpoints are taken from the work WP3 has undertaken for the UWEM 0.5.

7.1 Checkpoint 1.1

"Provide a text equivalent for every non-text element".

 element:

- Missing ALT.
- ALT is empty (e.g. image used as an image map, or decorative images which require an empty ALT tag).
- ALT is empty (for images contained in a link without textual content)
- ALT that is too long (more than 80 characters).
- Longdesc reference not available.
- Longdesc description too short (less than 80 characters).
- Multiple instances of an image (not part of a link) with different textual descriptions.

Purpose of test:

To test the accessibility of images that have potentially inappropriate ALT text.

The web page contains ALT text but does not fully conform to the checkpoint tested because on one image the description is unhelpful (it just says 'right image').

User/s who may experience problems:

Someone cannot see the screen and needs to use a screen reader/refreshable Braille may experience problems reading or interpreting images without ALT Text.

Task

Using the Thames Valley University home page, try to find out how many images are displayed.

Questions:

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), how satisfied are you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied

6. Satisfied
7. Very satisfied

What information was provided by the images on the page? Open question.

Please briefly describe what you did for this task. Open question.

On a scale of 1 – 7 (with 1 as not at all satisfied, and 7 as very satisfied) how satisfied were you with the presentation of images on the page?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe.
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.2 Checkpoint 3.4

"Use relative rather than absolute units in markup language attribute values and style sheet property values".

- Absolute font sizes used.
- Distances between borders and boxes expressed in px
- boxes and layout defined using px, pt.

Purpose of test:

To test a page that uses absolute font sizes.

The web page does not conform to the checkpoint tested.

User/s who may experience problems:

Someone who needs to use screen magnification to be able to read what is on the screen. Also by someone who needs to adjust the screen settings to very large size text.

Also people with low sight who uses the 480 x 640 screen definition.

Task:

Using the Lonely Planet Home Page: <http://lonelyplanet.com/>

Try to find out what is featured on the World Guide section of the Home page.

Questions:

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

What information was provided on the Shop section of the Home Page? Open question.

Please briefly describe what you did for this task. Open question.

On a scale of 1 – 7 (with 1 as not at all satisfied, and 7 as very satisfied) how satisfied were you with the display of information on the World Guide section of the Home Page?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe.
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.3 Checkpoint 3.5

"Use header elements to convey document structure and use them according to specification".

- No <h> element on page.

- Header levels skipped.
- Tables which contains data don't use the <th> element

Purpose of test:

To test whether correctly specified headings improve the accessibility of a page.

The web page does not conform to this checkpoint.

User/s who may experience problems:

People who use screen readers to navigate

Task:

Using Stanford's Home Page: <http://www.stanford.co.uk/>

Try and find out the main sections listed on this page using the Headers provided.

Questions:

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

What were the main sections listed on Stanford's Home Page? Open question.

Please briefly describe what you did for this task. Open question.

On a scale of 1 – 7 (with 1 as not at all satisfied, and 7 as very satisfied) how satisfied were you with the organisation of information on Stanford's Home Page?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe.
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.4 Checkpoint 5.5

"Provide summaries for tables".

<table>

- No summary for table.
- Empty summary.
- Summary too long.
- No caption element
- empty caption.
- wrong, long or inadequate caption.

Purpose of test:

To verify the presence of caption element.

The web page does conform to the checkpoint.

User/s who may experience problems:

Users who use screen reading technologies.

Task

Using the table on the Quios Press Page, try to find the date of the press release: 'Quios secures US \$ 1.5m Funding'.

Questions

How satisfied are you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

What was the date of the press release: 'Quios secures US \$ 1.5m Funding'? Open question.

Please briefly describe how you looked for the information for Task One. Open question.

How satisfied were you with the ability to find the information presented in the table?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe – (e.g. completely impedes/disrupts my progress with the task).
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.

5. Not a problem.

Overall, how satisfied were you with the accessibility of the table?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.5 Checkpoint 5.6

Abbreviations for header labels in tables

- <th> too long without abbr
- abbr too long

Purpose of test:

To test the accessibility of tables that do not have an abbreviation for table headers.

The web page does not conform to the checkpoint.

User/s who may experience problems:

Users who use screen reading technologies.

Task

Using the Center for Disease Control (CDC) report on the percentages of visits for influenza-like illness in 2005, try to find out the total number of patient visits in week 45.

Questions

How satisfied are you that you completed the task?

1. Not at all satisfied

2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

What was the total number of patient visits in week 45? Open question.

Please briefly describe how you looked for the information for Task Four. Open question.

How satisfied were you with the descriptions used in the tables?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

How satisfied were you with the presentation of sections on page?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe.
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.

5. Not a problem.

Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.6 Checkpoint 9.4

Logical tab order

- tabindex is NAN
- Multiple definition of tabindex values

Purpose of the test

To test the accessibility of navigation using the tab key only.

The web page does not conform to the checkpoint.

User/s who may experience problems:

Users who are unable to use a mouse and/or need to use keystrokes to navigate because of the assistive technology they use.

Task

Using the Section 508 Contact Us page, try to find the contact email address by navigating using the tab key only.

Questions

How satisfied are you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Was navigation using the tab key possible? Open question.

Please briefly describe how you looked for the information for Task Three. Open question.

How satisfied were you with the ability to navigate through this page using the tab?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe – (e.g. completely impedes/disrupts my progress with the task).
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied

7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.7 Checkpoint 12.1

"Title each frame to facilitate frame identification and navigation".

<frame> titles

- Missing title.
- Empty title.
- Title too long.
- Longdesc reference not available.
- Longdesc too short.
- Inadequate title

Purpose of the test

To verify whether Frames have been provided with appropriate titles.

The web page does not conform to the checkpoint.

User/s who may experience problems:

Users who cannot see the screen and needs to use a screen reader. Also by someone using the mouse to navigate.

Task:

This task has two parts, part a) and b). You should complete both parts before answering the questions.

Using the Keirsey.com About Us page: <http://keirsey.com/frame.html>

- a). Try and out when the AdvisorTeam were founded.
- b). Then try to find out if there is a link to Keirsey.com questionnaires.

Questions:

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied

5. Quite satisfied
6. Satisfied
7. Very satisfied

Describe how easily you were able to move between task a) and task b) and briefly describe what you did for this task. Open question.. Open question.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), how satisfied were you with the ability to use the whole page for both the tasks?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe – (e.g. completely impedes/disrupts my progress with the task).
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you with the task performed?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.8 Checkpoint 12.4

"Associate labels explicitly with their controls".

Explicit association of <input> and <label>

- Form element with missing id.
- Id definition without corresponding label.
- Label with missing for.
- Missing corresponding form element.

Purpose of test

To verify whether forms have been labelled correctly.

The web page does conform to the checkpoint.

Users who may experience problems:

A blind user may not be able to identify the correct label for each control if using a screen reader.

Task:

Using the Join Dove page: http://www.dove.com/your_dove/newmember.asp?cid=

Try and complete the form with your contact information (do not submit the form).

Questions:

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please briefly describe what you did for this task. Open question.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), how satisfied were you with the ability to use all the form fields?

1. Not at all satisfied
2. Not satisfied

3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

On a scale of 1 – 7 (with 1 as not at all satisfied, and 7 as very satisfied) how satisfied were you with the organisation of the Form on the Join Dove page?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe – (e.g. completely impedes/disrupts my progress with the task).
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.

7.9 Checkpoint 13.1

"Clearly identify the target of each link".

- Links with same textual description referring to different targets.
- Links with inadequate title attributes (like the "click here" links)
- links for download don't provide information about the file, size, format.
- Title attribute don't provide information on the page (the page opens in the same window? New window will open?)

Purpose of the test

To verify hypertext links have been described in a meaningful way.

Users who may experience problems:

Someone using a screen reader or Braille to interpret the data on screen. Possibly also someone using magnification or a program that does not allow you to view the whole screen at once.

Task:

Using the Winzip Home Page: <http://www.winzip.com/winzip.htm>

Follow the Link to find out more about the Winzip Self-Extractor product.

Can you find out what type of files are created by the Winzip Self-Extractor.

(NB: the target of both Winzip Self-Extractor links must be available in the same order as presented on the Home Page).

Questions:

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you that you completed the task?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please briefly describe what you did for this task. Open question.

On a scale of 1 – 7 (with 1 as not at all satisfied, and 7 as very satisfied) how satisfied were you with the description of links on the Winzip Home Page?

1. Not at all satisfied
2. Not satisfied

3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Using the following 5 descriptions, how would you rate any problems experienced:

1. Catastrophe – (e.g. completely impedes/disrupts my progress with the task).
2. Major problem.
3. Minor problem.
4. Cosmetic problem only.
5. Not a problem.

On a scale of 1 – 7 (with 1 as not at all satisfied and 7 as very satisfied), Overall, how satisfied were you with the accessibility of the page/web site?

1. Not at all satisfied
2. Not satisfied
3. Not very satisfied
4. Neither dissatisfied nor satisfied
5. Quite satisfied
6. Satisfied
7. Very satisfied

Please provide any further comments about your experience performing this task: Open question.