# Evaluation of accessibility testing methods: Which methods uncover what type of problems?

Aleksander Bai, Heidi Camilla Mork,
Trenton Schulz, Kristin Skeide Fuglerud

Norwegian Computing Center, e-inclusion group



UD2016, York, UK, 23 July 2016

# Overview

- Background: usable and technical accessibility
- Methods for testing accessibility
- Methods and procedure in our study
- Results
- Conclusion and future work

# Background: usable accessibility and technical accessibility

- A universally designed solution needs to be compatible with assistive technology (AT) **and** usable by as wide a range of people as possible.
- *Technical accessibility* refers to whether a solution is compatible with AT and follows guidelines for accessibility (WCAG).
- *Usable accessibility* is how usable a solution is for all people, including people with impairments.
- Fuzzy line between the these two categories

# Methods for testing accessibility

- Simulation kit
- Testing using automatic or semi-automatic tools
- Expert testing with guidelines/ heuristics
- Testing with diverse users

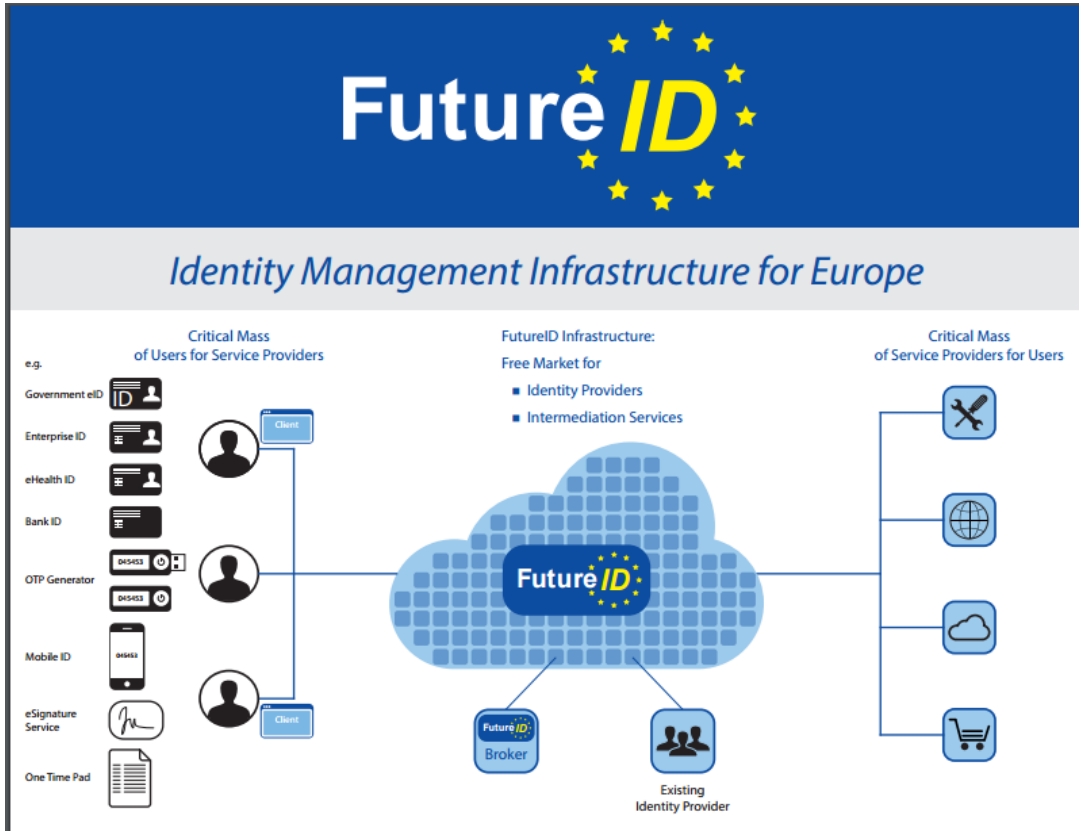
---

- Most methods address both technical and usable accessibility, but emphasis differ
- Methods cover different types of impairments

# We were going to evaluate an e-ID solution

An authentication process using a software certificate or an ID card with a card reader.

Java client and a web front-end.



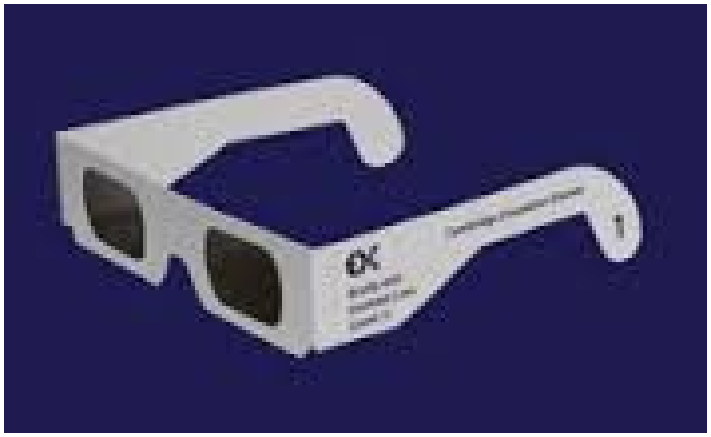
# Testing task scenarios

| no | description |
|---|---|
| 1 | Login with a invalid digital certificate |
| 2 | Login with a valid digital certificate |
| 3 | Login with invalid smart card |
| 4 | Login with valid smart card, but incorrect PIN |
| 5 | Login with valid smart card and correct PIN |

# Methods in our study

| Method | Type of impairments |
|---|---|
| Gloves and glasses<br>(Simulation kit) | – Reduced dexterity<br>– Reduced vision |
| VATLab<br>(testing tool / guidelines) | – Blindness<br>– Light sensitivity |
| Persona testing<br>(expert testing) | – Dyslexia<br>– Age related impairments |
| Manual WCAG eval.<br>(conformance check) | – Multiple impairments |

# Cambridge simulation gloves and glasses

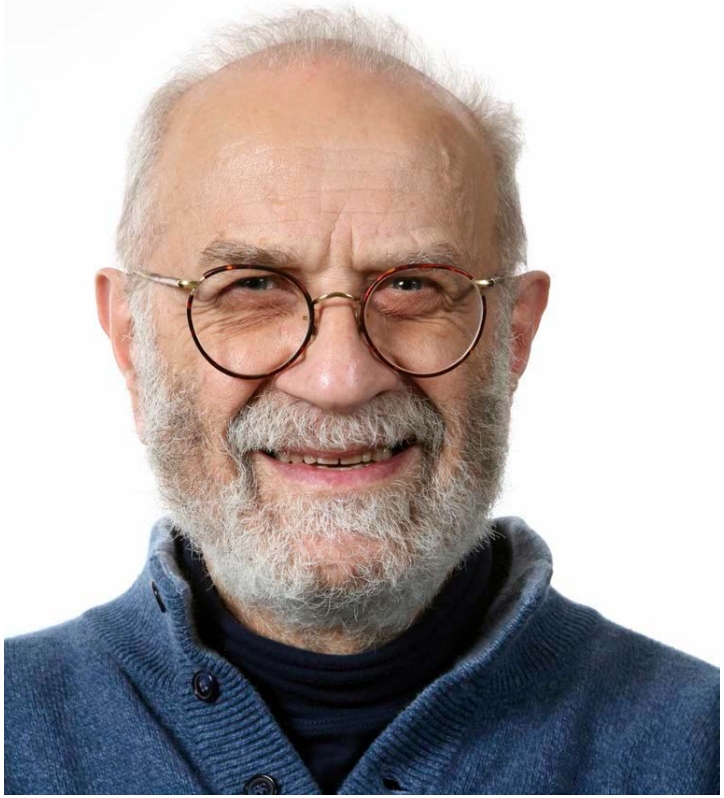



-8-

http://www.inclusivedesigntoolkit.com/betterdesign2/inclusivetools/inclusivedesigntools.html

# Persona testing: focus on cognitive issues

**Professor Friedhelm Krüger**



Age related impairments

**Jose Salazar**

Dyslexia

# VATLab – Virtual Assistive Technology Lab – Screen reader testing

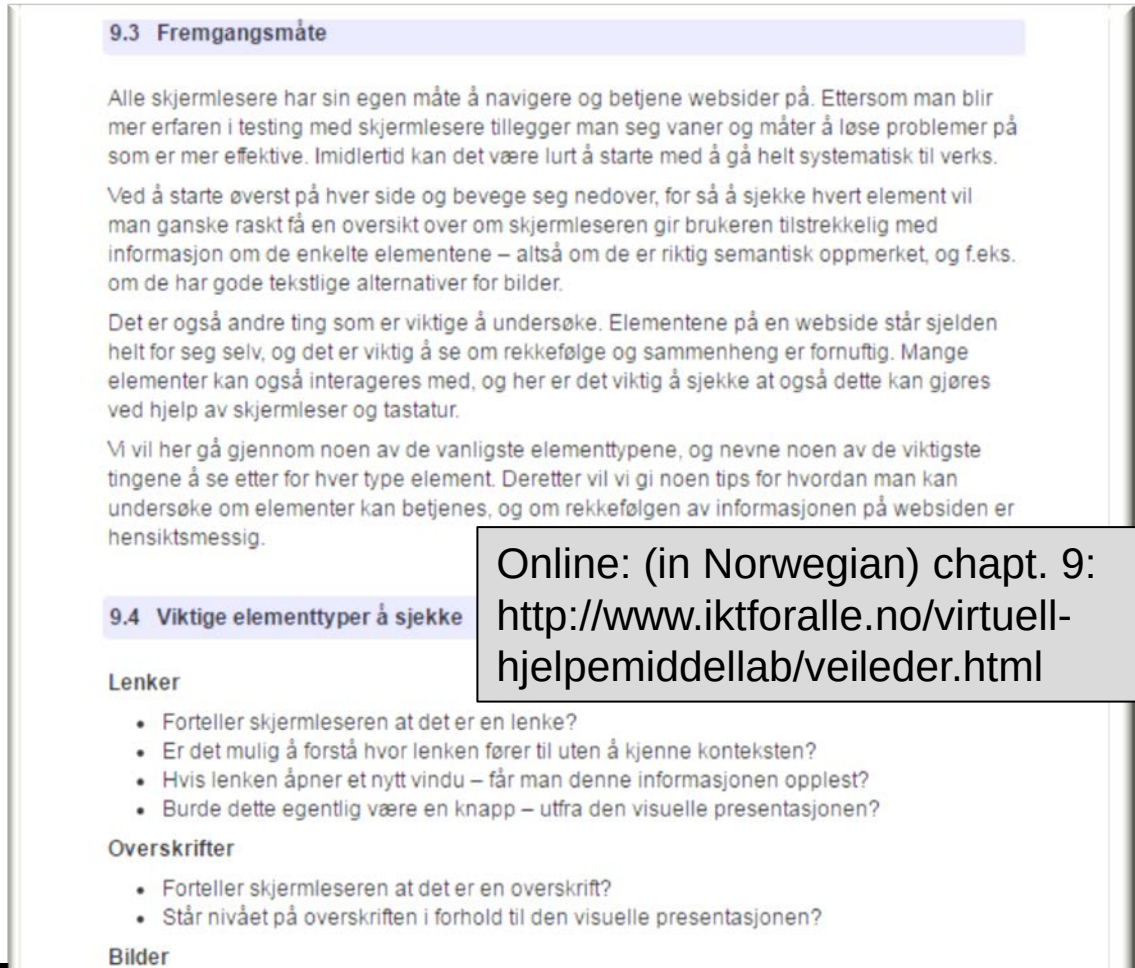
Screen readers

- SuperNova
- NVDA

Web browsers

- Internet Explorer
- Chrome
- Firefox
- Opera

Screen reader testing guideline / heuristics



### 9.3 Fremgangsmåte

Alle skjermlesere har sin egen måte å navigere og betjene websider på. Ettersom man blir mer erfaren i testing med skjermlesere tillegger man seg vaner og måter å løse problemer på som er mer effektive. Imidlertid kan det være lurt å starte med å gå helt systematisk til verks.

Ved å starte øverst på hver side og bevege seg nedover, for så å sjekke hvert element vil man ganske raskt få en oversikt over om skjermleseren gir brukeren tilstrekkelig med informasjon om de enkelte elementene – altså om de er riktig semantisk oppmerket, og f.eks. om de har gode tekstlige alternativer for bilder.

Det er også andre ting som er viktige å undersøke. Elementene på en webside står sjelden helt for seg selv, og det er viktig å se om rekkefølge og sammenheng er fornuftig. Mange elementer kan også interageres med, og her er det viktig å sjekke at også dette kan gjøres ved hjelp av skjermleser og tastatur.

Vi vil her gå gjennom noen av de vanligste elementtypene, og nevne noen av de viktigste tingene å se etter for hver type element. Deretter vil vi gi noen tips for hvordan man kan undersøke om elementer kan betjenes, og om rekkefølgen av informasjonen på websiden er hensiktsmessig.

### 9.4 Viktige elementtyper å sjekke

**Lenker**

- Forteller skjermleseren at det er en lenke?
- Er det mulig å forstå hvor lenken fører til uten å kjenne konteksten?
- Hvis lenken åpner et nytt vindu – får man denne informasjonen opplest?
- Burde dette egentlig være en knapp – utfra den visuelle presentasjonen?

**Overskrifter**

- Forteller skjermleseren at det er en overskrift?
- Står nivået på overskriften i forhold til den visuelle presentasjonen?

**Bilder**

Online: (in Norwegian) chapt. 9: http://www.iktforalle.no/virtuell-hjelpemiddellab/veileder.html

# The VATLab start screen

# Manual WCAG 2.0 evaluation

| Principles | Guidelines | Level A | Level AA | Level AAA |
|---|---|---|---|---|
| 1. Perceivable | 1.1 Text Alternatives | 1.1.1 | | |
| | 1.2 Time-based Media | 1.2.1 – 1.2.3 | 1.2.4 – 1.2.5 | 1.2.6 – 1.2.9 |
| | 1.3 Adaptable | 1.3.1 – 1.3.3 | | |
| | 1.4 Distinguishable | 1.4.1 – 1.4.2 | 1.4.3 – 1.4.5 | 1.4.6 – 1.4.9 |
| 2. Operable | 2.1 Keyboard Accessible | 2.1.1 – 2.1.2 | | 2.1.3 |
| | 2.2 Enough Time | 2.2.1 – 2.2.2 | | 2.2.3 – 2.2.5 |
| | 2.3 Seizures | 2.3.1 | | 2.3.2 |
| | 2.4 Navigable | 2.4.1 – 2.4.4 | 2.4.5 – 2.4.7 | 2.4.8 – 2.4.10 |
| 3. Understandable | 3.1 Readable | 3.1.1 | 3.1.2 | 3.1.3 – 3.1.6 |
| | 3.2 Predictable | 3.2.1 – 3.2.2 | 3.2.3 – 3.2.4 | 3.2.5 |
| | 3.3 Input Assistance | 3.3.1 – 3.3.2 | 3.3.3 – 3.3.4 | 3.3.5 – 3.3.6 |
| 4. Robust | 4.1 Compatible | 4.1.1 – 4.1.2 | | |

http://www.michaelgaigg.com/blog/tag/wcag-20/

# Testing procedure

- The 5 test scenarios were tested in same order each time
- Each test scenario was performed using the 4 methods.
- 2 testers for each method
- testers had varying degree of accessibility knowledge (from novice to experienced)
- One note-taker for each test

# Results (1): overview

- 425 issues were reported
- 213 distinct issues
  (even though we only tested a small bit of functionality)
- each method uncovered issues spread evenly between the different impairments.
  (less so for the reduced dexterity simulation)

# Results (2): number of issues found

| Method | No of issues | % of total | Critical % | Confusing % |
|---|---|---|---|---|
| Simulation kit | 58 | 27 | 24 | 33 |
| VATLab | 62 | 29 | **76** | 16 |
| Persona testing | 61 | 29 | 46 | **75** |
| WCAG | 32 | 15 | 22 | 22 |

# Results (3): Unique critical and confusing issues

| Method | Critical % | Confusing % |
|---|---|---|
| Simulation kit | 12 % | 24 % |
| VATLab | **61 %** | 12 % |
| Persona testing | 25 % | **59 %** |
| WCAG | 2 % | 5 % |

# Results (4): Issue coverage

| Method | Total coverage |
|---|---|
| Simulation kit | 42 % |
| VATLab | 45 % |
| Persona testing | 44 % |
| WCAG | 23 % |

# Conclusion and future work

- No single method that works best for finding both critical and confusing issues
- Include both methods that focuses on
  - technical accessibility and
  - usable accessibility
- All methods uncovered less than 50% of issues
- Future:
  - compare with user testing
  - improved experimental set-up

# Thank you for your attention!

Comments?

Questions?

Please contact

Kristin Skeide Fuglerud
Head of e-Inclusion
Norwegian Computing Center (NR)

E-mail: kristin.skeide.fuglerud@nr.no