

Distr.
GENERAL

Working Paper No.14
11 April 2007

ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Meeting on the Management of Statistical Information Systems (MSIS 2007)
(Geneva, 8-10 May 2007)

Topic (iii): Accessibility and usability of IT applications

EXPERIENCES USING THE EYE-TRACKING METHOD TO TEST WEBSITE USABILITY

Invited Paper

Prepared by Annegrete Wulff, Statistics Denmark, Denmark

I. INTRODUCTION

1. Testing the accessibility and usability of the products is a necessity for a statistical office. One of the core tasks is to disseminate the statistics to a variety of users ranging from professionals to the broad public. We need to know if this is done in an optimal way with respect to users' background and competences.
2. Statistics Denmark has tested the eye tracking method on ten users who were asked to carry out certain tasks on our website. Some of the results were quite persuasive and supported the comments we have received in previous tests. Others were surprising to us and made us consider changes in the way information is prioritized on the pages. The presentation will highlight some of the findings concerning the method as well as an evaluation of the user reactions.

II. PERSPECTIVE OF THE PROJECT

A. Background

3. A web site should be developed according to users' needs, preconditions and qualifications more than to the technical capabilities of the programmers and the statisticians wish promote his or her topic on the front page. To get a picture of the way users look at accessibility of the site we need to do some testing.
4. The web site is visited by 260,000 visitors monthly. It is of course unrealistic to communicate with every single one to learn about his or her view on our site. In accordance with Sundgren and Steneskog we assume that when choosing individuals who represent typical users of our site, we can expect their opinion and navigation habits to give a picture of most individuals belonging to that group.

5. According to Statistic Denmark's annual user surveys of the web site students count for close to 50 % of all users. A very big part is first-time users. (Wulff 2002, 2006). Our intension was to get information on how students would intuitively navigate our site when confronted with specific tasks.

6. Statistics Denmark has tested the eye tracking method on ten users who were asked to carry out certain tasks on our website. (Lund), (Wulff 2007). Some of the results were quite persuasive and supported the comments from other users. Others were surprising to us and make us now consider changes in the way information is prioritized on the pages.

B. Different usability test techniques used

7. Different methods to measure accessibility and usability have been carried out in Statistics Denmark: These include:

- (a) Usability test encouraging the users to "think aloud"
- (b) Pop-up web surveys with possibility for users to write comments
- (c) Analysing logs and search results
- (d) Hot line supporter interviews
- (e) Eye tracking

8. In the "*think aloud*" test a test leader sits next to the test person who carries out a task on the web site. While doing so, he/she comments on reflections concerning navigation, concepts and contents. Statistics Denmark considers this a valuable technique that gives a considerable good image of where the users do not follow our logic of thinking.

9. Another regularly used method is *user satisfaction surveys* conducted via pop-ups on the web. In general this method does not give much of information concerning which acts should be performed or changed to make the site more accessible. However, when we make the pop-up active very close to the investigated problem we get more - and more relevant - responses. User comments in free text are another source of information.

10. Analysing *log files* and *search results* have given an indication of the difficulties users meet when they come to a web site written in a "foreign language": To many users it is far from obvious what is hidden behind the statistical terms and concepts usually found in our statistics. In this method we have no direct contact with the user and can only guess what the intention behind the search has been.

11. The hotline support collects a lot of valuable information. Telephone calls and e-mails from users who have failed to reach the required information could be used a lot more systematically than is the case in Statistics Denmark. The hot line staffs are often guiding the users through the site and could through a simultaneous interview follow where the user is trapped.

12. Finally, in 2006 the usability testing using the eye tracking technique has been used on Statistics Denmark's web site, www.dst.dk and the on-line databank www.statbank.dk.

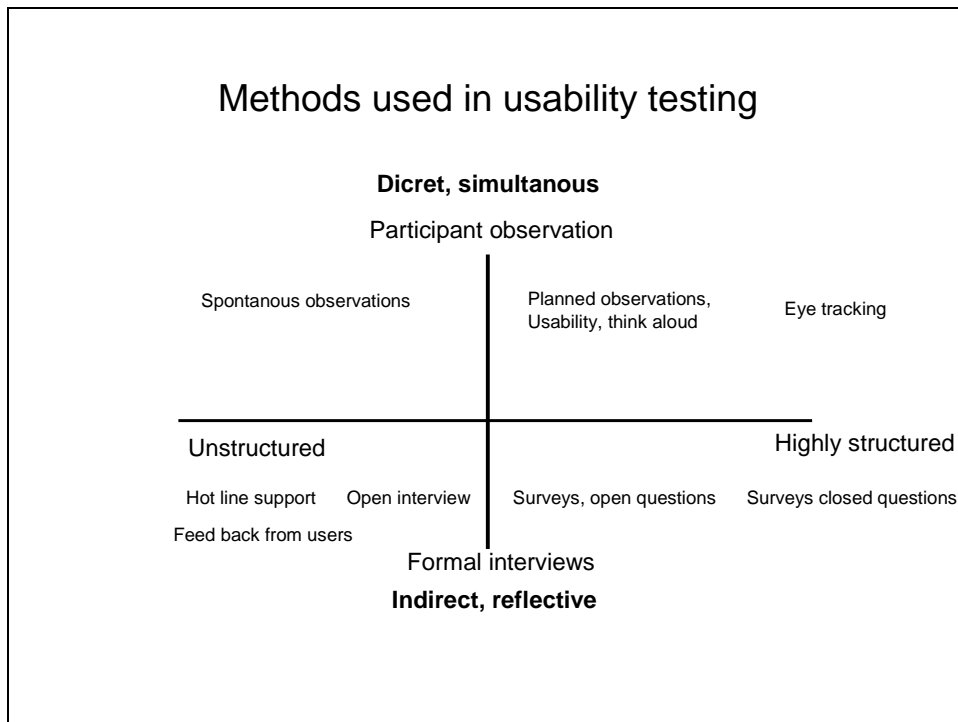


Figure 1 Characteristics of the different methods gathering information about user behaviour.

C. About the eye-tracking technique

13. The equipment used in the eye tracking test was a Tobii 1750 connected to a Dell Precision 370 equipped with a 3.2 Ghz Intel Pentium CPU and 2 Gb Ram. It was built into a 17 " flat screen with a high quality web camera and light sources of "Near infra-red light". The infrared light is reflected in the eye's cornea and pupil. The combination of a reflection in the dark pupil with the reflection in the cornea is used to calculate the focus of the test user's eyes. It is possible to eliminate the movements of the head. Before start the system is calibrated to match each test person's eyes.

14. The results are shown in "real-time videos" and in "gaze replays". They consist of recorded gaze plots that display the fixations of the eyes and scan paths. A scan path can be understood as a "view route". Another way to show the results is "hot spots" some times also called "heat maps". These give an accumulated picture of all test persons fixation points in one display. The parts where the most people have looked for the longest time are displayed as read areas, while green areas only have been passed at a glance. (The scale goes from red over yellow to green).

15. The test was carried out over three days outside Statistics Denmark by The Royal School of Library and Information Science [Lund, 2006] as a research project. Statistics Denmark had designed the tasks.

D. User navigation

16. A test manager was informed in advance about the intention with all exercises and success criteria were defined for each task. More precisely - we described 10 tasks to be solved. These tasks included some rather general and commonly used functions on web sites. Others would imply some knowledge of statistical terminology in order to navigate on Statistics Denmark's web site. Also a few more complex tasks were to be solved.

17. The test took place in two parts where the test persons were asked to:

- find information directly on the homepage (www.dst.dk) and use some of the functions available from there (print, search, ..), tasks 1-5.
- extract specific data from the StatBank (www.statbank.dk) and use some of the functions available there (file download, additional documentation, related publications...), tasks 6-10.

E. Some test results

18. Part one of the exercise was a rather simple one. Looking up information on the website was in general possible either through direct links or through the sitemap. Part two - consisting of retrievals from the StatBank - caused more challenges: all though the test person considered she had solved the task it was quite likely that she had not found the correct information. The follow up interviews would give us information on the problems encountered here. These could concern terminology, layout as well as structure of the site.

19. In general the test persons managed to solve the tests. The time spent finishing an exercise, however, varied considerably, for instance in task 5 between 29 seconds and 4 minutes 28 seconds:

Minimum and maximum time needed to solve the tasks:

Time in seconds	task 1	task 2	task 3	task 4	task 5	task 6	task 7	task 8	task 9	task 10
Minimum	0:32	0:17	0:29	0:34	0:29	0:57	1:22	0:28	1:07	1:31
Maximum	1:35	0:53	4:03	3:10	4:28	3:47	3:30	2:50	2:59	4:24
Gave up	1				3		1			1

In a test like this, the users expect there will be a correct solution. So they continue looking around. In real life some users would probably have given up earlier. Three to four minutes is a long time on the web.

20. Example 1:

Find information about changes in the local authority structure. Print the page. (task 1)

The information is directly available from the homepage at two places: (A),

(B)

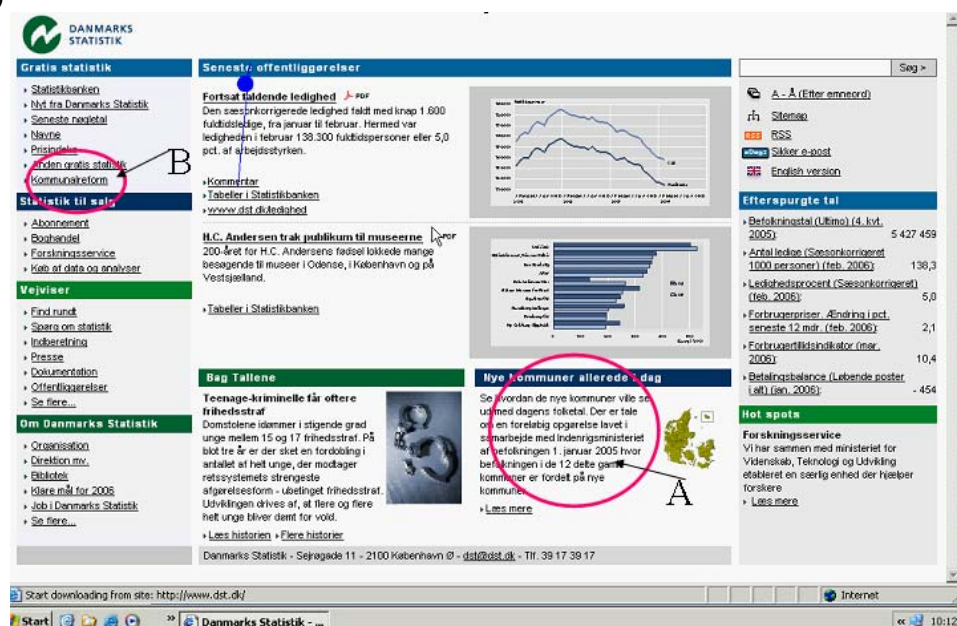


Figure 2

The picture in figure 2 is the first glimpse from a video and it shows the first fixation to be in the upper middle part and going downwards (see the blue dot and trace).

21. Four students solved the task by selecting the link with the map (A), three selected the left column (B). The rest chose to go via *Sitemap* in the upper right column. When it came to printing the information eight used the print icon on the site while two used the print icon in the browser. The Hot-spot map from the same task gives the result, where all students' actions are aggregated.

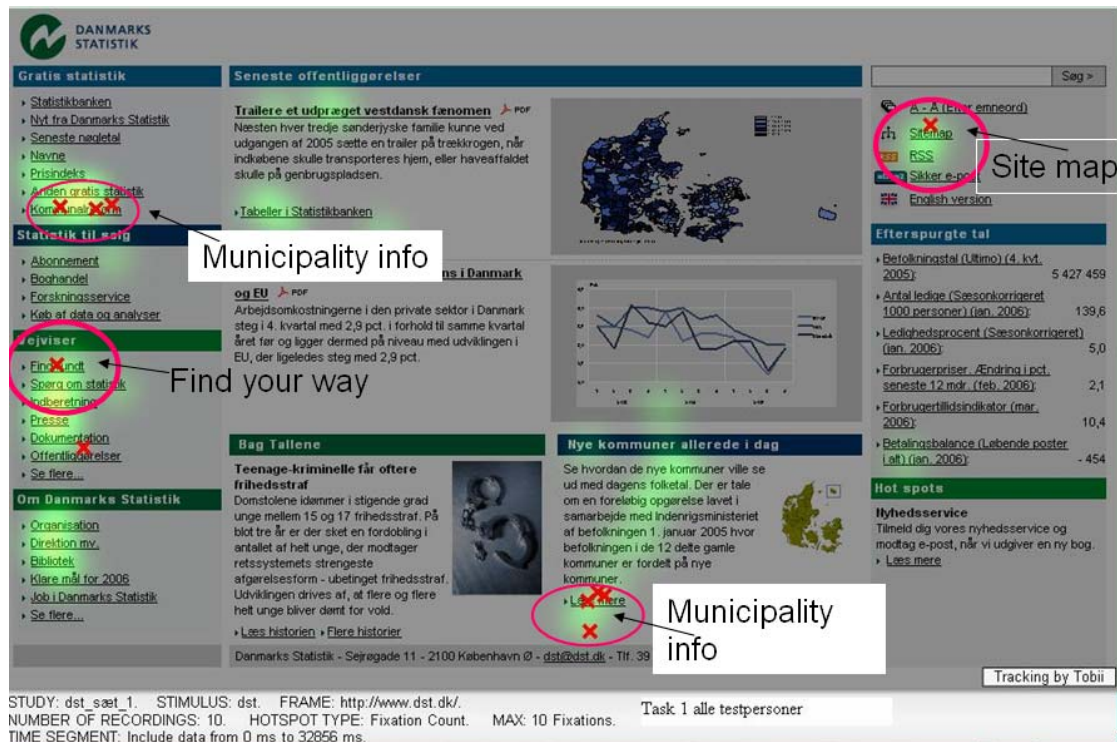


Figure 3, Task 1 Information about the municipality reform.

22. The pattern is similar when it comes to the more complex exercises. The video films show that a user can pass the correct link several times without actually noticing it. Some test persons do not even observe the relevant links. They choose to go to *Find your way* or *Sitemap*.

23 Example 2

The second example is from the StatBank. The task was to find the number of Swedish citizens in Denmark. Navigation here is by topic in the left side column. Swedish citizens are placed in the top placed topic, *Population*. The test persons read through the whole list of topics before returning to the top topic and click *Enter*. The large red spot in the left side shows that. Again, several eyes reach the upper middle and right part of the screen visualized by the yellow spots. The log on procedure in the down left corner is optional. Nobody seems to consider logging on - which they do not need to either.

24 Except from this first page of the StatBank the Heat map display does not work very well on pages with pop-ups. The version of the analysing tool, Clear View, that we used did not compensate for scrolling, frames and pop-up windows. The aggregation of test person results into a heat map was useless in such cases. On the other hand the videos recorded and displayed the navigation all right.

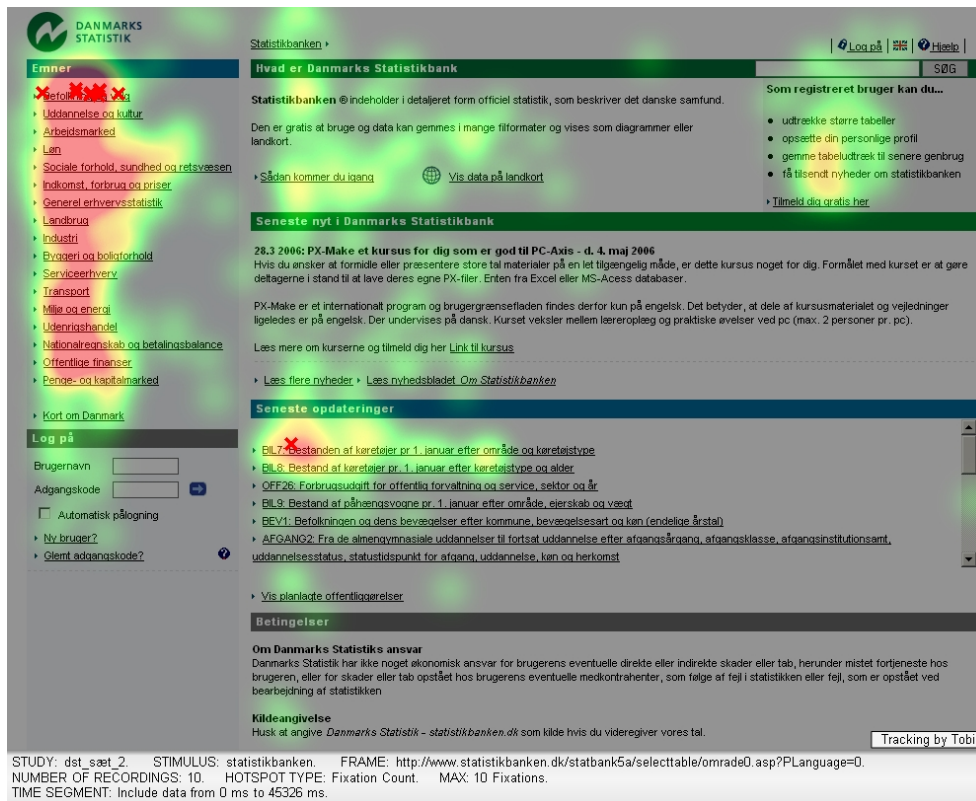


Figure 4 Task 6, Selecting a topic from the StatBank.

III. Evaluation and lessons learned

A. Evaluation of the method

25. When we compare the eye tracking method and the formalised “thing-aloud-tests” we do not find a lot of differences. A traditional participant observation method combined with concurrent think aloud and a follow-up interview produces partly the same results as the inclusion of eye tracking. Never the less, we postulate that the eye tracking has some advantages:

26 All though the eye tracker tool is some technical equipment it is discretely integrated into a monitor without any visible or moving “tracking devices”. Test persons are allowed to move freely in front of the screen. This non-intrusiveness ensures that respondents behave naturally. The test person does not feel he is over looked. Thus he does not feel he reveals any incompetence.

27 It allows for collecting eye tracking data simultaneously with standard test procedures, for instance think-aloud usability testing or follow-up interviews. This was also done in our case. The first part of the test (tasks 1-5) was followed up by an interview where the test persons draw the the recalled navigation on a sheet of paper. There was a rather good agreement between the actually eye movements and the persons understanding of behaviour. When there was a disagreement it was usually an opinion that she did “better” than the eye movements showed.

28 Documentation is, however, one of the big advantages of the eye tracking method. The results are so easily shared between team members working with the web as well as a convincing document to the managers. Web sites lay out is often an area where individual opinions of taste fight each other. The videos and the heat maps help us to focus on how users tend to see and interpret the site.

B. Some lessons learned

29. The test has confirmed several assumptions about how users navigate on the site – and also demonstrated some erroneous assumptions. Firstly, there seems to be some weak points in the structure of the general home page www.dst.dk. The interviews showed difficulties in managing the left side column that is used in navigation. It is very compact with a lot of links grouped in four categories:

- Statistics free of charge
- Statistics for sale
- Guidance to statistics
- Organisation

30. Within each of these groups 4-7 links are visible from the front page. For instance *StatBank*, *Researchers Forum*, *Book shop*, *Management*,...while others are even one more click down hidden behind a *See more* link. This is a solution that was selected in order to avoid scrolling on the front page. A user friendliness that apparently does not work very well.

31. To some users it is not evident what is behind the groups in this left hand column of the page. None of the test person chose the *See more* link. Most visitors are looking for statistics; statistics within a specific topic. However, the structure on the site is “product” oriented: *Statistics for free*, *StatBank*, *News releases*, *Statistics for sale* etc. The sheet anchor is the *Sitemap* and *How to get around*: from there users get an overview of the possibilities on the site. The heat maps indicate that nobody looks in the upper left corner where we have placed the logo. Tests from other web sites show a similar experience.

- (a) In a future layout revision of dst.dk we should consider a home page based on subject areas.
- (b) Search results should be presented in a way where the content is obvious from the very start of the line.
- (c) Hidden alternatives (*See more* links) in the navigation are relatively useless. What we want the user to see should be visible and comprehensible from the first page. Giving meaningful labels to the links is the be-all and end-all of the usability.
- (d) Do not place anything important in the right side column. Only the upper part of the right side is noticed
- (e) Is there a need for more help info or is there a need for more simplicity? Maybe much of the information is entered onto the pages in order to give the producer a good conscience while it disturbs the clarity for the users? A bad structure or layout cannot be repaired by extending the texts.

IV. CONCLUSIONS

32. Is the Eye tracking tool second to none or is it just another way of aloud thinking with the eyes? In our test we followed the eye tracking up with an interview where the test person described her choices of navigation by memory. All though some details were lost in the interview the agreement was dominating.

33. We find that the two methods make a richer understanding when complementing each other. The explanatory remarks from the interview and the convincing picture that a heat map presents makes in one glimpse a very good overview of reactions of a larger group of users.

References

Lund, Haakon (2006). *Brugertesting via Eye-tracking*, The Royal School of Library and Information Science, Copenhagen 2006

Sundgren, Bo & Steneskog Gösta: *Information Systems for Concerted Actions*. In Sundgren, B., Mårtensson, P., Mähring, M. & Nilsson, K. (Eds.) (2003). *Exploring Patterns in Information Management: Concepts and Perspectives for Understanding IT-Related Change*, Economic Research Institute, Stockholm School of Economics, Stockholm.

Wulff, Annegrete (2007): *Eyes Wide Shut- or Using Eye tracking Technique to test a Web Site* in International Journal of Public Information Systems, IJPIS 2007:1, (www.ijpis.net).

Wulff, Annegrete (2006). *User satisfaction survey of StatBank 2001-2006*, www.dst.dk/usersurveys, Danmarks Statistik/Statistics Denmark.

Wulff, Annegrete (2002). *Usability test af Statistikbankens grænsesnit*, Danmarks Statistik/Statistics Denmark.
