DET NATURVIDENSKABELIGE FAKULTET
KØBENHAVNS UNIVERSITET

Usability-evaluering af NemID

- en belysning af autentifikation i borgervendte it-systemer

Kandidatspeciale af
Anders Bjerg Pedersen
Marts 2011

Datalogisk Institut
Københavns Universitet
Vejleder: Erik Frøkjær

Forord

Dette speciale er udarbejdet i perioden fra den 20. september 2010 til den 20. marts 2011 på Datalogisk Institut ved Københavns Universitet.

Emnet for specialet er udsprunget af en artikel for godt et år siden på dr.dk, der ganske kort og anonymt fortalte om NemID – den kommende afløser for den digitale signatur, som skulle lanceres inden for få måneder. Artiklen påpegede, at meget få danskere overhovedet kendte til denne nye løsning og de konsekvenser, det ville få for dem. Jeg måtte erkende, at jeg var én af disse. På denne baggrund besluttede jeg at kaste mig ud i at se nærmere på denne næste mastodont inden for offentlig IT i Danmark. Kunne den modsat så mange andre kuldsejlede projekter faktisk gå hen og blive en succes? Alene navnet forpligtede! I løbet af projektet er fokus mere og mere drejet over på slutbrugerens indtryk og holdninger til produktet NemID, fremfor at rette fokus på de mere tunge aspekter såsom teknik eller sikkerhed. Disse overvejelser er mundet ud i denne afhandling, som for mig er endt med at blive væsentlig mere spændende og lærerig, end jeg havde turdet håbe på.

Der er mange mennesker, som jeg uden gråd i stemmen og forgyldte statuetter vil tillade mig at sige tak til. En stor tak til min vejleder Erik Frøkjær for at turde begive sig ud på tur med en indenfor usability uprøvet sønderjysk knægt – de mange hyggelige møder har i den grad hjulpet! Desuden skal der lyde en stor tak til min kæreste, min familie og mine venner, som har gidet høre mig fortælle om ord med alt for mange stavelser. En særlig tak skal dog lyde til de 161 testpersoner og respondenter, som tog sig tid til at deltage i mine undersøgelser, samt til Peter Lind Damkjær fra DanID for at rive en time ud af sin travle kalender. Sidst, men ikke mindst, tak til Kasper Hansen for stor hjælp til tests og efterbearbejdning, samt til mit yndlings-statistikerpar for at lade sig bestikke til at hjælpe mig godt igang med denne vigtige del.

København, marts 2011

Anders Bjerg Pedersen

Indhold

Bilag (se seperat hæfte)

Resumé

Omdrejningspunktet for dette speciale er gennemførselen af en usability-evaluering af NemID, afløseren for den digitale signatur. Vi indleder afhandlingen med en gennemgang af både den digitale signatur og NemID i forhold til særligt teknik, sikkerhed, udbredelse og brugervenlighed, ligesom vi undersøger baggrunden for deres indførsel. Herefter udvælger og gennemgår vi teori og metode for tænke-højt-tests, *cooperative usability testing* samt spørgeskemaer og anvender denne viden til at planlægge og gennemføre både kvalitative tænkt-højt-tests og kvantitative online spørgeskemaundersøgelser af og om NemID. Resultaterne analyseres herefter grundigt og relateres desuden til lignende relevante undersøgelser på området. Efter analysen diskuterer vi resultaterne og konkluderer på baggrund heraf, om NemID kan siges at være eller at blive et succesfuldt offentligt IT-projekt. Resultaterne viser, at der i et forholdsvis gennemtestet produkt som NemID stadig er en del usability-problemer, hvoraf for mange må anses som kritiske, ligesom testpersonernes vurdering af systemet er middelmådig og antyder mangel på forståelse af de underliggende koncepter. Der er desuden udbredt utilfredshed med det omdiskuterede OTP- (one-time password) nøglekort, som er den vigtigste faktor i at gøre NemID sikrere end sin forgænger. Vi konkluderer på basis af ovenstående, at NemID på visse punkter kan siges at være en (tvungen) succes, mens der på andre punkter er brug for mere arbejde.

Abstract

The pivot point of this thesis is performing a usability evaluation of NemID, the successor to the Danish public digital signature. We open the thesis with an examination of both the original digital signature and NemID with particular focus on technical mode of operation, security, propagation, and usability, as well as an investigation of the incentives to introduce them. We then select and examine theory and practice of think-aloud testing, *cooperative usability testing*, and questionnaires and use this knowledge to carefully plan and conduct both qualitative think-aloud tests as well as quantitative online questionnaires on NemID. The results are thoroughly analysed and related to similar investigations into the subject. Following the analysis we discuss the results and on the basis of these conclude whether or not NemID can be said to be or become a succesful public IT project. Results show that in a thoroughly tested product such as NemID a lot of usability problems can still be identified, including too many classified as critical, just as test subjects' assessments of the system are mediocre at best and indicate a lack of understanding of the underlying concepts. Also, a widespread dissatisfaction is seen with the much discussed cardboard OTP (one-time password) keycard which is the single most important factor in making NemID more secure than its predecessor. Based on these findings we conlude that in some respects NemID can be seen as a (forced) success, whereas in others much work is still needed.

1 Indledning og motivation

I 2003 blev den digitale signatur indført i Danmark. Den markerede en landvinding inden for autentifikation og signering på det hastigt ekspanderende internet. Forventningerne til den digitale signatur var store, omend dens potentiale ikke på daværende tidspunkt fuldt ud kunne overskues. Den digitale signatur er sidenhen blevet klandret for sin til tider kryptiske installation, manglende kompatibilitet, manglende forståelighed af de bagvedliggende koncepter samt en skuffende udbredelse og manglende accept fra centrale offentlige og private spillere på markedet for digital selvbetjening. Netbankerne har kørt med proprietære systemer, og de offentlige myndigheder har enten ikke været gearet til at håndtere digital selvbetjening eller har tilbudt 4-5 forskellige login-metoder – her var digital signatur blot én blandt mange og hvorfor så skifte, når nu de andre virkede fint?

Vi skriver nu 2011, og NemID er blevet lanceret. Denne afløser for den digitale signatur er blevet præsenteret som en mere brugervenlig og sikker løsning, der samler netbanker, den offentlige forvaltning og deltagende private virksomheder under ét fælles login-system – tilsyneladende til gavn for både virksomheder, offentlige instanser og frem for alt slutbrugerne. NemID har dog haft en svær fødsel og er i medierne blevet sablet ned for gentagne svipsere i sikkerheden og kompleksiteten af systemet, samt en generel utryghed omkring selve systemet. Men navnet forpligter: NemID er *nødt* til at være nemt! I dette speciale vurderer vi denne præmis og forsøger i sidste ende at svare på, om NemID bliver en succes eller ej – primært set ud fra 'den almindelige slutbrugers' synsvinkel. Vi tager med andre ord et grundigt kig på kerneydelserne i systemet og ser på, om disse er til at finde ud af; hvorvidt de udviser god *usability*.

Specialets målsætning I specialet sammenligner vi indledningsvist både den gamle digitale signatur og det nye NemID i forhold til faktorer som teknik, brugervenlighed og udbredelse. Vi undersøger også kort motivationen for overhovedet at udvikle en ny løsning, samt sætter kort satsningen i forhold til de ønskede mål og den aktuelle IT-politik på området. Vi inddrager her relevante juridiske dokumenter, evalueringsrapporter og interviews med aktører på området. Hovedbidraget består af to forskellige slutbrugertests af NemID, med sigte på hver sin målgruppe. Baseret på analysen af disse undersøgelsers resultater forsøger vi at komme med et bud på, om og hvorfor NemID kunne tænkes at blive et hårdt tiltrængt *succesfuldt* offentligt IT-projekt. Der vil i specialet blive lagt stor vægt på metodiske overvejelser samt grundig efterbearbejdning af data fra de forskellige undersøgelser. Hvor det er muligt, vil vi forsøge at relatere og perspektivere resultaterne til den brede befolkning, og i diskussionen vil vi sammenholde vores egne resultater med lignende undersøgelser, særligt de af DanID udførte prototypetests.

Afgrænsning af emne I forhold til den oprindelige problemformulering er der lagt mindre vægt på de politiske sider af indførelsen af både den oprindelige digitale signatur og NemID. Det har desværre ikke været muligt at skaffe tilstrækkelig information om dette emne, og desuden var dette aldrig ét af specialets fokuspunkter. Generelt er specialet heller ikke tænkt som en stringent gennemgang af de tekniske aspekter af begge

loginsystemer, men vi præsenterer detaljer i det omfang, det er passende og nødvendigt for forståelsen af de to systemers virkemåde. Sidst, men ikke mindst, er det i løbet af processen blevet besluttet ikke at relatere opgaver og succeskriterier for tænke-højt-tests til de kriterier for usability af sikkerhedssoftware, som er blevet udredt i et tidligere litteraturstudie.[1] Disse kriterier egner sig langt bedre til f.eks. en heuristisk gennemgang eller et expert review, og det blev i stedet fundet nyttigt at afprøve *Cooperative Usability Testing*-teknikken (som præsenteres i afsnit 4.1.2) og dennes resultater.

Specialet indeholder bevidst ikke noget indledende litteraturstudium af borgervendte autentifikationssystemer. Dette er fravalgt, da der for det første er tale om et speciale omhandlende et system, som ikke tidligere er blevet evalueret, og for det andet finder vi det mere kontruktivt og læsevenligt at inddrage litteraturen, der hvor den anvendes – i specialets metodiske og analytiske afsnit. Her er til gengæld lagt vægt på at basere argumentation og analyse på veldokumenteret, videnskabelig funderet teori.

NemID er meget mere end blot login og signering af transaktioner, og DanID har netop lanceret sikker krypteret email, medarbejdersignaturer er lige om hjørnet, og desuden er der hele aspektet med tilslutningen af private virksomheder til systemet. Fokus i dette speciale vil dog komme til at ligge på slutbrugerens anvendelse af de helt basale funktioner: aktivering, brug og holdninger til NemID.

Specialets opbygning I afsnit 2 præsenterer vi kortfattet baggrundsviden om den oprindelige digitale signatur. Afsnittet vil primært beskrive virkemåden og sikkerheden af signaturen, men vi kommer også kort ind på brugervenlighed og udbredelse. I afsnit 3 gør vi det samme for NemID, blot går vi her mere i dybden og inddrager interviews og materiale fra DanID i processen.

Afsnit 4 tager sig af de metodiske overvejelser og litteratur hørende til de benyttede brugertests af NemID. Vi gennemgår tænkt-højt-tests, kooperative evalueringsteknikker og spørgeskemaer og illustrerer undervejs fordele og ulemper ved de enkelte metoder. Vi beskriver desuden opbygning og overvejelser omkring de konkret udførte tests, ligesom vi indsnævrer og udvælger målgrupperne til de forskellige undersøgelser.

I afsnit 5 bearbejder og analyserer vi grundigt de resultater og data, som er indsamlet ved de udførte brugertests. De fundne usabilityproblemer identificeres, klassificeres og grupperes, og data fra de to spørgeskemaer analyseres, bearbejdes og illustreres. I afsnit 6 diskuteres de førnævnte resultater og relateres til lignende undersøgelser, og i afsnit 7 konkluderer vi overordnet på resultaterne af specialet.

Der er til specialet knyttet et meget omfangsrigt bilagsmateriale, som blandt andet indholder udarbejdet materiale til de udførte undersøgelser (testplaner, spørgeskemaer, udleverede opgaver, etc.), de indsamlede og bearbejdede data samt diverse interviews og relaterede undersøgelser. En liste over inkluderede bilag kan ses i umiddelbar forlængelse af indholdsfortegnelsen.

[1] Mere specifikt tænkes her på eksempelvis Yee [41].

2 Den digitale signatur, 2003-2012

I dette afsnit skal vi se nærmere på den 'gamle' digitale signatur, som blev igangsat i 2002 og indført i Danmark i 2003, og som bliver udfaset gradvist indtil udgangen af 2012 til fordel for afløseren NemID. Denne er specialets omdrejningspunkt og vil blive præsenteret i afsnit 3. Vi redegør kort for den digitale signaturs virkemåde, teknik og brugervenlighed samt opridser indledningsvist baggrunden for dens indførsel. Vi ser desuden på anvendelser og kendte trusler mod sikkerheden i signaturen. Formålet med afsnittet er at give et billede af det system, som NemID skal afløse, samt introducere de tekniske termer, som er nødvendige for at forstå virkemåden af begge systemer.

2.1 Baggrund og politiske mål

Digitale signaturer var ikke noget nyt begreb i 2003, men deres udbredelse i Danmark var meget begrænset. Loven om elektroniske signaturer (Ministeriet for Videnskab, Teknologi og Udvikling [19]) fra 2000 må ses som det politiske grundlag og blåstempling af det arbejde, som Telestyrelsen påbegyndte med den fælles digitale signatur og OCES-certifikaterne (Offentlige Certifikater til Elektronisk Service). Loven har som formål at *"[...] fremme en sikker og effektiv anvendelse af elektronisk kommunikation gennem fastsættelse af krav til visse elektroniske signaturer og til nøglecentre, der udsteder certifikater til elektroniske signaturer"* ([19], §1) og er stadig gældende idag. Eftersom online selvbetjening og handel så småt var ved at blive mere udbredt, var der et behov for en fælles standard for sikkerhed og autentifikation. Den stigende digitalisering af den offentlige sektor og dennes behov for sikker kommunikation fremsatte også øgede krav til sikringen af kommunikation over internettet og email generelt.

IT- og Telestyrelsen [12] peger på fordele ved den digitale signatur i form af bedre økonomi og bedre service. For forbrugerne nævnes øget tilgængelighed af de offentlige myndigheder som en fordel i forbindelse med sikker elektronisk kommunikation. Borgeren kan hermed komme i kontakt med myndigheden 24 timer i døgnet, da den digitale signatur er garant for både krypteret, signeret og autentificeret kommunikation. Det nævnes desuden, at digital kommunikation kan være med til at *"nedbryde nogle af de mentale skranker mellem myndighed og borger"*. I forhold til elektroniske blanketter ses det ligeledes som en fordel, at der kan hjælpes mere undervejs. Sidst, men ikke mindst, kan digital signatur øge antallet af logon-tjenester, som brugeren har adgang til døgnet rundt og hele tiden har mulighed for at tilgå opdateret og redigérbar information fra. Tilgængelighed og sikkerhed er altså nøgleordene for borgerne.

Udover de umiddelbare fordele for brugerne af digital signatur er der for offentlige myndigheder også store økonomiske gevinster at hente i form af besparelser på interne postgange, porto og konvolutter, mindsket antal personlige henvendelser og mindre arbejde med at fremsøge informationer, samt en generel mindskning af sagsbehandlingstider grundet mindre journalisering, indscanning og brevveksling. Det er i forbindelse med de listede fordele interessant, at mange ikke direkte er relateret til indførelsen af digital signatur som sådan – et hvilket som helst andet autentifikationssystem havde kunnet give næsten samme fordele.

I beskrivelsen af OCES-standarden (se IT- og Telestyrelsen [11]) er det angivet, at

baggrunden for indførelsen af en fælles offentlig standard for elektroniske certifikater bunder i ønsket om øget interoperabilitet og en mere bred anvendelighed blandt borgerne. Særligt nævnes kravet om, at borgeren personligt skal møde op for at kunne få udstedt et certifikat som en begrænsende faktor i udbredelsen.[2]

2.2 Teknik og virkemåde

Vi vil i dette afsnit kort gennemgå teknikken bag og funktionerne i den digitale signatur. Vi baserer primært vores oplysninger på IT- og Telestyrelsen [10], suppleret med Schneier [33]. Afsnittet er medtaget for at opnå tilstrækkelig forståelse for de koncepter og begreber, som vi vil forudsætte kendt i senere afsnit.

Kryptering og signering

Generelt indføres en PKI (*Public Key Infrastructure*, eller i folkemunde blot digital signatur) for at sikre følgende fire faktorer i kommunikationen mellem to parter, f.eks. en borger og en offentlig myndighed ([10], s5):

- **Autenticitet:** Autencitet sikrer modtageren af en meddelelse (f.eks. en email), at afsenderen er den person, som vedkommende giver sig ud for at være. Dette aspekt er særlig vigtigt, når kommunikationen foregår elektronisk, da der her ikke umiddelbart er mulighed for visuel eller fysisk identifikation. Den digitale signatur er her en slags erstatning for et id-kort eller pas.
- **Integritet:** Integritet sikrer, at den modtagne meddelelse er uændret i forhold til den form, den blev afsendt i. Hvis Alice sender en besked til sin bank om at overføre 100 kroner til Bob, må Bob eller andre ikke kunne gå ind i beskeden og ændre dette til 1000 kroner. Integritet er altså en sikring mod eventuelle mellemmænd, der kunne have interesse i at foretage ændringer i de afsendte beskeder, og kan tjene som verifikation af, at data er intakt.[3]
- **Fortrolighed:** Et særdeles vigtigt aspekt i digitale signaturer er deres mulighed for at sikre kommunikationen mellem to parter mod udenforstående overvågning. Her sikres altså, at det kun er afsender og modtager af en meddelelse, som i stand til at afkode eller læse den – konceptet benævnes ofte blot *kryptering*.
- **Uafviselighed:** Selvom denne faktor nok vil have størst praktisk betydning for myndigheder og virksomheder, er den alligevel væsentlig, idet den kan garantere, at en afsendt meddelelse eller givet samtykke ikke kan tilbagetrækkes eller benægtes. I kombination med en tidsstemplingstjeneste og arkiveringsfunktion finder denne faktor f.eks. anvendelse ved bedømmelse af rettidigt (eller for sent) indgivne ansøgninger.

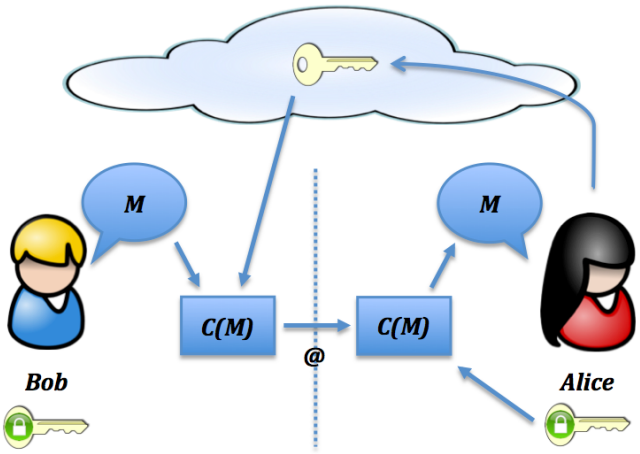
Vi vil nu klarlægge, hvordan disse faktorer sikres i praksis. Uafviseligheden behandler vi ikke videre, da den teknisk er forholdsvist kompleks og for specialet uinteressant.

[2] En fin lille historisk oversigt over den digitale signatur i Danmark kan ses på: http://www.itst.dk/sikkerhed/digital-sikkerhed/digital-signatur/historisk-gennemgang-af-oces-digital-signatur/.

[3] Som er service til mere kompetente IT-brugere offentliggør mange softwareudviklere ofte *checksummen* eller *hash-værdien* af den software, de stiller til rådighed som download. Dermed kan brugeren efter endt download forsikre sig om, at det hentede arkiv er intakt og identisk med det, der måtte forventes.

Grundlæggende skelner man mellem *symmetrisk* og *asymmetrisk* kryptering. Et velkendt eksempel på symmetrisk kryptering er Cæsar-kryptering (eller substitutionskryptering), som ganske simpelt virker ved at tage en tekst og erstatte hvert bogstav med det, som ligger n tegn længere fremme i alfabetet. På denne måde konverteres *kildeteksten* (den oprindelige besked) til en krypteret meddelelse, som for udenforstående vil virke som volapyk. Den afkodes dog nemt på samme måde som kodningen – blot ved at rykke bogstaverne n tegn tilbage igen. Vi anvender altså samme *nøgle* til både kryptering og dekryptering af meddelelsen. Her vil den konkrete udveksling af nøglen mellem afsender og modtager være det primære sikkerhedsproblem.

I asymmetrisk kryptering er krypteringsnøglen forskellig fra dekrypteringsnøglen. PKI bygger på unikke *nøglepar*, hvor den ene nøgle oftest er offentlig, og den anden er strengt privat. Her kan begge nøgler bruges til kryptering, men kun den tilhørende anden nøgle i parret kan bruges til dekryptering. Krypterer man eksempelvis en meddelelse med den offentlige nøgle, er det kun den private nøgle, der kan dekryptere meddelelsen igen. Den grundlæggende krypteringsproces er gengivet i Figur 1, og på denne måde sikres fortroligheden i meddelelsen. Autenticitet er delvist sikret, idet vi for asymmetriske nøglepar har mulighed for at bekræfte, at en besked for eksempel er krypteret med den private nøgle, der hører til den offentlige nøgle, som man har anvendt til at dekryptere meddelelsen. Der er dog ikke nødvendigvis garanti for, at denne offentlige nøgle rent faktisk tilhører den rette person. ([10], s5f)

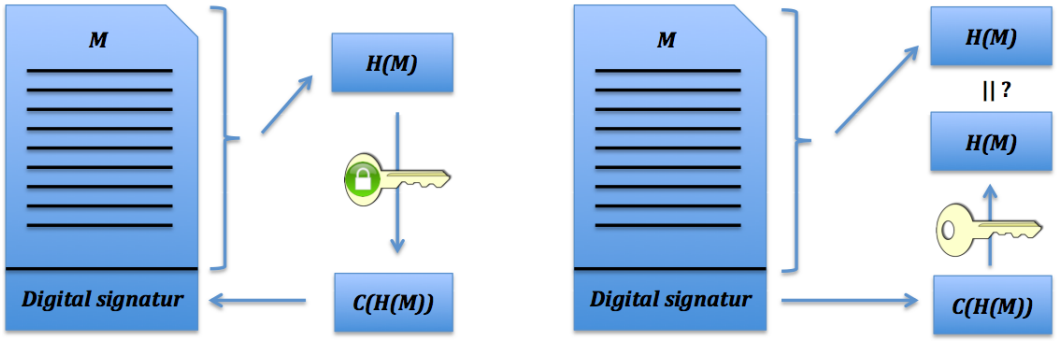


Figur 1: Forsimplet kryptering. Bob ønsker at sende meddelelsen M til Alice og henter derfor hendes offentlige nøgle, som hun eksempelvis har publiceret på en nøgleserver eller sendt til Bob i en email. Bob bruger denne nøgle til at kryptere meddelelsen til kodeteksten $C(M)$, som han sender til Alice. Alice bruger derefter sin private nøgle til at dekryptere meddelelsen igen.

Den private nøgle er elektronisk og opbevares af brugeren, oftest på brugerens PC eller på en USB-stick eller et fysisk nøglekort. Den offentlige nøgle kan være tilgængelig på en offentlig nøgleserver eller hjemmeside eller ligge sammen med den private nøgle og kan så distribueres til folk, der ønsker at sende krypterede meddelelser til én. Der

er hertil forbundet en del væsentlige sikkerhedsaspekter, som vi ser nærmere på i afsnit 2.3. Der findes mange forskellige algoritmer til selve krypteringen, men den klart mest udbredte er RSA, som bygger på særdeles sikre men principielt ikke ubrydelige matematiske algoritmer baseret på primtalsfaktorisering. Se eventuelt mere i Rivest et al. [29]. I virkeligheden vil man dog af hensyn til performance ikke benytte asymmetrisk kryptering til selve meddelelsen, da denne er væsentlig mere beregningstung end symmetrisk kryptering. Man vil i stedet generere en tilfældig, meget stor symmetrisk nøgle, som meddelelsen krypteres med. Herefter vil man kryptere denne nøgle asymmetrisk og vedhæfte den til den symmetrisk krypterede meddelelse.

For yderligere at sikre integritet, kan meddelelsen påtrykkes en *digital signatur*. Denne digitale signatur er mere kompleks at fremstille, og involverer blandt andet hashing-algoritmer, som er i stand til hurtigt at generere et unikt digitalt fingeraftryk af f.eks. en tekststreng.[4] I Figur 2 er gengivet principperne i fremstilling og påtrykning af en digital signatur til en meddelelse. ([10], s7f)



Figur 2: Processen for generering (venstre side) og verificering (højre side) af digital signatur, frit efter figur 2 og 3 i IT- og Telestyrelsen [10]. Alice genererer en hash-værdi $H(M)$ af meddelelsen, som krypteres med hendes private nøgle. Resultatet $C(H(M))$ vedhæftes den oprindelige meddelelse, som sendes til Bob. Bob udfører nu selv en hashing af meddelelsen, som sammenlignes med hashværdien genereret hos Alice, der er blevet dekrypteret med Alices offentlige nøgle. Stemmer de to overens, er der garanti for meddelelsens autenticitet og integritet.

Bemærk, at digital signering i sig selv ikke krypterer selve meddelelsen – der er ingen sikring af fortroligheden. Kombinerer man processerne i Figur 1 og 2 ses det dog tydeligt, hvordan en meddelelse kan krypteres *og* signeres i samme ombæring.

Aktører og certifikater

Som nævnt sikrer ovenstående processer integritet og fortrolighed af meddelelsen, hvorimod autenticitet kun er sikret på nøglebasis. Hvordan ved Alice med sikkerhed, at meddelelsen er kommet fra Bob? Her kommer *certifikater* på banen. Et PKI består både

[4] Hashing kan hurtigt og let transformere en tekststreng til en *næsten* unik og entydig værdi af ganske kort længde, mens den anden vej rundt er nærmest umulig. Med næsten mener vi, at det er særdeles usandsynligt at to forskellige input giver samme hash-værdi. Se http://en.wikipedia.org/wiki/Hash_function for flere detaljer. De mest anvendte hashing-algoritmer er MD5 og SHA-1.

af teknologi, som vi har gennemgået ovenfor, samt *aktører*. Vi har allerede præsenteret én slags, nemlig brugerne. En anden meget væsentlig aktør i et PKI er *certificeringscentret* (eller på engelsk *certification authority*, CA), som står for generering og udstedelse af nøglepar til brugerne. Som supplement til dette udsteder og vedligeholder CA'en certifikater for hver enkelt aktør i PKI'et. En brugers eller nøgleindehavers certifikat indeholder foruden brugerens offentlige nøgle også entydige oplysninger om den pågældende bruger, f.eks. fulde navn og adresse, CPR-nummer eller et andet unikt identifikationsnummer. Dette elektroniske certifikat bliver derefter signeret med CA'ens private nøgle, som dermed står inde for sammenhængen mellem brugerens identitet og dennes offentlige nøgle, som er vedhæftet certifikatet. Ved hjælp af CA'ens offentlige nøgle kan ægtheden af certifikatet og dets oplysninger altså verificeres, hvormed meddelelsens autenticitet kan bekræftes. CA'en tilbyder desuden opslagsmuligheder i dette certifikatregister, ligesom den også varetager spærrelister over udløbne eller kompromitterede certifikater. For den digitale signaturs vedkommende var TDC oprindeligt CA, men opgaven blev i juni 2008 overtaget af DanID, som nu også fortsat er eneste CA på NemID. ([10], s13ff)

Der er fra IT- og Telestyrelsens side opsat meget strenge krav til en CA for håndtering, indhold og beskyttelse af borgernes certifikater (IT- og Telestyrelsen [14]). For at sikre konsistens på tværs af tjenester og virksomheder, blev det fra offentlig side besluttet at indføre de såkaldte *OCES-certifikater* (se IT- og Telestyrelsen [11]). De oprindelige OCES-certifikater levede faktisk ikke op til kriterierne i Lov om elektroniske signaturer ([19]) af den simple årsag, at det af udbredelsesmæssige årsager var vigtigt at kunne undgå personligt fremmøde for at få udstedt et certifikat. I stedet bestiltes et certifikat online, og der blev så sendt et særligt link til brugerens emailadresse, samtidig med at en midlertidig pinkode sendtes til vedkommendes folkeregisteradresse. Indtastning af den modtagne kode på hjemmesiden hørende til linket i mailen startede så en generering af brugerens unikke nøglepar og certifikat på serveren og efterfølgende download af den private nøgle til brugerens computer.

Der findes tre typer OCES-certifikater: til privatpersoner, til virksomheder/myndigheder og til medarbejdere i virksomheder/myndigheder. Et OCES-personcertifikat indeholder udover CA'ens signatur og personens offentlige nøgle som standard personens fulde navn eller pseudonym, en landekode samt et personligt identifikationsnummer (PID-nummer), som kan veksles til f.eks. et CPR-nummer i IT- og Telestyrelsens database. Man kan læse meget mere om OCES-certifikater i IT- og Telestyrelsens beskrivelser og retningslinier (se [11] og [14]).

Tjenestebaseret autentifikation

Tilføjelsen af certifikater tillader os at indlemme den sidste væsentlige anvendelse af digitale signaturer: sikker identifikation på online tjenester. Processen, der sættes i værk, når en bruger ønsker at identificere sig overfor en tjeneste/server over internettet er enkel og hurtig, når ovenstående begreber og teknikker er på plads: ([10], s11f)

1. Serveren sender et meget stort, tilfældigt tal til brugeren.

2. Brugeren krypterer det modtagne tal med sin private nøgle og sender resultatet sammen med sit offentlige certifikat retur til serveren.

3. Serveren verificerer ved opslag hos CA'en, at det modtagne certifikat er gyldigt, ikke er udløbet og ikke er spærret.

4. Til sidst dekrypterer serveren det returnerede tal med brugerens offentlige nøgle, som er indeholdt i certifikatet. Matcher tallet det oprindeligt udsendte, kan serveren antage, at brugeren er den, vedkommende udgiver sig for at være.

5. Ved den omvendte proces kan brugeren verificere identiteten af serveren.

Selve forbindelsen mellem bruger og server vil undervejs sædvanligvis være krypteret, som regel ved hjælp af HTTPS-protokollen.

For den digitale signaturs vedkommende foregår denne proces ved, at der på den pågældende tjenestes hjemmeside er en Java-baseret fildialog, hvor man skal angive stien til sin private nøglefil på computeren. Denne skal så eventuelt låses op af det tilhørende password, inden den kan anvendes. Herefter foregår processen som nævnt ovenfor. I Figur 3 ses et eksempel på et sådant login, her taget fra det fællesoffentlige login-system.

Figur 3: Login med den gamle digitale signatur. 'Gennemse...'-knappen åbner en standard fildialog, hvor brugeren skal navigere frem til sin nøglefil, der kan ligge på PC'en eller på et eksternt medie. Herefter skal kodeordet hørende til signaturen tastes ind.

2.3 Sikkerhed og trusler

I dette afsnit skal vi kort se på sikkerhed og trusler mod digitale signaturer som helhed, og igen vil vi nøjes med at præsentere det, som er relevant for den videre forståelse af projektets resterende afsnit.

Når man taler om autentifikation, kommer begrebet *n-faktorsikkerhed* til at spille en stor rolle. Man skelner her mellem forskellige faktorer, som er med til at beskytte f.eks. en brugers identitet mod tyveri. Et eksempel på *et-faktorsikkerhed* er f.eks. brugen af brugernavn og password – her er der tale om noget, som brugeren *ved*. Selvom der er to individuelle faktorer (brugernavn og password), omtales de under et, da de begge tilhører samme kategori. I *to-faktorsikkerhed* taler man ofte om noget, som brugeren *ved*, og noget som brugeren *har*. Dette kunne f.eks. være kombinationen af et password og en elektronisk kodegenerator eller et id-kort. En tredje faktor, som nogle gange finder anvendelse, er noget, som brugeren *er* – f.eks. i form af et fingeraftryk, et iris-scan

eller et stemmemønster. Et almindeligt kreditkort er sædvanligvis to-faktorsikret, idet der både er selve kortet og den tilhørende pinkode. Den gamle digitale signatur bliver ofte omtalt som 'halvanden-faktor', idet en elektronisk nøglefil ikke bør anses som noget fysisk, brugeren er i besiddelse af.[5]

En oplagt sikkerhedsfaktor, man umiddelbart kan se på, er sikkerheden af den udførte kryptering. Denne afhænger i høj grad af størrelsen på de valgte nøgler, som defineres i bits og angiver, hvor mange forskellige nøgler der kan genereres. RSA i den digitale signatur benytter sig af 1024 bits nøgler, mens den i 2010 største brudte nøglelængde var 768 bits, som blev foretaget med state-of-the-art distribueret udstyr. Der er altså stadig et godt stykke op til de 1024, men enkelte eksperter forudser, at nøgler af denne størrelse *måske* vil kunne brydes inden for en overskuelig årrække. Selve krypteringen (og dermed også signeringen) er altså stadig forholdsvis sikker. Schneier [33] påpeger dog i tråd med mange andre, at nøglelængde i sig selv ikke giver et fuldstændigt billede af graden af sikkerhed i digitale signaturer og kryptering generelt, da man også er nødt til at se på, om systemerne for eksempel udnytter nøglerummet fuldt ud.

Da brute-force ikke umiddelbart er vejen frem for at bryde kryptering, kan man i stedet se på de alternative måder som hackere bruger til at omgå at skulle bryde krypteringen. Man skelner ofte mellem to typer angreb:

- **Online-angreb** foretages, samtidig med at brugeren anvender sin digitale signatur. Man prøver med andre ord at interagere med brugeren og eventuelt opsnappe eller forpurre den information, der flyder mellem brugeren og den tjeneste, brugeren forsøger at tilgå.
- **Offline-angreb** kan derimod foretages, uden at brugeren behøver være online. Et eksempel er at sætte en computer til forsøge alskens forskellige kombinationer af brugernavne og passwords til en bestemt tjeneste – et klassisk eksempel på brute-force.

I den første kategori hører blandt andet man-on-the-wire-angreb (MOTW-angreb), hvor en hacker "lytter med på linien" mellem bruger og server. Hvis denne mellemmand kan opsnappe det store tal sendt fra serveren *og* det retursendte certifikat og krypterede tal, kan han overtage brugerens login. Brugeren vil så sandsynligvis blot få en fejl. Man-in-the-middle-angreb (MITM-angreb) ligner meget, men her befinder mellemmanden sig 'fysisk' mellem brugeren og serveren og overtager og styrer hermed kommunikationen mellem dem. Denne metode bliver dog besværliggjort en smule af eventuel logning af IP-adresser, timeouts og lignende forholdsregler fra serveren, ligesom mellemmanden skal kunne udgive sig korrekt for at være både bruger og server på præcis det rette tidspunkt. Desuden skal mellemmanden være på samme netværk som enten brugeren eller serveren – noget som dog forholdsvis let kan lade sig gøre med åbne trådløse netværk.

Den optimale situation for en hacker er dog at kunne overtage brugerens computer og dermed få adgang til brugerens private nøgle direkte. Da denne ofte er beskyttet af et password, skal dette dog først brydes, men her er der tale om et offline-angreb og en mere overkommelig opgave. Ligger nøglen f.eks. på en ekstern disk, må hackeren tålmodigt

[5] Se eventuelt mere på http://en.wikipedia.org/wiki/Two-factor_authentication

vente, til brugeren forsøger at logge sig på en tjeneste (nøglen skal i så fald gøres tilgængelig for computeren) og herefter overtage kommunikationen helt. DanID anslår, at der er i omegnen af 100.000 inficerede computere i Danmark, der tillader sådanne overtagelser eller overvågning.

Den nok største trussel mod et PKI er tiltroen til, at brugeren kan opbevare den private nøgle forsvarligt og sikkert. Så snart den private nøgle er blotlagt for omverdenen, er der dybest set fri adgang for hackere til de oplysninger, som nøglen giver adgang til (vel at mærke efter at have brudt passwordet). I forlængelse af ovenstående har langt de fleste brugere deres nøgle liggende på deres PC, og er denne inficeret, er nøglen potentielt blotlagt og kan misbruges. Det er derfor ekstremt vigtigt, at brugeren sørger for at holde sin PC opdateret og fri for virus, spyware og lignende ting og i det hele taget udvise forsvarlig online opførsel.

Der findes andre, mere subtile trusler mod RSA og PKI generelt, men disse vil vi ikke se nærmere på her. Har man mod på mere om sikkerhed generelt, kan vi særligt anbefale to bøger, som hver især på deres egen måde giver et billede af IT-sikkerhed. Schneier [33] gennemgår letforståeligt og humoristisk mange af de forskellige typer angreb (og angribere), tager et mere filosofisk blik ud over landskabet og kommer med masser af gode råd til, hvordan man bedst beskytter sig i en moderne, forbundet verden. Pfleeger og Pfleeger [28] ser mere konkret på forskellige typer af sikkerhed, angreb og hvordan man beskytter sig mod dem.

2.4 Distribution og udbredelse

Den digitale signatur blev introduceret i et landskab, hvor den bestemt ikke var den eneste spiller på markedet. Hos SKAT kunne man for eksempel logge ind med digital signatur, TastSelv-kode, den fælles pinkode eller 'autorisation' (hvad det så end var). Noget lignende sås hos de fleste andre offentlige spillere på markedet, ligesom kun få private virksomheder udbød services rettet mod brugen af den digitale signatur. En artikel i Ingeniøren 12. oktober 2007[6] opsummerede situationen efter 4-5 år med digital signatur: kun 4% af danskerne havde taget signaturen til sig og kunne kaldes regelmæssige brugere, på trods af at der på dette tidspunkt var udleveret en knap 900.000 signaturer, hvilket vel at mærke ikke levede op til målet om en million signaturer i juni 2006. Den eneste kommentar til artiklen er skrevet af en ikke helt uinteressant person i denne sammenhæng, Palle H. Sørensen (den nuværende centerleder for Center for Digital Signatur i IT- og Telestyrelsen), og kommer med lidt andre påstande, nemlig at 17% af befolkningen har taget en digital signatur i brug, og at 11% benytter den mindst én gang om måneden – vel at mærke på daværende tidspunkt i 2007. Danmarks Statistik har ligeledes i 2008 foretaget en undersøgelse blandt 1054 brugere af digital signatur.[7] Her angiver 80%, at de anvender signaturen, 92% angiver, at den er sikker eller meget sikker, og 85% er tilfredse eller meget tilfredse med signaturen.

Da den digitale signatur blev indført, kom det nok som en overraskelse for netban-

[6] Se: http://ing.dk/artikel/82239-kun-fire-procent-af-danskerne-bruger-den-digitale-signatur

[7] Her citeret fra http://www.itst.dk/sikkerhed/digital-sikkerhed/digital-signatur/historisk-gennemgang-af-oces-digital-signatur/

kerne, som havde indført deres eget fælles NetID. Noget, som dog sidenhen er gået i sig selv igen. Kritikere angiver den manglende understøttelse af funktioner fra blandt andet kommunernes side som afgørende for den ringe udbredelse og accept af signaturen, selvom den i sig selv beviseligt har sparet det offentlige for anselige beløb i administration.

Det var op til den enkelte borger selv at anskaffe sig en digital signatur. Den blev ikke på nogen måde påtvunget og var anlagt som en frivillig tilbud om at øge borgerens selvbetjening på nettet. Som udgangspunkt bestilte man sin digitale signatur på en hjemmeside, og herefter startede den forholdsvis komplicerede process med postbaseret pinkode, registrering og installation. Vi ser lidt nærmere på disse faktorer i næste afsnit, men de har helt givet været begrænsende for udbredelsen af den digitale signatur. En anden betydningsfuld faktor kan være befolkningens manglende kendskab og forhold til IT-sikkerhed. Noget som er blevet udredt i den såkaldte KFE-undersøgelse [13], som angav, at 46% af danskerne i 2006 havde en ringe forståelse for IT-sikkerhed.

2.5 Brugervenlighed

Eftersom vi i dette speciale primært fokuserer på brugervenligheden af NemID, vil vi her kort skitsere nogle punkter om brugervenlighed for den gamle digitale signatur, som kunne have relevans for sammenligningen med NemID. Dette område er noget af en gråzone, eftersom vi efter at have rådspurgt IT- og Telestyrelsen fik at vide, at der ikke eksisterer evalueringsrapporter for den gamle digitale signatur (de henviste af uvisse årsager ikke til undersøgelsen fra Danmarks Statistik nævnt ovenfor). Vi vil dog illustrere nogle pointer og uhensigtsmæssigheder, der er af mere generel karakter, og som til dels har hæftet sig til den digitale signaturs ry i mange år.

Mange mennesker anser installationsprocessen som en væsentlig byrde ved signaturen. De fleste vejledninger og support til den gamle digitale signatur kan stadig findes på DanIDs hjemmeside,[8] og mange brugere vil sandsynligvis kunne nikke genkendende til den til tider særdeles komplekse proces, man skulle igennem for endelig at kunne anvende sin digitale signatur. DanID har i årenes løb strømlinet processen meget, men et tydeligt tegn på kompleksitet er de hele seks forskellige muligheder for installation på Windows. I de tidlige år var digital signatur slet ikke understøttet på Mac eller Linux, selvom det var muligt at få systemet til at fungere også på disse platforme. Her måtte brugerne yde hjælp til selvhjælp, men vejledningerne og understøttelsen er dog med tiden blevet meget bedre. Selvom de fleste brugere i de seneste par år sikkert er kommet smertefrit gennem installationsprocessen, er denne hængt ved den digitale signatur som en uønsket sygdom.[9]

En anden begrænsende faktor i den digitale signatur er dens manglende portabilitet. I og med at den er baseret på fysiske nøglefiler, som skal beskyttes, har signaturen som udgangspunkt været låst til den computer, den oprindeligt blev installeret på. Igen er det sidenhen blevet muligt at få signaturen på et cd-kort, som man kan tage med sig, og efterfølgende er f.eks. en USB-stick også blevet en mulighed. Ulempen herved er dog så, at sikkerheden potentielt er i fare, da signaturen så skal beskyttes mod tyveri flere

[8] Se: https://danid.dk/export/sites/dk.danid.oc/da/support/
[9] Se f.eks. http://www.version2.dk/artikel/5648-digital-signatur-er-for-boevlet.

steder, selvom nøglefilen så er utilgængelig for eventuelle vira og spyware.

Opsamling Vi har i dette afsnit set nærmere på den digitale signatur, dens virkemåde, udbredelse, brugervenlighed og teknik. Vi har observeret, at den ikke har nået den ønskede udbredelse, primært på grund af manglende understøttelse og kompleks installation, men at den til gengæld må siges at være forholdsvis sikker i nogle år endnu. Vi har ligeledes udredt grundprincipperne for kryptering, signering og autentifikation af kommunikation ved brug af asymmetrisk kryptering og nøglepar og skitseret de potentielle trusler, som et PKI står overfor. I det kommende afsnit skal vi se nærmere på NemID, som lover forbedringer på de fleste af disse områder.

3 NemID, 2010 - nu

I dette afsnit ser vi nærmere på NemID, og ligesom ved behandlingen af den digitale signatur vil der ikke blive tale om en stringent *teknisk eller sikkerhedsmæssig* gennemgang af NemID. En stor del af afsnittet baserer sig på det udførte interview med Peter Lind Damkjær, som er PKI Manager og Senior Security Officer hos DanID, der har udviklet NemID. Interviewet udsprang af manglen på teknisk dokumentation omkring NemID samt en række ubesvarede spørgsmål. Det transskriberede interview samt de stillede spørgsmål kan ses i Bilag P. Formålet med afsnittet er at sætte læseren ind i det system, som resten af specialet handler om og grundigt vil teste. Dette er nødvendigt for at forstå de problemstillinger, som senere vil blive udredt og analyseret.

3.1 Baggrund og politiske mål

I forbindelse med udløbet af regeringens kontrakt med TDC i 2008 om udvikling og drift af den digitale signatur, blev der opstillet et EU-udbud i sommeren 2007, som skulle dække udvikling og drift af den næste generation af digitale signaturer i Danmark. DanID vandt udbuddet, og kontrakten blev underskrevet 25. juni 2008, som efterfølgende har mundet ud i lanceringen af NemID. Udbuddet var blandt andet defineret ud fra en meget omfattende og detaljeret kravspecifikation, som vi venligst har fået udleveret fra IT- og Telestyrelsen. Prisen for udvikling, drift og vedligeholdelse af NemID sniger sig ifølge DanIDs estimater op på ca. 850 millioner kroner over en femårig periode, hvor ca. halvdelen er dækket af bankerne, en fjerdedel af den offentlige sektor, og resten gennem decentral finansiering i form af f.eks. tilslutningsaftaler for private virksomheder. (Ministeriet for Videnskab, Teknologi og Udvikling [20], s14n)

Udover sammenfaldet med udløbet af den tidligere kontrakt har der fra regeringens side været opstillet politiske mål for digitaliseringen af den offentlige sektor. Stat, regioner og kommuner er enige om Globaliseringsstrategiens e2012-mål:[10]

> *"Senest i 2012 skal al relevant skriftlig kommunikation mellem virksomheder, borgere og den offentlige sektor kunne foregå digitalt. For borgerne er der tale om en mulighed, ikke en pligt. I forhold til erhvervslivet er målet, at alle relevante breve, indberetninger, ansøgninger mv. herefter kun kan sendes via digitale kanaler."*

Disse mål hænger sammen med det såkaldte eDag-projekt, som består af tre overordnede milepæle:[11]

- **eDag1 (2003):** E-post anerkendes som officiel korrespondance
- **eDag2 (2005):** Sikker e-post med digital signatur
- **eDag3 (2010):** Nem adgang til det offentlige på nettet. Delmål:
    1. Alle borgerrettede selvbetjeningsløsninger med national udbredelse, hvor der er behov for sikker identifikation, skal alene anvende NemLog-in med Digital Signatur.

[10] Se: http://modernisering.dk/fileadmin/user_upload/documents/Projekter/eDag3/Overordnet_om_eDag3.ppt
[11] Her hentet fra: http://modernisering.dk/da/projektside/andre_projekter/edag3/

2. Alle borgerrettede selvbetjeningsløsninger med national udbredelse skal integreres visuelt i borgerportalen.
3. Alle myndigheder kan kontaktes og svare via dokumentboksen, og alle myndigheder skal afsende alle relevante masseforsendelser via dokumentboksen til tilsluttede borgere og virksomheder.

Her henviser 'borgerportalen' til www.borger.dk, og særligt delmål 1 er interessant, da NemID spiller en væsentlig rolle i opfyldelsen af dette. Dog er NemID som sådan kun en digital signatur og har ikke direkte noget med NemLog-in at gøre: NemLog-in er låsen, og NemID er nøglen.

I Finansrådets regi besluttede man sig for, at man gerne ville have én fælles løsning for både netbanker og offentlige myndigheder og institutioner, men da kravene til løsningerne var en smule forskellige, besluttede man sig for at lave en løsning, som udadtil virkede som én, men i virkeligheden var to:

> *"Det er simpelthen et spørgsmål om, at dengang NemID blev tænkt op, da var bankerne igang med at lave en fælles sikkerhedsmodel. De kunne se, at nøglefilbaserede systemer var under angreb fra især Østeuropa, hvor der blev lagt keyloggere ind, og så stjal de nøglefilen, og så begyndte de at udgive sig for at være den her netbankbruger. Det foregik i regi af Finansrådet, at man ville have en fælles løsning, som var baseret på det, man kalder ægte to-faktor. Lidt efter at de her beslutninger er lavet i bankverdenen, så kommer det offentliges digitale signatur i udbud igen, fordi den oprindelige aftale var baseret på et EU-udbud, og så har man så brugt lejligheden til at sige "hvad kan vi så risikere, at den her infrastruktur, vi har nu, kan blive udsat for; hvad er det for nogle erfaringer, vi har gjort; lad os opgradere den". Men bankernes behov og det offentliges behov var ikke fuldstændig identiske, og der var nogle ting, som sagde, at de funktionelle krav ikke svarede ikke til hinanden. Så det, vi lavede, var én løsning, hvor vi så blev enige om at have en fælles grænseflade. Så de brugere, der ikke interesserer sig for det, vil opleve, at de har én løsning. Når jeg træder på speederen i en hybridbil, er jeg lidt ligeglad med, om det er den ene eller den anden motor. Men det virker som én."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Møberg et al. [21] har i deres årlige undersøgelse af det danske IT-landskab konkluderet, at målsætningerne i globaliseringsstrategien og eDag-konceptet er godt på vej til at blive opfyldt, og kommunerne forventer, at antallet af digitale administrative henvendelser fra borgerne vil stige fra de nuværende 20% til mindst 30% i 2012 (Møberg et al. [21], s102). Andre interessante observationer er, at 43% af befolkningen foretrækker at kommunikere digitalt med det offentlige, og at 62% mener, de har tilstrækkelige IT-kompetencer til at kunne dette. Et tal, som dog falder drastisk, når vi ser på personer over 65 år (Møberg et al. [21], s110f). I det hele taget giver publikationen et interessant overblik over det landskab, som NemID er kommet til verden i.

De forventede fordele ved indførelsen af et så komplekst og dyrt system ligger nogenlunde i forlængelse af den digitale signaturs: større tilgængelighed for brugerne og

effektiviseringsgevinster for offentlige myndigheder og nu også bankerne. NemID ses også som løsningen på den manglende udbredelse af den digitale signatur og efterfølgende noget sløve digitalisering af den offentlige sektor. Mere herom i afsnit 3.4.

3.2 Teknik og virkemåde

Den overvejende del af dette afsnit bygger på interviewet med Peter Lind Damkjær (se Bilag P). I forhold til den oprindelige digitale signatur er der teknisk set sket store ændringer. De grundlæggende principper for kryptering og signering (som gengivet i afsnit 2.2) er ganske vist stadig gældende. NemID er et autentifikations- og signeringssystem baseret på PKI og *to-faktorsikkerhed*. Grundlæggende kan NemID bruges til at logge ind på forskellige online tjenester, blandt andet netbanker og offentlige myndigheder, ligesom det også kan anvendes til kryptering af f.eks. emails. De største forskelle til den digitale signatur er placeringen af brugerens nøgler, samt måden man anvender nøglerne til f.eks. at autentificere sig.

For den digitale signaturs vedkommende var nøglefilen personlig og blev opbevaret af brugeren selv, enten på dennes PC eller på et eksternt medie, f.eks. et smartcard eller en USB-stick. For at øge sikkerheden har man for NemID valgt at lagre samtlige brugeres nøgler centralt på en server administreret af DanID. Man skal med andre ord koble op til denne server for at få adgang til ens nøgler og bruge dem til f.eks. login hos SKAT. Mere herom om et øjeblik. Denne beslutning er taget ud fra den betragtning, at DanID vil være væsentlig bedre i stand til at beskytte borgernes nøgler, end den enkelte borger selv er. Denne fremgangsmåde har vakt stor harme blandt mere IT-kyndige borgere, og vi tager diskussionen op igen i afsnit 3.3.

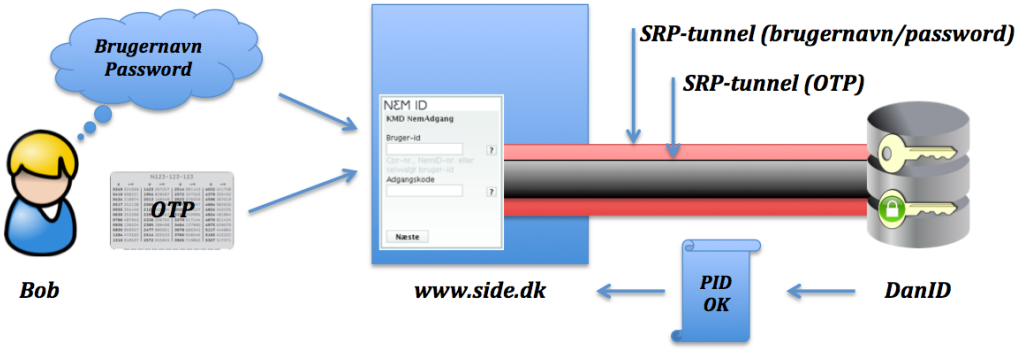
Den anden væsentlige forskel er måden, man anvender NemID på. I det gamle system var den private nøgle beskyttet af et enkelt password. For NemIDs vedkommende er der i stedet tale om at beskytte *kommunikationen* mellem borgeren og dennes nøglepar på DanIDs server. Denne beskyttelse består af tre faktorer: et unikt brugernavn (som kan være selvvalgt, CPR-nummer eller NemID-nummer), et selvvalgt password (som opfylder visse krav) samt en *engangskode* (eller *one-time password*, OTP). Sidstnævnte findes på et foldbart nøglekort i kreditkortstørrelse, som tilsendes den enkelte borger med post. Kortet indeholder 148 kodepar, som hver kan anvendes én gang, og et eksempel kan ses i Figur 4a. Inden koderne på kortet er opbrugt, sendes automatisk et nyt kort til brugerens folkeregisteradresse. Nøglekortet er den ekstra faktor, som gør NemID til ægte to-faktorsikkerhed: kortet er noget, brugeren *har*, mens brugernavn og adgangskode er noget, brugeren *ved*. Nøglekortet er strengt personligt og må ikke kopieres eller digitaliseres på nogen måde. Et fysisk nøglekort sikrer, at f.eks. indbrud på borgerens PC ikke umiddelbart kan resultere i identitetstyveri, da hackeren ikke har adgang til denne faktor.

Processen med autentifikation online foregår gennem en Java-applet (se Figur 4b), som hentes ned fra DanIDs server via en sikker SSL-forbindelse, appletten sørger dermed selv for holde sig opdateret. Processen er skitseret i Figur 5. På tjenesten indtaster Bob sit brugernavn og adgangskode i appletten. Er disse korrekte, etableres der en direkte SRP-

(a) Nøglekort til NemID **(b)** DanIDs Java-applet

Figur 4: Nøglekort og Java-applet hørende til NemID.

tunnel[12] fra appletten til DanID. Gennem denne tunnel sendes nu en OTP-challenge fra serveren, som Bob svarer på ved at indtaste den tilhørende nøgle fra nøglekortet. Er denne korrekt, etableres endnu en SRP-tunnel indeni den eksisterende, og Bob er nu autentificeret overfor DanIDs server. Gennem denne tunnel sendes information om, hvilken tjeneste Bob ønsker at logge sig ind på. Serveren anvender nu Bobs private nøgle til at oprette et signeret certifikat, der indeholder Bobs unikke identitet (i form af et PID-nummer), et tidsstempel og en godkendelse af, at Bob ønsker at logge ind på den pågældende tjeneste på det pågældende tidspunkt. Dette certifikat sendes gennem appletten tilbage til den pågældende side, som Bob nu har autentificeret sig overfor. Offentlige myndigheder samt private virksomheder, der måtte have behov for det, kan veksle det PID-nummer, som ligger i certifikatet, til et CPR-nummer i en database, som IT- og Telestyrelsen driver, og dermed få adgang til de fulde oplysninger om brugeren.



Figur 5: Proceduren for login med NemID på en online tjeneste.

Modsat den digitale signatur er der til NemID en del forskellige indstillinger, som

[12]SRP – *Secure Remote Password* – er en protokol til sikker kommunikation over nettet på baggrund af et password. Selve protokollen regnes for at være særdeles sikker (selvfølgelig afhængig af passwordet) og er modstandsdygtig overfor *dictionary attacks*. Se mere i Wu [40].

kan foretages på en online selvbetjening på www.nemid.nu. Her kan man blandt andet bestille ekstra nøglekort, spærre sin signatur, ændre adgangskode og andre ting. DanID har for nylig også meget anonymt introduceret sikker email, dvs. signering og kryptering af e-post fra brugerens egen computer. Til dette formål skal der downloades et program, der håndterer selve den kryptografiske signering og/eller kryptering af meddelelsen, og processen vil foregå meget lig den for online login – man kan forestille sig programmet som Java-appletten hentet ned på brugerens computer. Nøglerne er stadig fysisk udelukkende placeret på DanIDs server. DanID varetager desuden en certifikatdatabase, som man kan foretage opslag i for at hente offentlige certifikater på personer, man ønsker at sende signeret eller krypteret e-post til. Der vil desuden inden for kort tid komme både elektroniske nøglegeneratorer og smartcards på markedet; førstnævnte som alternativ til nøglekortet, og sidstnævnte for at imødekomme visse brugeres ønske om fuld kontrol over deres private nøgle.

3.3 Sikkerhed og trusler

Ligesom i afsnit 2.3 vil vi her se nærmere på de sikkerhedsmæssige foranstaltninger og trusler, som er knyttet til NemID. Da systemet er så nyt, som det er, er der indtil videre heldigvis ikke registreret deciderede brud på den *tekniske* sikkerhed, men der har været en del kritik af den for brugeren rent praktiske sikkerhed. Sikkerheden generelt i NemID har desuden været særdeles omdiskuteret i medierne, og derfor vil vi her forsøge at klarlægge en del af problematikkerne og imødekomme nogle af mange usaglige påstande, der florerer på nettet. Igen bygger vi mange af vores oplysninger på interviewet med Peter Lind Damkjær samt [15].

Vi starter med at se på de rent tekniske aspekter af systemet. Her er nøglestørrelsen som nævnt i afsnit 2.3 en væsentlig, men ikke afgørende, faktor i sikkerheden hørende til den konkrete signering og kryptering af information. Brugernes nøgler i NemID er dobbelt så store som i den gamle digitale signatur, nemlig 2048 bit. Eftersom certifikaterne desuden højst kan blive 15 år gamle, må denne faktor siges at være særdeles sikker et godt stykke ud i fremtiden.[13]

Ser vi på selve to-faktorsikkerheden består denne som sagt af et brugernavn og password samt en engangskode (OTP). Her er der helt konkret tale om en forbedring i forhold til den digitale signaturs 'halvanden-faktor'. For passwordets vedkommende er der opstillet nogle krav til, hvordan det skal udformes, f.eks. i forhold til længde, specialtegn og lignende. En meget væsentlig detalje er dog, at der ikke skelnes mellem store og små bogstaver – noget som ellers er en almindelig måde at gøre passwords mere komplekse på. For NemIDs vedkommende vil hele det indtastede password blive lowercased, inden det sendes videre i systemet. Selvom dette umiddelbart er en klar svækkelse af sikkerheden af passwordet (og dermed også af den yderste SRP-tunnel i Figur 5), begrunder DanID det med hensynet til brugeren – mange brugere vil konsekvent ikke skelne mellem store og små bogstaver:

"Hvis du taster store eller små bogstaver, så lowercaser vi det hele. Og det

[13] DanID er ved at opbygge et hierarki af CA'er over brugerens certifikater. De udstedende CA'er (som udsteder brugernes certifikater) er ligeledes på 2048 bit, mens de overliggende CA'er, der udsteder disse CA'ers certifikater, er på hele 4096 bit.

er simpelthen fordi, en stor del af befolkningen ikke skelner mellem store og små bogstaver.
ABP: Men det virker på dem, som ved noget om passwords, som en svækkelse af sikkerheden, så jeg undrer mig over, hvorfor I har valgt at skrive specifikt, at der ikke skelnes mellem store og små bogstaver, når I nu alligevel bare skriver 'bogstaver og tal'. Så folk selv kunne vælge, om de ville have lowercase hele vejen.
PLD: Det er netop fordi, vi lowercaser det hele.
ABP: Og hvorfor er det, I gør det?
PLD: Jamen det er fordi, folk de glemmer, at de har lavet et af bogstaverne med stort.
ABP: Så det er af hensyn til brugeren?
PLD: Det er af brugerhensyn, ja. Og så en samlet sikkerhedsvurdering der siger, at vi har alligevel brute-force [-sikring, red.] på, så det er altså væsentlig nemmere at angribe et kreditkort for eksempel, end det ville være at angribe det her system med brute-force. [...] Hvis du ser på det samlet set, så har du en voldsomt meget højere entropi, end du har på mange andre systemer, hvor du for eksempel kræver, at det skal være store og små bogstaver og otte tegn. Her der har vi faktisk 12 tegn, og så ved jeg godt, at de seks af dem det er cifre, men der er faktisk relativt meget entropi på.
ABP: Hvis du blander faktorerne sammen, ja, men hvis du holder faktorerne adskilt, så er man nødt til at kigge isoleret på passwordet.
PLD: Ja, og der vil du kunne opnå samme sikkerhed ved typisk at lægge et eller to tegn ekstra på.
ABP: Men det er så ikke noget, brugerne ved, kan man sige?
PLD: Nej. Men det er jo også derfor vi skriver, at vi ikke skelner, så brugerne de er klar over, at entropien er ikke stærkere, end at vi ikke skelner mellem store og små bogstaver, så nej. Hvis vi ikke skrev det, så kunne nogle blive narret ved at tro, at vi havde en entropi på en vis størrelse.
ABP: Men det er jo så fordi, I har valgt at lowercase passwordet. Det ved brugeren jo heller ikke, at I gør.
PLD: Det er af brugerhensyn, da der en stor del, der ikke skelner, og ikke kan huske, at de har valgt nogle af bogstaverne med stort."

Peter Lind Damkjær, DanID (Bilag P)

Det giver altså umiddelbart god mening at ville hjælpe brugerne ved ikke at *kræve* forskel på store og små bogstaver, men for mere kyndige brugere giver det i vores øjne ikke mening at fjerne *muligheden*, når systemet teknisk set understøtter brugen af store og små bogstaver. DanID gør sig her måske lidt for store antagelser om den generelle brugers viden om entropi og styrker af passwords.

Nøglekortet udgør udover brugernavn og adgangskode den anden væsentlige faktor i to-faktorsikkerheden, men dermed stilles der ligesom for nøglefilen i den digitale signatur også visse krav til opbevaringen af kortet. DanID fraråder enhver form for kopiering eller digitalisering af nøglekortene, da de i så fald vil kunne komme til at ligge på f.eks. brugerens PC, og dermed skal der ikke mere end en simpel keylogger til, før en hacker vil kunne misbruge brugerens identitet. Den afledte fordel ved to-faktorsikkerheden ligger så i, at hvis nøglekortet skulle blive stjålet, har tyven stadig brug for den anden

faktor (brugernavn og password) for at kunne udnytte systemet og omvendt. En interessant pointe er dog, at en bruger ikke nødvendigvis har mulighed for at opdage det, hvis vedkommendes nøglekort er blevet kopieret – noget som en elektronisk nøglegenerator derimod ville være modstandsdygtig overfor.

Med hensyn til offline- og online-angreb er der også sket forbedringer i NemID. Offline-angreb er principielt ikke længere mulige, da der er en fysisk sikkerhedsfaktor i form af nøglekortet, som ikke kan tilgås elektronisk, med mindre selvfølgelig nøglekortet er blevet kompromitteret. Med andre ord er det nærmest umuligt at foretage brute-force-angreb på systemet, selvom brugerens PC skulle være inficeret. Desuden spærres brugerens NemID i 8 timer efter fem fejlslagne forsøg, og efter yderligere fem forsøg kan der kun åbnes op igen med en ny midlertidig adgangskode.

Valget af Java som redskab til selve autentifikationen har fået stor kritik af blandt andet eksperter på området og på IT-relaterede fora rundt omkring på nettet. Vi er ikke i stand til konkret at vurdere sikkerheden i den valgte løsning, men i en artikel i PROSA-bladet (december 2010) udtaler Peter Lind Damkjær, at en løsning baseret udelukkende på HTTPS og JavaScript ikke ville yde tilstrækkelig beskyttelse mod kendte MITM- og MOTW-angreb, ligesom de underliggende kryptografiske beregninger ville være for tunge at afvikle i JavaScript. Man kan forholdsvis nemt se, at MOTW-angreb er så godt som umulige, men et simpelt MITM-angreb, hvor en mellemmand via en vellignende men falsk kopi af en kendt tjenestes hjemmeside, som han har lokket brugeren ind på, formidler kommunikationen mellem bruger og server, kan principielt godt lade sig gøre:

> *"Ja, det er muligt at lave sådan et scenarie. Vi har lavet nogle ting i appletten, der gør det besværligt at lave det angreb. Og det er vigtigt også at vide, at i det her NemID...det vi har startet, det er en proces. Vi har ikke lavet en endelig løsning, der er færdig, og så er den mejslet ind i sten - så vi kender godt de angreb, man kunne forestille sig, og vi overvåger også for, hvordan verden ser ud, og hvordan det udvikler sig, og vi har nogle skruer, vi kan skrue på, så vi kan fange de her ting, efterhånden som vi ser, at det bliver et scenarie, som vil ske. Men vi har også skrevet på vores FAQ, at man godt kan forestille sig man-in-the-browser og man-in-the-middle angreb, altså online phishing-angreb. [...] Så vi overvåger, og vi har nogle mekanismer, vi kan skrue på, men det vil gøre, at der kommer en brugerinvolvering, og det er altid et spørgsmål om at finde niveauet. Der er ikke noget, der hedder 100% sikkerhed. Vi kan aldrig gardere os 100% - om ikke andet så er det, brugeren ser på, en glasplade, og han har et tastatur og et interface. Det kan man altid mingelere med, så brugeren tror, der sker noget andet. Man kan jo ikke sidde og kigge på de strømme, der løber nede i computeren, så det er et spørgsmål om at sige, hvor skal man sætte barrieren. Og den barriere vil flytte sig over tid."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Eksemplet i medierne med www.nemid.dk viste desuden, hvor forholdsvis let det kan være at lave en hjemmeside, som folk blindt stoler på og få dem til at forsøge at logge ind med deres NemID. Til NemIDs forsvar skal dog nævnes, at det med engangskoderne er blevet sværere for en hacker at anvende de opsnappede informationer på vedvarende

basis.

En mere praktisk sikkerhedsfaktor er udsendelsen af selve nøglekortene og primært samkøringen af denne med udsendelsen af den midlertidige adgangskode. Der har af flere omgange været episoder, hvor mange tusinde nye brugere har fået spærret deres NemID lige efter oprettelsen, fordi de havde modtaget velkomstbrevet med nøglekortet og den midlertidige adgangskode på samme dag, hvilket ifølge Datatilsynet er en alvorlig sikkerhedsbrist, da de to breve tilsammen kan være nok til at misbruge brugerens identitet. Der er her tale om praktiske fejl på grund af overbelastning og menneskelige fejl, men der er stadig tale om en væsentlig sikkerhedsbrist, som DanID har reageret forholdsvis korrekt på hver gang ved at spærre og genudstede de berørte NemID – dog med enkelte forsinkelser.

Den sidste, men måske mest omdiskuterede, faktor i sikkerheden af NemID er placeringen af brugerens private nøgler på DanIDs servere. Her har vi fat i den altoverskyggende forskel mellem den digitale signatur og NemID: hvilken betydning har det for brugeren, at vedkommende ikke har direkte fysisk adgang til sin personlige nøgle? Praktisk: ingen. Sikkerhedsmæssigt: nogen. Personligt: for nogle bestemte brugere en hel del. For den praktiske dels vedkommende er det umiddelbart kun blevet nemmere for brugeren, hvis man ser på den rene nøglehåndtering (introduktionen af nøglekortet er en helt anden sag). Der er ikke de samme problemer med installation, generering af nøgler og lignende som med den gamle signatur. DanID har ligeledes fået foretaget en juridisk vurdering af, om det overhovedet er tilladt at opbevare borgeres private nøgler centralt (se Udsen [38]). Notatet konkluderer, at der ikke umiddelbart er problemer forbundet med denne opbevaringsform. Desuden behøver man ikke længere være så nervøs for, om man har beskyttet sin personlige nøgle godt nok. Og dette leder os hen til den sikkerhedsmæssige betydning: ligger nøglerne så sikkert nok hos DanID? Ifølge DanID selv ligger de bedre end hos den enkelte bruger:

> *"Jeg har hørt nogle kritikere sige til mig, at de kan beskytte nøglen bedre end man kan på en HSM [hardware security module, red.] hos DanID. Jeg har været i sikkerhedsbranchen i nogle år, så jeg er ret sikker på, at jeg selv ikke ville være i stand til at beskytte nøglen så godt, som man kan. [...] Så jeg er glad for, at jeg ikke har mine ting på min PC mere, for jeg er ikke sikker på, det er et godt sted at have dem."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Der bliver ligeledes gået meget op i den konkrete beskyttelse af nøglerne, eller med andre ord 'muren omkring dem':

> *"Hele vores infrastruktur er baseret på, at den mur skal være sikker. Så vi har gjort os meget meget umage. Vi har haft alle de kritiske faktorer, vi har lavet, til review hos tredjeparter og skal have en rating på alt det, vi laver. Man rater efter, hvor godt man synes, det er. De protokoller, vi laver, har vi haft både indenlandske og udenlandske eksperter til at kigge på og sige, hvor er det, at sårbarhederne er i det her, og hvor realistisk er det, at de vil blive brugt. Hele vores driftsapparat er bygget op om, at alle de steder, hvor der er nogle kritiske ting, f.eks. vores CA-nøgler eller de steder, hvor vi har brugerens private nøgler liggende, der har vi en udpræget brug af dualitet,*

> *dvs. at der skal mere end én person til at gøre nogle bestemte ting. Vi har zoneopdelt det, og i det inderste i vores CA, der kan der ikke gå én mand ind alene - der skal der to mand til. Og det skal ikke være to driftsfolk, det skal være en driftsperson og en anden, typisk en security officer eller en, der har en applikationsvej [uklar optagelse, red.]. Der skal være to til stede, og de skal have forskellig funktion.*
> *ABP: Så du vil mene, at muren er særdeles holdbar?*
> *PLD: Der er nogle folk, der synes, vi er hysteriske, men det er vores forbandede pligt at være hysteriske.*
> *ABP: Er muren ubrydelig?*
> *PLD: Der er ikke noget i verden, der er 100% sikkert, men vi har gjort os meget store anstrengelser i at gøre det usædvanlig svært at angribe det her. Vandet vil søge hen de steder, hvor det er nemt at løbe igennem, og det er* ikke *server-side."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Så selve sikkerheden omkring brugernes og CA'ens nøgler bliver der åbenbart gjort meget for at sikre, men der er selvfølgelig stadig en potentiel mulighed for, at muren gennembrydes, hvilket i så fald kunne medføre en spærring og efterfølgende genudstedelse af samtlige digitale signaturer i Danmark.

Noget helt andet er de forholdsvist mange IT-kyndiges kritik af, at de ikke fysisk har kontrol over deres private nøgle. Med indførelsen af smartcards senere på året tager DanID denne kritik til sig og muliggør dette, men indtil da ligger den private nøgle udelukkende på DanIDs server. DanID foretager dog en skelnen mellem at have *adgang* til og at have *kontrol* over ens personlige nøgle:

> *"Det, der er meget væsentligt, det er, at det er brugeren, der har kontrol over den private nøgle. Det er et af de kritikpunkter, der har været, og derfor er der nogen, der ikke mener, at det er PKI det her, fordi den private nøgle ikke ligger hos brugeren. Men det, mener jeg måske nok, er en meget firkantet måde at se det på, for det, der er væsentligt, er, at det kun er brugeren selv, der kan bruge den her private nøgle – at systemet er sat op på den måde."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Distinktionen mellem adgang og kontrol er selvfølgelig en mere akademisk en af slagsen, og vi vil ikke her vurdere, hvad der er klogest eller sikrest. Vi er dog enige med DanID i, at nøglerne *generelt* sandsynligvis vil være bedst beskyttet på en central og stærkt sikret server – selvfølgelig med forbehold for, at en del IT-kyndige mennesker er udmærket i stand til at beskytte den selv.

Den sidste knast i mediernes sikkerhedsdækning af NemID er af mere konspiratorisk art: tør man stole på, at DanID ikke misbruger sin magt til at stjæle brugeres identiteter, i al hemmelighed logge brugerens transkationer eller usynligt inficere brugerens PC med keyloggere eller spyware? Det kunne de sagtens, men gør de det så? Med enormt overvejende sandsynlighed nej. Den blandt mere IT-kyndige brugere nærmest konspiratoriske mangel på tillid til DanID er i vores øjne stærkt overdrevet, og vi er ikke et sekund i tvivl om, at DanID som virksomhed aldrig bevidst ville bryde deres forpligtelser som certificeret CA. At enkelte medarbejdere måske vil kunne gøre dette er næsten umuligt

at forhindre, men det er der heller ikke noget nyt i. De fleste ansatte i offentlige myndigheder har opslagsrettigheder til personfølsomme oplysninger, men igen mangler den fysiske faktor i form af nøglekortet stadig for at kunne fuldbyrde forbrydelsen. I forhold til den centrale lagring af de personlige nøgler ville DanID nemt kunne udstede et nyt sæt nøgler for en hvilken som helst person i kongeriget til eget brug, men i forhold til den gamle digitale signatur er der her igen intet nyt under solen, og en stor del af sikkerheden i et PKI bygger derfor netop på borgernes tiltro til certificeringscentrene.

3.4 Distribution og udbredelse

Distribution og udbredelse af NemID må siges at være radikalt anderledes end for den digitale signaturs vedkommende. Fra den 1. juli 2010 og året ud blev der hver måned udsendt ca. 600.000 velkomstbreve med nøglekort og et seperat brev med en midlertidig adgangskode. Brevene blev automatisk sendt til samtlige netbankbrugere i Danmark[14] i tilfældig rækkefølge. Den faktiske ibrugtagning var dog i høj grad afhængig af, hvor hurtigt de enkelte banker kunne få integreret NemID i deres systemer. På denne måde er det lykkedes NemID at komme ud til ikke mindre end ca. 3 millioner danskere inden for et halvt år. Det er dog langt fra alle, som har modtaget deres NemID, som også har taget det i brug. Der er stadig i februar mellem en halv og en hel million ikke-aktiverede NemID blandt befolkningen, men systemet har effektivt opnået større udbredelse og anvendelse på et halvt år, end den digitale signatur kunne nå i løbet af 7 år.

Udover den automatiske distribution til netbankkunderne har alle borgere desuden kunnet bestille et NemID online fra og med lanceringsdatoen 1. juli 2010. Noget som ganske mange har benyttet sig af, enten fordi de ikke var netbankkunder, eller fordi de ikke gad vente på, at deres NemID kom med posten automatisk. Desuden er det muligt at få udleveret sit NemID hos Borgerservice- og skattecentre rundt omkring i landet, ligesom man som netbankkunde kan få det ved personligt fremmøde i sin bank.

Landskabet er dog også et væsentlig andet, særligt i forhold til alternativerne til NemID. For netbankkunder er der ingen! NemID vil afløse samtlige andre loginmuligheder på samtlige netbanker i Danmark. Herudover skifter samtlige offentlige myndigheder og tjenester til NemID i løbet af ganske kort tid, dog vil den gamle digitale signatur være understøttet, så længe der stadig er aktive certifikater. Der er altså for NemIDs vedkommende tale om en langt mere agressiv distribution og understøttelse i forhold til den mere passive form anvendt under udbredelsen af den digitale signatur.

3.5 Brugervenlighed

I beskrivelsen af brugervenligheden af NemID vil vi holde os til de funktioner, som er direkte rettet mod brugervenlighed samt de testrapporter, som DanID har fået udarbejdet inden systemets lancering. Afsnittet vil være forholdsvis kort, da resten af specialet særskilt handler om netop dette emne.

[14] Ifølge Møberg et al. [21] har ca. 80% af danskerne mindst én netbankløsning.

3.5.1 Features direkte rettet mod brugervenlighed

I det aktstykke fra Ministeriet for Videnskab, Teknologi og Udvikling, som er fremlagt for Finansudvalget omhandlende indgåelse af kontrakten med DanID om udvikling og drift af den kommende generation af digital signatur, står følgende beskrevet:

> *"Etableringen af en fælles digital signatur med finanssektoren har en række fordele for borgerne, herunder øget brugervenlighed, idet signaturen anvendes hos både banker og det offentlige. Den tilbudte løsning medfører et sikkerhedsløft af den digitale signatur, da denne fremover opbevares centralt hos leverandøren frem for på borgerens egen computer. Sikkerhedsløftet er endvidere baseret på, at borgeren skal indtaste to koder ved anvendelse af signaturen frem for en kode som i dag. Dette kan virke mere besværligt for brugeren, men det vurderes, at borgeren hurtigt vil vænne sig til den nye løsning, da signaturen anvendes i både banker og det offentlige."*
>
> Ministeriet for Videnskab, Teknologi og Udvikling [20], s14m

Den øgede brugervenlighed er dermed opsat som et af de direkte mål med det nye system og vil delvist blive udmøntet gennem en harmonisering af login-mulighederne mellem det offentlige og netbankerne. Sloganet for NemID er da også 'Et log-in til det hele', hvilket signalerer, at brugeren nu kan nøjes med at huske ét brugernavn og password til både sin netbank, offentlige myndigheder og selvbetjening samt eventuelle tilknyttede private virksomheder. Kombinerer man dette med det offentlige single-signon-system (NemLog-in), er der i sandhed tale om bred og lettilgængelig autentifikation. Der bliver ligeledes slået på genkendeligheden af Java-appletten og de omkringliggende elementer på de til NemID tilknyttede tjenester. Endvidere er der i kravspecifikationen for NemID i afsnit 3.1.4 om brugervenlighed udpeget følgende:

> *"IT- og Telestyrelsen ønsker at styrke brugervenligheden i den nye PKI-infrastruktur (specifikt i forhold til de administrative funktioner som bestilling, installation, flytning af certifikater m.m.), og ser denne styrkelse som en vigtig forudsætning for at indfri målsætninger for digital signatur. Brugervenligheden kan bl.a. øges gennem en bedre anvendelse af teknologier og mekanismer, hvor der f.eks. ikke kræves en endelig installation af certifikater."*
>
> NemID kravspecifikation, s12

Her er det altså nærmere en fjernelse af nogle af de knap så brugervenlige faktorer fra den digitale signatur, der er tale om. På https://www.nemid.nu/privat/fordele_med_nemid/ listes endvidere fire konkrete fordele ved NemID:

- **Brugervenlighed:** Argumentet herfor er, at NemID "kan du bruge på de fleste computere med internetadgang".
- **Fælles login:** Er som ovenfor nævnt, at NemID sammen med NemLog-in kan bruges på samtlige offentlige sider og netbanker. Genkendelighed og hurtig tilpasningsevne, eller *learnability*, er altså nøgleordene her.
- **Portabilitet:** DanID slår her på, at NemID er uafhængig af installation på en bestemt PC. Det kan med andre ord bruges fra næsten hvilken som helst computer, blot den har internetadgang. Både i Danmark og i udlandet.

- **Øget sikkerhed:** Her skitseres ganske kort, at to-faktorsikkerheden i systemet er med til at højne sikkerheden og beskytte mod "indbrudstyve og hackere".

Selvom brugervenligheden står øverst på NemIDs officielle liste over fordele, er det altså slående lidt, der konkret står beskrevet om emnet. Navnet i sig selv hentyder også til brugervenlighed, men igen er det svært at læse sig til, *hvorfor* NemID er 'nemt'.

En anden ting, som DanID reklamerer med, er den nemme og hurtige aktiveringsproces, samt at processen med nøglekort, der løber tør, forløber transparent og smertefrit: brugeren behøver ikke tænke på, om vedkommende løber tør for koder – et nyt kort tilsendes automatisk, inden dette sker. Kortet i sig selv er lille og foldbart og kan dermed ligge i brugerens pung sammen med kreditkort og lignende.

3.5.2 Brugertests udført af DanID inden lancering

DanID har inden lanceringen af NemID testet produktet grundigt, og vi vil i dette afsnit nævne nogle af konklusionerne fra den testrapport, som firmaet Userminds har udarbejdet i juni 2009 på baggrund af 22 udførte tænke-højt tests af DanIDs logon-applet. Der har for denne rapports vedkommende været tale om en prototype af appletten, som derfor ikke bør sammenlignes med det lancerede produkt. Testrapporten er sammen med et antal tilhørende dokumenter blevet udleveret ved aktindsigt fra IT- og Telestyrelsen og kan ses i Bilag Q.[15]

Testen blev udført på et bredt udsnit af befolkningen i forhold til køn, alder, IT-kompetencer og erhvervsbeskæftigelse og mundede ud i identifikation af 56 usabilityproblemer samt anbefalinger til udbedringer, som allesammen efterfølgende er blevet gennemgået af DanID, IT- og Telestyrelsen samt repræsentanter fra et udvalg af banker. Vurderingen og rettelsen af de identificerede problemer har ikke efterfølgende givet anledning til yderligere tests af systemet (så vidt vides ud fra det udleverede materiale). De primære identificerede udfordringer i rapporten er forståelsen af NemID-konceptet generelt, ligesom en del deltagere er utilfredse eller usikre på konceptet med at anvende et nøglekort – begge dele ting, som ikke bare kan fikses med en softwareopdatering. Man må på denne baggrund kunne antage, at DanID på det rent praktiske område har gjort, hvad de syntes var nødvendigt for at teste for og imødekomme eventuelle usabilityproblemer, der måtte være i det lancerede produkt.

IT- og Telestyrelsen og DanID har ligeledes inddraget relevante parter i høringsprocesser og lignende, blandt andet ÆldreSagen og Dansk Blindesamfund, for at højne brugervenligheden for ældre og svagtseende/blinde. Der er til disse personer udviklet et særligt system, hvor man enten kan få et nøglekort med større tal eller en speciel løsning til blinde, hvor man kan få sin OTP-kode ved at ringe til en telefonservice. Disse løsninger har DanID også bedt Userminds om at teste, og vi har fået udleveret testrapporterne, men vi vil i dette projekt af tidsmæssige årsager ikke komme nærmere ind på dem.

[15] Der blev i forbindelse med aktindsigten ligeledes udleveret testrapporter for den særlige løsning til blinde og svagtseende, test af nemid.nu og dennes funktioner, præsentationer og opsamlinger heraf samt diverse interessenters (primært de største danske bankers) kommentarer til disse. Vi vil dog lægge vores fokus på testrapporten om appletten, da det er her, vores egne undersøgelser primært kommer til at befinde sig.

Opsamling Vi har i dette afsnit set nærmere på NemID og særligt lagt vægt på brugervenlighed og teknisk virkemåde. For førstnævntes vedkommende er de politiske forventninger store, men det har været svært at se, hvordan disse forventninger konkret er blevet implementeret. De udførte tests inden produktets lancering indikerer primært forståelsesmæssige problemer i forhold til koncepterne samt utilfredshed med nøglekortet. Dette kort er en væsentlig faktor i den øgede sikkerhed i systemet, ligesom placeringen og opbevaringen af brugerens nøglepar er markant anderledes end for den digitale signatur. Den opsamlede viden skal vi i de efterfølgende afsnit bruge til at designe meningsfulde og realistiske testopgaver, ligesom vi skal indsnævre de dele af systemet, som vi ønsker at teste.

4 Brugerevaluering af NemID

I det følgende afsnit udvælger og beskriver vi de forskellige metoder, vi har valgt at brugerteste NemID med. Grundet projektets primære fokus på *brugerens* oplevelse af produktet, har vi udvalgt tre forskellige testmetoder, som beskrives nærmere nedenfor. Vi vil også komme ind på, hvorfor disse metoder er anvendelige til formålene, ligesom vi grundigt beskriver, hvorfor og hvordan de forskellige metoder er anvendt og tilpasset i dette konkrete tilfælde. Afsnit 4.1 er en generel præsentation af metoderne, hvorimod det efterfølgende afsnit ser på den konkrete implementering af disse i dette projekts undersøgelser.

For at få en fælles referenceramme at tale ud fra, vil vi kort præsentere den definition af usability, som vi vil arbejde efter gennem resten af afhandlingen. Den bredest anvendte definition af usability er ISO 9241-11-standarden, der opdeler usability i *effectiveness*, *efficiency* og *satisfaction*. Vi vil primært arbejde ud fra denne definition, dog vil vi også lejlighedsvist anvende Schneidermans og Nielsens udvidelser/underopdelinger (se f.eks. Schneiderman og Plaisant [32]) til at præcisere, hvad vi ønsker at teste, ligesom de på nogle punkter er mere målbare. Disse forskellige definitioner kan ses i Tabel 1.

ISO 9241-11	**Schneiderman**	**Nielsen**
Efficiency	Speed of performance	Efficiency
	Time to learn	Learnability
Effectiveness	Retention over time	Memorability
	Rate of errors by users	Errors/Safety
Satisfaction	Subjective satisfaction	Satisfaction

Tabel 1: Oversigt over tre forskellige definitioner af usability, gengivet fra Tabel 1 i van Welie et al. [39].

Efficiency drejer sig om, hvor hurtigt brugeren kan udføre de stillede opgaver, både i forhold til hvor lang tid opgaven konkret tager at løse, samt hvor længe brugeren er om at finde ud af hvordan. *Effectiveness* handler primært om, hvor godt de indlærte opgaveløsninger sidder fast hos brugeren og kan huskes over længere tid, samt i hvor stort omfang brugeren genererer fejl undervejs. *Satisfaction* drejer sig om brugerens umiddelbare eller mere langsigtede tilfredshed med systemet, dets virkemåde, udseende eller opførsel. Modsat de to foregående kategorier måles denne kategori næsten udelukkende subjektivt.

Vi vil indledningsvist også knytte et par kommentarer til, hvorfor vi overhovedet bruger så meget energi på at teste NemID – DanID har jo udført forholdsvis omfattende tests af produktet inden lanceringen (se afsnit 3.5.2). Et af motivationspunkterne for denne afhandling er dog, at modtagelsen af NemID i befolkningen mildest talt har været lunken og mestendels kritisk; i hvert fald hvis man kigger på mediernes dækning af lanceringen og den efterfølgende opstartsfase. Ligeledes har mange elektroniske medier og fora rundt om på nettet flydt over med kritik af både teknikken, sikkerheden og brugervenligheden af systemet, endda før det overhovedet blev lanceret. Den megen mediedækning har skabt politisk bevågenhed fra den siddende opposition, som nu er begyndt at stille kritiske §20-spørgsmål om NemID til den ansvarlige minister. Noget kunne altså tyde på, at NemID har fået en særdeles vanskelig start, og alene dette berettiger et ekstra eftersyn

af produktet. Vi forholder os kritisk til DanIDs tests og foretager egne undersøgelser, der forhåbentlig kan være med til at belyse, om nogle ting er røget for hurtigt igennem udviklings- og testfasen.

Greenberg og Buxton [6] argumenterer i tråd med ovenstående for, at der er en udbredt mangel på replikering af mange af de resultater, som fremkommer i HCI-relaterede artikler baseret på f.eks. tænke-højt tests. Der er ifølge forfatterne ingen status eller umiddelbar publiceringsværdi i at forsøge at genskabe resultater, med mindre de oprindelige undersøgelser har vist opsigtsvækkende resultater. I NemIDs tilfælde er replikering derfor i vores øjne en god ide, da de positive resultater af DanIDs egne tests er opsigtsvækkende på den måde, at de tilsyneladende harmonerer dårligt med befolkningens opfattelse og brug af systemet. Noget som usabilitytests ellers skulle være med til at forhindre.

Vi vælger dog ikke 'bare' at gentage DanIDs undersøgelser. Dette er i vores tilfælde praktisk og tidsmæssigt umuligt. I stedet vælger vi at konstruere vores egne tests, baseret bl.a. på studier af den ønskede funktionalitet af systemet og de af medierne udpegede kritikpunkter. Vi holder os heller ikke kun til tænke-højt tests, men ønsker at udbrede undersøgelsen til en større skala med mange brugere – noget som DanID ikke havde samme muligheder for under udviklingsfasen inden lanceringen. Tullis et al. [36] påpeger i tråd med dette, at flere typer af tests vil give anledning til større og mere varieret udpegning af usabilityproblemer, ligesom der argumenteres for, at et lille antal testpersoner i f.eks. en tænke-højt test ikke nødvendigvis giver specielt pålidelige konklusioner omkring *subjective satisfaction*. Her er der behov for at supplere med en større respondentgruppe.

Under en normal testforberedelse er det meget vigtigt at holde sig nøje for øje, hvad resultaterne af de planlagte tests skal bruges til. I de fleste tilfælde skal resultaterne anvendes til videre arbejde med fejludbedring eller forbedringer af det testede produkt før en eventuel lancering, og det vil derfor være af stor betydning, at udviklere, usabilityeksperter og lignende inddrages i forberedelsen af testen. Omvendt *evaluerer* vi i dette projekt et færdigt og lanceret produkt med cirka et halvt år på bagen, og vi er ikke nødvendigvis interesseret i at ende med en udpegning af konkrete bugs eller eventuelle anbefalinger til ændringer i produktet. I og med at NemIDs grundydelser allerede er færdigudviklet, testet og lanceret i stor skala, kan vi i disse tests fokusere på retrospektivt at forsøge at bedømme det færdige produkts kvalitet ud fra brugernes synspunkt, samt udpege eventuelle svagheder og styrker i håb om at kunne vurdere kvaliteten af produktet på udvalgte fokusområder.

4.1 Testmetoder til slutbrugerevaluering

For at give det bedst mulige billede af brugerens anvendelse af NemID både i introduktionsfasen (eller aktiveringen) og i den efterfølgende daglige brug har vi valgt at anvende forskellige typer undersøgelser, som beskrives nærmere nedenfor. Til den *kvalitative* del, er den primære metode *tænke-højt tests*, suppleret med *cooperative usability testing*. Til den mere *kvantitative* del, som berører både aktiveringen og den efterfølgende brug, har vi valgt at anvende to tidsmæssigt adskilte *spørgeskemaer* på samme målgruppe. Distinktionen mellem kvalitative og kvantitative data er i dette tilfælde dog ikke helt sort/hvid, da tænke-højt principielt kan levere kvantitative data for et lille antal personer, ligesom spørgeskemaernes fritekstfelter kan levere mere kvalitative input til undersøgelsen. Ved

at foretage denne adskillelse sikres det dog, at der anvendes de metoder, der giver størst udbytte i forhold til de resultater vi ønsker at analysere og studere. Desuden sikres både bredde og dybde i undersøgelserne, og metodevalget gør, at vi kommer rundt i alle de tre usability-kategorier i Tabel 1.

4.1.1 Tænke-højt test

Tænke-højt test (THT) bliver af mange i HCI-miljøet anset som værende standarden inden for test af en lang og bred vifte af forskellige produkter inden for teknologi, IT, husholdning og mange andre produktkategorier. Metoden kan anvendes i mange forskellige faser af en udviklingsproces, lige fra tegnede prototyper til release candidates og færdige produkter. Teorierne bag THT bygger på koncepterne i *Verbal Protocol Analysis* (Ericsson og Simon [4]), men anvendes dog i sin nuværende form i en noget mindre teoretisk udgave. Der findes mange forskellige udgaver af THT, men den i dette projekt anvendte tager udgangspunkt i den mere klassiske metode, som er gengivet i Rubin og Chisnell [31] og Molich [22].

Overordnet går THT ud på, at et antal nøje udvalgte testpersoner stilles en række realistiske og meningsfulde opgaver, som skal løses, mens en testleder og eventuelle observatører overvåger processen og kan yde assistance undervejs. I den klassiske udgave må testlederen kun komme med opfordringer til at tænke højt og ikke yde nogen anden form for assistance. Det er dog de færreste testledere, der følger denne metode stringent. Under opgaveløsningen bliver testpersonen bedt om at *tænke højt*, altså i videst muligt omfang hele tiden fortælle, hvad der foregår, hvad personen foretager sig, og hvad personen har af overvejelser og tvivl. Efter løsningen af opgaverne afsluttes forløbet normalt med en form for debriefing, hvor testlederen stiller spørgsmål til forløbet, får opklaret eventuelle uklarheder undervejs, og hvor testpersonen har mulighed for at kommentere og uddybe forløbet. Dette forløber oftest som en variant af et *semi-struktureret interview*.

Det overordnede forløb for en standard THT kan skitseres som følger:

1. Modtagelse og velkomst af testperson
2. Introduktion til forløbet og afklaring af eventuelle spørgsmål
3. Eventuel screening af testperson og underskrift af samtykkeerklæring
4. Løsning af foruddefinerede opgaver
5. Debriefing og eventuelle spørgsmål

Vi kommer nærmere ind på, hvordan de enkelte punkter er implementeret i denne undersøgelse i afsnit 4.2. Forløbet tager normalt mellem 60 og 90 minutter, hvor hovedparten af tiden bruges på punkt 4 ovenfor, afhængigt af det testede produkt samt opgavernes sværhedsgrad og omfang. Vi vil nu se på nogle af de vigtigste af de mange tommelfingerregler, der findes for gennemførsel af en THT. Disse (og flere) vil blive taget op igen, når vi beskriver opbygningen af vores undersøgelser af NemID i afsnit 4.2.

Testpersoner En væsentlig faktor i en succesfuld THT er forberedelsen. Det er særdeles vigtigt at udvælge sin målgruppe korrekt, så man ikke sidder med testpersoner, som f.eks. aldrig ville kunne komme til at skulle bruge det testede system eller testpersoner, som alle sammen er overkvalificerede til at bruge systemet, så der derved ikke konstateres nogen som helst problemer undervejs. Det er altså vigtigt at holde sig for øje, hvilken målgruppe der forventes at komme til at anvende det testede system – at teste på andre end den giver ingen umiddelbar mening eller brugbare resultater. Udover disse krav kan testpersoner dog være alt fra ansatte i den virksomhed, som udvikler produktet, til ens egne kolleger, familie eller venner – så længe de passer ind i målgruppen for det testede produkt.

Det nødvendige antal testpersoner har i mange år skabt stor diskussion blandt testledere og forskere inden for HCI-miljøet. Flere (bl.a. Nielsen [25]) har plæderet for, at fem testpersoner vil være nok til at finde langt de fleste (80-85%) af usabilityproblemerne i f.eks. software eller hjemmesider. Andre er mere ambitiøse og mener, at mindst 8-10 testpersoner er nødvendige. Lindgaard og Chattratichart [18] har testet, hvorvidt der er en sammenhæng mellem antallet af testpersoner og antallet af fundne problemer, og de finder, at der ingen umiddelbar statistisk sammenhæng er. Deres undersøgelse er baseret på CUE-4 undersøgelserne (se Molich og Dumas [24]), hvor 17 uafhængige eksperthold blev bedt om at teste de samme to hjemmesider på forskellige måder. Lindgaard og Chattratichart [18] baserer deres resultater på de 9 hold, der valgte at lave usabilitytests med testpersoner, i modsætning til f.eks. heuristisk inspektion eller expert reviews. De fandt dog i tråd med ovenstående, at antallet og udformningen eller *dækningen* af opgaverne var en væsentlig faktor for antallet af identificerede usabilityproblemer. Så hvor mange testpersoner, der er nødvendige for at opdage så og så mange usabilityproblemer, må bero på en vurdering i hvert konkret tilfælde og vil ofte afhænge meget af målgruppen for og sværhedsgraden af det testede produkt. Uanset hvad er forskere dog enige om, at man med stor sandsynlighed vil skulle bruge mindst fem testpersoner, og at det er usandsynligt, at man ved at anvende mange testpersoner vil kunne identificere *alle* usabilityproblemer i et produkt. En pragmatisk tilgang, hvor man tilføjer testpersoner, indtil ingen nye usabilityproblemer dukker op, ses også som en løsning.

Opgaver Med hensyn til selve opgaverne er der også en del forholdsregler. Opgaverne skal være præcist definerede, de skal kunne relateres til produktet og brugerens hverdag af testpersonen selv, ligesom unødvendig humoristisk tilgang bør undgås; de skal med andre ord føles realistiske for testpersonen. At give en økonomiansvarligs daglige arbejdsopgaver til en receptionist giver i de fleste tilfælde ingen mening og vil føre til misvisende resultater af THT'en. En anden tommelfingerregel er, at den eller de første opgaver i rækken skal være lette at løse for alle testpersoner. Dette får testpersonen til at føle sig tryg ved forløbet og kompetent til at fortsætte den videre opgaveløsning. Denne første opgave kan dog tit være ganske vanskelig at konstruere (Molich [22], s103ff). Opgaverne kan med fordel uddeles på papir, da dette modsat en oplæsning fra testlederens side mindsker dennes indflydelse på en eventuel forståelse eller opfattelse af opgaven. En anden fordel ved at udlevere opgaverne på papir er, at testpersonen kan blive instrueret om at give opgaven tilbage til testlederen, så snart den efter testpersonens egen mening er løst. Dette hjælper med at måle faktorer som task completion time og tydeliggør, om der er tale om task failures.

Under udformningen af opgaverne skal der lægges stor vægt på ikke at hjælpe unødigt meget med løsningen af opgaven, da dette vil forvanske billedet af, hvorvidt testpersonen selv er i stand til at udføre den opgave, der egentlig ønskes testet. Det kan dog i mange tilfælde være nødvendigt at yde tilstrækkelig hjælp i opgaven, så testpersonen ikke tvinges ud i nye eller ukendte tiltag, der ikke har relevans for indholdet af opgaven eller benytter sig af features, der ikke vil blive brugt af testpersonen.[16]

Testlederens rolle En meget væsentlig faktor i af få en THT til at lykkes er testlederens rolle og opførsel under forløbet. Som nævnt ovenfor er han i den klassisk teoretiske model ude af stand til at give andet end opfordringer til testpersonen om at tænke højt. I de fleste tilfælde er der dog en form for kommunikation mellem de to, uden at testlederen dog kommer med konkret hjælp til opgaverne, før testpersonen specifikt har opgivet (og dermed har givet anledning til en task failure). Testpersonen har normalt lov til at stille spørgsmål til testlederen under forløbet, som testlederen skal svare neutralt og opfordrende på og f.eks. bede testpersonen prøve sig frem, hvis der er tvivl, eller spørge ind til hvad testpersonens næste mulige trin kunne være. Testlederens primære rolle er dog at få testpersonen til at føle processen som værende naturlig og i videst muligt omfang foregå som om, situationen ikke var opstillet. Sekundært kan testlederen intervenere og eventuelt yde assistance til testpersonen, når det er tydeligt, at et problem ikke kan overkommes uden hjælp. En intervention bør dog altid være sidste udvej, idet det bryder det naturlige forløb i processen. Der findes enddog specifikke lister med typer af interventioner, som en testleder kan foretage, se f.eks. Tabel 1 i Zhao og McDonald [42].

Resultater De resultater, der kan opnås af en THT afhænger i høj grad af, hvordan og hvilke data der opsamles. En THT kan gennemføres i større eller mindre skala, og med få eller mange hjælpemidler. I komplekse tests kan det være en fordel at bruge et decideret usability-laboratorium, hvor testpersonen og dennes skærm kan optages, der kan optages lyd af forløbet, og indtil flere observatører kan følge med på fjersynsskærme eller gennem et semi-transparent vindue fra et tilstødende lokale. En sådan opsætning kan dog i mange tilfælde virke voldsom på testpersonerne, og testholdet kan derfor med fordel udføre THT for eksempel på testpersonens arbejdsplads eller i en brugers eget hjem. På denne måde virker situationen mere vant, og der er færre forstyrrende elementer. Til gengæld er mulighederne for dataopsamling typisk mindre.

Valget af opsamlede data vil være en vægtning, der skal foretages i hvert enkelt tilfælde, men som hovedregel er det bekvemt for testholdet at have en lydoptagelse af testpersonens kommentarer, som bagefter kan transskriberes og sammenholdes med f.eks. en observatørs noter eller blot de enkelte trin i hver opgave. Man kan med fordel anvende specielt udviklede programmer til tid- og notetagning, så man efterfølgende kan beregne task times, task completion og andre relevante metrikker. Overordnet leverer THT derfor både objektive data (såsom task time, task completion, antal identificerede usabilityproblemer, etc.) såvel som mere subjektive data i form af testpersonens udtalelser, tvivl og

[16] Et eksempel herpå kunne f.eks. være at hjælpe testpersonen til at arbejde med faner i en browser for at teste kommunikation mellem to hjemmesider. Noget som vi vil gøre brug af i opgaverne til denne undersøgelse.

holdninger til produktet. Sidstnævnte kan ofte give anledning til større designændringer end førstnævnte, men de tager dog også væsentlig længere tid at bearbejde og springes ofte over. De subjektive resultater er modsat de objektive heller ikke nemme at prædefinere i form af succeskriterier (de er svære at gøre målbare), hvilket gør det sværere for et testhold at argumentere for den ekstra tid, det tager at behandle dem.

Beskrivelsen af efterbehandlingen af de opsamlede data udskydes til den konkrete analyse i afsnit 5.1. Vi vil blot nævne en væsentlig pointe i forbindelse med fortolkningen af de opsamlede data: oftest vil testlederen sammen med en eller flere eventuelle observatører udgøre det hold, der ud fra f.eks. en videooptagelse af testen skal identificere usabilityproblemer opstået undervejs i forløbet. Jacobsen et al. [17] konkluderer på baggrund af en interessant mindre undersøgelse, at der er markant forskel i typen og antallet af problemer, der findes af forskellige, men alle sammen velkvalificerede evaluatorer, og at antallet af fundne problemer er proportionalt med antallet af evaluatorer. Det vil derfor altid være en fordel at få flere end én kompetent person til at gennemgå resultaterne fra THT'en, da man ellers risikerer at overse muligvis kritiske opståede problemer.

Fordele og ulemper THT er som nævnt tidligere en meget anvendt teknik til at teste produkter fra en bred vifte af kategorier, både inden de udvikles, og inden de lanceres, og har vist sig at give særdeles brugbare resultater og er i sin opbygning forholdsvis simpel at gå til og særdeles veldokumenteret. Der er dog som nævnt ovenfor også en del regler, der bør overholdes. Hvis en eller flere af disse regler brydes, kan THT give misvisende eller decideret ubrugelige resultater.

Den måske væsentligste anke ved THT er muligheden for, at testpersonernes opførsel og opgaveløsning afviger væsentligt fra, hvad der ville være observeret, hvis de havde siddet selv i deres vante omgivelser. Der kan gøres meget for at forhindre dette, hvor det mest oplagte selvfølgelig er at udføre testen i testpersonens egne omgivelser og med dennes egen computer, men i lige så høj grad er det testlederens opgave at få testpersonen til at *føle* sig hjemmevant og nedtone eventuelle konsekvenser eller forstyrrelser som følge af setuppet. Spørgsmålet rejser sig, om hvorvidt den kunstigt opsatte situation har en reel indflydelse på resultaterne af testen. I en undersøgelse af Hertzum et al. [8] påvises det, at i den klassiske form (som defineret af Ericsson og Simon [4]), hvor samtale mellem testleder og testperson minimeres, er indflydelsen på resultaterne minimal. Derimod er der i en mere relakseret udgave, hvor der føres mere dialog mellem de to parter markant forskel i blandt andet task completion-tider og mental arbejdsbyrde. Umiddelbart kræver verbalisering af tanker under opgaveløsning selvfølgelig ekstra hjernekapacitet, og testpersonerne vil derfor være længere tid om at løse opgaverne under en THT end normalt. Dette er i vores øjne forventeligt, og der skal altså korrigeres for dette, hvis man er særligt interesseret i task completion-tider. *Mental workload* blev af testpersonerne også vurderet højere i de tilfælde, hvor der blev tænkt højt. Selvom undersøgelsen baserer sig på et lille antal testpersoner i en meget snæver aldersgruppe, giver den en løftet pegefinger om, at påpasselighed i gennemførslen og analysen af resultaterne af en THT er nødvendigt.

En anden indvending ved THT er, at mange af testpersonens ytringer er uanvendelige i forbindelse med en efterfølgende udpegning af usabilityproblemer, og at databehandlin-

gen derved gøres unødigt kompleks. Zhao og McDonald [42] påviser i deres undersøgelse dog, at *antallet* af brugbare kommentarer[17] stiger, når man tillader interventioner i forhold til den klassiske metode, men at *procentdelen* af disse er nærmest uændret. Der kommer med andre ord mere 'fyld' i en relakseret THT. Omvendt viste undersøgelsen også, at selvom en THT forsøgtes kørt efter de klassiske principper fra Ericsson og Simon [4], kom testpersonerne alligevel med kommenterende eller direkte samtalende ytringer undervejs.

Baggrund for metodevalg til undersøgelse THT er i denne undersøgelse valgt som den metode, der primært skal give kvalitativt input omkring brugervenligheden af NemID. Undersøgelserne er udført nogenlunde i stil med Rubin og Chisnell [31], da denne dels er meget gennemprøvet og giver gode muligheder for at få et indblik i testpersonernes opfattelse og holdning til systemet samt deres forståelse af de underliggende koncepter (såsom single-signon og to-faktorsikkerhed), som de tvinges ud i at anvende og beherske. For de fleste testpersoners vedkommende vil opgaverne ligeledes virke realistiske, da der arbejdes med velkendte hjemmesider, som testpersonerne ofte også har anvendt før og derfor kan relatere til. Altsammen noget som potentielt kan resultere i større verbalisering af processen undervejs.

THT er i dette specifikke tilfælde også en god mulighed for at kunne supplere de særlige grupper af befolkningen, som er underepræsenteret i de resterende undersøgelser, primært fordi de kan være svære at nå ud til. Hvor spørgeskemaerne kan udpege generelle problemområder eller holdninger for store grupper af testpersoner, kan THT udpege specifikke problemer for mere snævre grupperinger.

THT er i form af dets opgaveløsning også væsentlig mere specifik hvad angår udpegning af usabilityproblemer i NemID *under* løsning af nøje udvalgte opgaver, hvorimod spørgeskemaerne primært *retrospektivt* undersøger, hvordan 'opgaven' er forløbet. Her spiller samtalen mellem testleder og testperson også en stor rolle, særligt i debriefingen til sidst i undersøgelsen. I det hele taget spiller det en stor rolle for resultaterne af en test, i hvor høj grad man har adgang til testpersonen, og under en THT er adgangen noget nær optimal. Dog giver THT i dets almindelige form ikke de store muligheder for at gå i dybden med ytringer eller handlinger fra testpersonens side undervejs i forløbet. Vi skal i næste afsnit se på en udvidelse af THT, der tillader dette.

I forhold til dataindsamling og udpegning af usabilityproblemer er THT væsentligt overlegen på detaljegraden, hvorimod spørgeskemaerne er overlegne på antallet af respondenter. På dette punkt er THT ekstremt dyr, tidskrævende og decideret urealistisk at få op på samme antalsmæssige niveau, ligesom det kan være svært at rekruttere nok testpersoner – særligt hvis der ingen gevinst eller kompensation er i det for dem. Selvom vi i dette projekt ingen ressourcer har til rådighed, var det alligevel muligt at rekruttere nok testpersoner, da tidsforbruget per test blev holdt indenfor en rimelig grænse.

[17] Hvor 'brugbare' skal ses som 'af størst værdi for den efterfølgende udpegning af problemer'.

4.1.2 Kooperative evalueringsteknikker

I vores øjne bygger en ikke uvæsentlig del af brugerens oplevelse af et IT-system på, hvor godt de bagvedliggende koncepter forstås og tages i anvendelse. Når IT-systemer ikke opfører sig, som brugeren forventer, er det værdifuldt for en testleder at kunne fastholde situationen og ikke blot fortsætte til næste opgave. Det er en af debriefingens opgaver gennem spørgsmål og tilbageblik at forsøge at undersøge dette, og der findes mange forskellige metoder hertil. Dog kan debriefingen efter vores mening i mange tilfælde komme for sent, ligesom detaljegraden i testpersonens hukommelse hurtigt mindskes, og dermed kan værdifuld information gå tabt. Da vores undersøgelse af NemID ikke primært handler om fejlfinding men er af en mere eksplorativ karakter, ønsker vi at fastholde disse overvejelser og handlinger undervejs i forløbet.

Frøkjær og Hornbæk [5] foreslår en udvidelse af THT, den såkaldte *Cooperative Usability Testing* (CUT), hvor testleder og testperson efter løsning af opgaverne indgår i en mere naturlig dialog med hinanden, så det bliver en fælles opgave at udpege usabilityproblemer. CUT udspringer af ovennævnte ulemper ved THT og består af en *interaction session* (IAS), efterfulgt af en *interpretation session* (IPS) og en afsluttende debriefing. IAS kan eksempelvis bestå af en THT, og debriefingen vil så være kortere, da meget af indholdet ligger i IPS-delen. Denne del gennemføres efter en kort pause efter IAS-delen og bygger på videooptagelsen af denne. Testleder og testperon gennemgår sammen IAS-delen og diskuterer væsentlige punkter undervejs i forløbet, på baggrund af videooptagelsen og observatørens noter undervejs. IPS-delen skal dermed bruges til at opnå en fuld forståelse af de mest kritiske usabilityproblemer og give testpersonen en ekstra mulighed for at forklare deres handlinger foretaget undervejs i testforløbet.

Baggrund for metodevalg til undersøgelse For NemIDs vedkommende ønsker vi at undersøge, hvordan testpersonerne reagerer på og tolker single-signon-konceptet. I og med at single-signon fungerer 'usynligt' for brugeren, er det vigtigt at fastholde de tanker og handlinger, der følger umiddelbart efter introduktionen til konceptet i opgaverne – hvad tænker testpersonen, når noget pludselig sker 'automatisk'. Vi foreslår derfor at anvende en alternativ udgave af CUT til at fastholde disse overvejelser og udvider derfor THT med korte, specifikke sessioner med samtale lige efter de opgaver, som omhandler single-signon og single-signoff. Disse sessioner indledes, så snart opgaven er løst af brugeren og har en varighed på ca. 2-3 minutter. Det primære formål er at få testpersonen til i frisk erindring at genfortælle, hvad der skete i den foregående opgave, og hvorvidt opførslen var forventet. Gennem samtale og uddybende spørgsmål kan testlederen nu få præciseret, om og i så fald hvordan testpersonen opfatter og anvender de afprøvede koncepter. Vi indlægger med andre ord korte IPS-sessioner lige efter udvalgte opgaver.

Disse særlige opgaver vil være af en sådan karakter, at de er forholdsvis korte og lette at udføre, så testpersonens tankevirksomhed i videst muligt omfang kan fastholdes på de koncepter, der ønskes afprøvet, og hele opgavens forløb kan fastholdes i testpersonens erindring. Vi vil ikke anvende metoden til mere almindelige opgaver, hvor testpersonen blot skal udføre en sekvens af handlinger korrekt for at løse opgaven. Her vil debriefingen fortsat være det primære værktøj til afklaring. Vi kommer nærmere ind på den konkrete brug af metoden og de tilhørende opgaver i afsnit 4.2.

4.1.3 Spørgeskemaer

De ovenfor beskrevne metoder er gode til at producere kvalitative data. THT er velkendt for at kunne levere gode resultater, hvis man afsætter tid nok til at gennemlytte eller -se optagelser fra seancerne, så testpersonernes ytringer og holdninger medtages – så undersøgelsen ikke blot bliver en udredning af usabilityproblemer og task completion-tider. Cooperative Usability Testing i vores præsenterede udgave hjælper her sideløbende med at fastholde de indtryk og overvejelser, der er vigtige at få vendt og registreret straks undervejs i forløbet. Som nævnt ovenfor er THT dog en meget dyr og tidskrævende metode, og det vil derfor næsten aldrig være muligt at gennemføre nok tests, til at man kan sige noget statistisk signifikant om resultaterne. Her kommer spørgeskemaerne ind som en noget anderledes og også billigere dataindsamlingsmetode.

Spørgeskemaer findes i et hav af forskellige former, både i forbindelse med form (papirbaseret, elektronisk, læst op, etc.), længde, formål, indhold og spørgsmålstyper. Overordnet kan spørgeskemaer inddeles i tre kategorier (Ozok [26], s1152):

- **Brugers evalueringer:** Her anvendes spørgeskemaet til at få brugerens feedback på f.eks. et softwareprodukt. Her er der oftest tale om et konkret produkt, i vores tilfælde NemID, som brugeren kan relatere til og har prøvet. Her kan spørges til mål, forventninger, og mere direkte eller indirekte til termer som *effectiveness*, *efficiency* og *satisfaction*. Denne type spørgeskema vil ligge til grund for vores mere kvantitative undersøgelse af respondenternes holdning til NemID generelt og brugen af dette.

- **Brugers meninger:** Modsat ovenstående type af spørgeskema er denne møntet på brugerens holdninger og forventninger til mere generelle koncepter og produkter som helhed. Dette kunne for eksempel indeholde forventninger eller ønsker til offentlig selvbetjening i næste årti.

- **Andet:** Den sidste kategori består mestendels af demografiske spørgeskemaer, der ikke i nær så høj grad spørger til respondentens holdninger, men mere til egenskaber og opførsel. Denne type vil vi anvende forholdsvist grundigt til at få det demografiske overblik over respondenternes svar, så det potentielt er muligt at udrede forskelle inden for demografisk eller erfaringsmæssigt adskilte grupper.

I denne undersøgelse vil vi kun anvende første og sidste kategori, selvom et sideløbende studie af NemID-brugernes holdninger til IT generelt ligeledes ville have været interessant.

Karakteristika Normalt skelner man for indsamlingsmetoder mellem papirbaserede (eng: *paper-and-pencil*) og elektroniske spørgeskemaer. De papirbaserede har i mange år været dominerende, men de elektroniske, web-baserede har været på kraftig fremmarch. De udgjorde i 1999 ca. 15% af alle udførte spørgeskemaundersøgelser – et tal, der i 2004 var vokset til 70% og i dag må formodes at ligge endnu højere. (Ozok [26], s1168)

For alle typer gælder dog, at spørgeskemaet og dets spørgsmål i videst muligt omfang skal opfylde nogle helt generelle betingelser, her hentet fra Stone [34]:

- **Passende:** De stillede spørgsmål skal være i stand til at give svar på de spørgsmål, der ønskes undersøgt. Heri ligger også, at unødvendige spørgsmål skal undgås, så længden og dermed gennemførselstiden for spørgeskemaet kan minimeres.
- **Forståelige:** Spørgsmålene skal sikkert kunne forstås af den målgruppe, man udsender spørgeskemaet til. Det giver for eksempel kun et lille udbytte at bruge tunge tekniske termer i en undersøgelse, som den brede befolkning skal deltage i. På den anden side er der fare for at falde i den anden grøft og ubevidst komme til at tale ned til respondenternes forståelse eller intelligens ved at overforklare eller udpensle spørgsmål eller bruge unødigt store mængder tekst til indledninger og lignende.
- **Entydige:** Spørgsmålene og de tilhørende svar skal udformes således, at de for det første har samme betydning for spørger og adspurgte, og for det andet ikke på nogen måde kan misforstås. Et spørgsmål som "I hvilken del af landet bor du?" er upræcist formuleret og kan give anledning til massive udsving i svarene (Nordsjælland, Thisted, Region Syddanmark, ...), selvom spørgeren måske havde forestillet sig, at respondenterne automatisk ville svare, hvilken region de bor i. Her vil man typisk give respondenten relevante svarmuligheder for at undgå tvetydighed.
- **Uforudindtagede:** Ingen af spørgerens (eller nogen andres) holdninger må komme til udtryk i spørgsmålene. Spørgeren må altså ikke tilskynde til én type svar frem for andre.
- **Omnikompetente:** Spørgsmålene skal kunne dække alle tænkelige svarmuligheder. Brugeren skal altså i alle tilfælde kunne placere sit svar inden for en i spørgsmålet given mulighed. Føler brugeren sig stadig i tvivl ved at placere sit svar, kan man tilføje muligheden "Andet" eller "Ved ikke", men dette vil særligt i sidstnævnte tilfælde ikke være af stor værdi for spørgeren. Efter Stones mening bør denne tvivlsfaktor dog ikke ignoreres.
- **Korrekt kodede:** Et eksempel på en klassisk fejl her er spørgsmålet "Hvor gammel er du?" med svarmulighederne f.eks. "Under 20 år", "20-30 år" og "Over 30 år". Hvis man er 20 eller 30 år gammel, hvilket svar bør man så give? Som hoveregel skal svarmulighederne derfor være *udtømmende* og *uforenelige*, med tydelige skel.
- **Præ-testede:** Ethvert spørgeskema bør og skal testes på gerne flere testpersoner, inden det sendes ud. Dette er i særlig grad gældende, hvis spørgeskemaet skal udsendes til mange hundrede eller enddog tusinde respondenter. Her kan en lille fejl potentielt ekspandere til frustrationer eller tvivl hos et stort antal brugere, hvilket igen kan føre til fejl i resultaterne eller en højere frafaldsprocent (respondenter, som har startet men ikke gennemført undersøgelsen).
- **Etiske:** Undersøgelsen skal kunne redegøre for dens berettigelse, tidsforbrug, dens videnskabelige belæg og metode samt dens anvendelse af de opsamlede data. God skik dikterer desuden, at potentielle respondenter på forhånd informeres om disse faktorer, samt hvordan deres anonymitet og svar sikres og behandles.

Svartyper og redskaber Værktøjskassen, som anvendes til at opbygge og strukturere spørgeskemaer har udviklet sig gevaldigt, efter elektroniske spørgeskemaer kom på banen. Det har dog generelt ikke ændret voldsomt på de typer af spørgsmål, der stadig primært

anvendes. Hovedindholdet i et spørgeskema er selvfølgelig spørgemålene, og disse findes i et væld af udgaver, alt efter det ønskede svar. Vi gengiver her nogle af dem, vi senere selv kommer til at anvende, og præsenterer dem i form af deres svarmuligheder. De fleste svartyper kan principielt suppleres med en 'Ved ikke'-mulighed.

- *Tekstfelter*, hvor respondenten i princippet kan skrive hvad som helst, f.eks. navn eller beskæftigelse. Der er som udgangspunkt frit slag i forhold til, hvad feltet kan indeholde, men spørgsmålet kan også være specifikt (f.eks. alder eller antal hjemmeboende børn), hvor brugeren så forventes at indtaste et tal. En sådant typekrav kan ofte håndhæves i elektroniske spørgeskemaer. Tekstfelter findes også i længere udgaver, hvor brugeren f.eks. kan skrive kommentarer eller brainstorme på et bestemt emne.

- *Ja/nej-spørgsmål*, hvor de to oplagte muligheder eventuelt kan suppleres med et 'Ved ikke'. Heri ligger det implicit, at brugeren kun vælger én af de viste svarmuligheder. I modsætning til papirbaserede spørgeskemaer kan dette håndhæves i elektroniske, hvor man ofte vil anvende radiobuttons til formålet. I det helt simple tilfælde kan en afmærket checkboks markere et (aktivt) ja-svar, hvorimod en uafmærket vil symbolisere et (potentielt passivt) nej-svar. Ja/nej-spørgsmålet er en forsimplet udgave af et mere generelt *multiple-choice* spørgsmål, hvor respondenten skal vælge én blandt flere muligheder.

- *Multipel multiple-choice* er den variant af ovenstående, hvor brugeren kan vælge én eller flere muligheder. I den elektroniske udgave vil man hertil anvende checkbokse, ligesom det ofte vil være nødvendigt med en 'Andet'-kategori, evt. med mulighed for at skrive i et tekstfelt, for at sikre dækning af alle svarmuligheder.

- *Skalaer* anvendes ofte til at tilkendegive graden af enighed i det stillede udsagn. Til skalaen kan anvendes tal, gradbøjninger af enighed, omfang eller lignende, eller simpelthen en linie, hvor respondenten bliver bedt om at sætte et kryds et sted mellem to navngivne poler. Sidstnævnte kan være svær og tidskrævende at afkode i papirbaserede spørgeskemaer. I den elektroniske udgave anvendes ofte radiobuttons eller sliders. Skalaernes udformning er et meget omdiskuteret emne, men som tommelfingerregel skal der være et neutralt midtpunkt, hvilket i de fleste tilfælde vil betyde, at skalaen har et ulige antal svarmuligheder (oftest 5 eller 7). Grunden hertil er primært, at respondenter, som f.eks. hverken er enige eller uenige i det pågældende udsagn, skal have en relevant svarmulighed. Et eksempel, hvor dette *ikke* bliver fulgt, kan ses i Figur 6 herunder. Tullis og Albert [35] lægger skalaerne i kategorien *intervaldata*, hvis der er et meningsfyldt midtpunkt mellem to svarmuligheder på skalaen, og de forskellige svarmuligheder kan siges at have nogenlunde samme 'afstand'. En skala, hvor der ikke nødvendigvis er samme afstand mellem svarmuligheder karakteriseres som *ordinale data*. For ordinale data kan der beregnes frekvenser, mens det kun for intervaldata giver mening at beregne gennemsnit, hvis svarmulighederne hver tildeles et tal indenfor et bestemt interval. Forskellige skalaer (som kan fortolkes som ordinale data eller intervaldata, alt efter forgodtbefindende) kan ses i Tabel 58.3 i Ozok [26].

- *Matrix-spørgsmål* er en samling af spørgsmål med samme svarmuligheder, f.eks. skalaer. Denne type er meget kompakt, men kan spare en del plads og kan være

nemmere at overskue, hvis der er mange spørgsmål med samme svarmulighed lige efter hinanden.

Herudover er der i de elektroniske spørgeskemaer endnu større mulighed for at inddrage multimedier (billeder, lyd og video).

Figur 6: Eksempel på spørgsmål med manglende neutralt midtpunkt. Desuden er der flere klik end højst nødvendigt i nederste del – brugeren kan blot præsenteres for et tekstfelt, som han eller hun kan vælge at skrive kommentarer i eller ej. Spørgsmålene er fra brancheforeningen FTFs Ungdomsundersøgelse 2011.

Der findes et hav af programmer og hjælpemidler til at konstruere og designe både papirbaserede og elektroniske spørgeskemaer, ligesom der også er et stort udvalg af statistiske programmer og visualiseringsværktøjer, der kan hjælpe med efterbearbejdning og præsentation af data. Et stort antal firmaer tilbyder også at designe og gennemføre sådanne undersøgelser for privatpersoner, foreninger eller virksomheder. Som udgangspunkt er man dog som spørger nødt til enten at lægge en væsentlig sum penge på bordet eller gøre et stort stykke arbejde selv, hvis man ønsker at udvise høj professionalitet.

Med hensyn til udvælgelsen af spørgsmål findes inden for HCI flere 'præfabrikerede' løsninger til måling af forskellige faktorer. Heriblandt kan nævnes forskellige modeller til måling af brugertilfredshed med forskellige produkttyper. Tullis og Stetson [37] sammenligner en håndfuld forskellige, heriblandt *SUS-skalaen* (*System Usability Scale*, Brooke [2]), som er en samling af 10 skalaspørgsmål, der kan omsættes til en score mellem 0 og 100. Scoren angiver, hvor god usability det bedømte produkt har, med 100 som bedste score. SUS-spørgsmålene spørger modsat de andre studerede metoder kun til brugerens *overordnede* reaktion på produktet, og har desuden det laveste antal spørgsmål. Den anvendte femtrinsskala går fra "Meget uenig" til "Meget enig", og SUS har i undersøgelsen vist sig at være mest pålidelig med hensyn til resultater opnået med et forskelligt antal respondenter. Halvdelen af spørgsmålene er positivt ladede, halvdelen er negativt ladede, og disse spørgsmål kommer skiftevis og giver en svarværdi mellem 1 og 5. Scoren

beregnes på følgende måde:

$$SUS = 2.5 \left(\sum_{n \text{ ulige}} (svar_n - 1) + \sum_{n \text{ lige}} (5 - svar_n) \right)$$

Tullis og Stetson [37] konkluderer dog, at man har brug for mindst 12-14 respondenter for overhovedet at kunne komme med pålidelige resultater. Vi skal senere anvende SUS-skalaen i et af spørgeskemaerne.

Rekruttering af respondenter Med hensyn til at udvælge de rette personer til at svare på ens spørgeskema gælder selvfølgelig de samme principper og procedurer som for THT. Selve måden, rekrutteringen foregår på, kan dog være væsentlig anderledes. For papirbaserede spørgeskemaer skal respondenten have skemaet i hånden, udfylde det med blyant eller kuglepen, og på en eller anden måde returnere det til spørgeren (f.eks. via en frankeret svarkuvert eller aflevering i postkasse på arbejdspladsen). Spørgeskemaerne kan enten uddeles af spørger, ligge tilgængelige i mapper eller på borde eller distribueres med intern eller almindelig post. For elektroniske spørgeskemaer er der flere muligheder. Skemaet kan fremsendes per mail som f.eks. et Word-dokument, der skal udfyldes og returneres per mail igen. Der kan også fremsendes et link til et web-baseret spørgeskema, som brugeren kan udfylde på en hjemmeside indrettet til formålet. Dette link kan også lægges på hjemmesider, blandt andet på chatfora og lignende, samt udsendes med medlemsblade eller nyhedsbreve for at sikre større udbredelse, ligesom mange analyseinstitutter har opbygget omfattende databaser med internetbrugere, der aktivt har tilmeldt sig som potentielle respondenter. Adgang hertil er dog ofte en bekostelig affære. En anden fordel ved elektroniske skemaer er, at der indledningsvist kan screenes for, om respondenten er i den ønskede målgruppe eller ej. Hvis ikke kan undersøgelsen afbrydes på stedet, uden større gene for respondenten.

Det nødvendige antal respondenter vil variere meget, alt efter med hvor stor sikkerhed man ønsker at konkludere noget på baggrund af resultaterne, samt hvor stor en del af den samlede population man ser på. Laves der komparativ statistik på resultaterne, ved f.eks. en adskillelse på køn eller alder, skal man derfor sikre sig, at de opdelte grupper ikke bliver for små.

Fordele og ulemper Fordelene ved de elektroniske spørgeskemaer er oplagte: de koster ikke noget at masseproducere og udsende, store mængder data fra tusindvis af respondenter kan indsamles, bearbejdes og visualiseres nogenlunde med det samme, og de anvender elementer, som er kendt af alle, der har bare en anelse erfaring med internettet. Som nævnt ovenfor har den elektroniske udgave af et spørgeskema visse fordele fremfor den papirbaserede i forhold til svartyper. En væsentlig generel fordel er dog også, at elektroniske spørgeskemaer kan håndhæve og gøre respondenten opmærksom på, hvis der ikke er svaret på et obligatorisk spørgsmål og endda henlede opmærksomheden på det pågældende spørgsmål, f.eks. ved at farve spørgsmålet rødt. Den umiddelbart største ulempe ved elektroniske spørgeskemaer er distributionen. I disse tider med stort fokus på spam-mails, virus og lignende kan det være svært for spørgere at få hul igennem til et stort nok antal respondenter. Det kan ligeledes være svært rent praktisk at skaffe emailadresser på større befolkningsgrupper, med mindre man får et link med i en betroet

kilde, f.eks. et nyhedsbrev. Desuden er det selvfølgelig nærmest umuligt at få besvarelser fra folk, som ikke anvender internettet.

For de papirbaserede spørgeskemaer er fordelen, at de er særdeles velkendte. Enhver formular eller ansøgning, der skal udfyldes, vil indeholde væsentlige elementer fra spørgeskemaer, og de fleste mennesker vil nikke genkendende, omend til tider også opgivende, til sådanne opgaver. Den store ulempe ved de papirbaserede er dog klart omkostningerne ved at producere, distribuere og indsamle skemaerne. Noget som de elektroniske ikke i samme grad lider under.

Generelt kan det også ske, at respondenter ikke tager undersøgelsen seriøst, hvilket kan være meget svært at opdage, ligesom en præ-test ikke kan garantere at fange alle uklarheder og afvigelser fra de ovenstående retningslinier. Er karakteristika om hver respondent ikke gode nok, vil der også være mulighed for, at forskelle på baggrund af disse overses, ligesom det selvfølgelig ikke kan tjekkes, om respondenter svarer ærligt og korrekt på spørgsmålene. Af fordele kan derimod generelt nævnes, at spørgeskemaer tillader hurtig indsamling af store mængder data fra en bred gruppe af mennesker på en forholdsvis ligetil måde, der ligeledes kan sikre anonymitet af respondenten. I opbygningen af skemaerne kan man som ovenfor nævnt bygge på veludviklet teori og standardisering af data (Robson [30], s233f). En sidste fordel er, at man mindsker eller helt undgår *social desirability bias*, altså at respondenter føler sig forpligtet til at svare mere positivt, når der sidder en testleder og observerer eller spørger. På grund af fraværet af en sådan i et spørgeskema, der udfyldes selvstændigt af respondenten, minimeres dette fænomen (Tullis og Albert [35], s126f).

Baggrund for metodevalg til undersøgelse Elektroniske spørgeskemaer er som ovenfor nævnt medtaget som den primære metode til at indsamle kvantitative data om brugeres opfattelse og anvendelse af NemID på grund af deres simplere distribution, mindre *social desirability bias* og efterfølgende lettere databehandling. De belaster desuden respondenter væsentligt mindre, end hvis man eksempelvis skulle indsamle data fra 25-50 THT'er. Selvom skemaer kan tage meget lang tid at udfærdige, er de øjeblikkeligt skalerbare til selv et meget stort antal respondenter. Det er også langt mere bekvemt for respondenten at kunne besvare spørgsmålene hvor som helst og når som helst, ligesom vi også vil anvende metoden af to omgange. Først udsendes et større skema, som tager sig af brugerdata, aktiveringsproces og generelle holdninger og problemer. Dernæst udsendes på et senere tidspunkt et opfølgende kortere skema til interesserede respondenter, som aktivt har sagt ja til at deltage. Dette opfølgende skema har til hensigt at måle den mere langsigtede tilfredshedsgrad med produktet for at se, om der sker ændringer i den samme respondentgruppe. Dette giver gode muligheder for sammenligning af tidsmæssigt adskilte data baseret på de samme spørgsmål.

Som en sidste detalje vil vi også anvende et papirspørgeskema i slutningen af hver THT til at indsamle demografiske oplysninger om testpersonen samt få svar på nogle af de samme spørgsmål, som bliver stillet i den elektroniske udgave. Vi vil også i dette spørgeskema lade testpersonen bedømme produktet ud fra en let modificeret SUS-skala. Mere herom i næste afsnit.

4.2 Planlægning af tænke-højt test

Vi har i det foregående afsnit præsenteret de testmetoder, som vi ønsker at benytte i de undersøgelser, som skal ligge til grund for resten af projektet. Vi skal i dette afsnit tilpasse metoderne til den aktuelle kontekst samt opbygge det konkrete testmateriale, der skal anvendes til de forskellige typer af undersøgelser. Undervejs argumenterer vi for de mange valg, der er truffet undervejs i opbygningsprocessen.

4.2.1 Definition af testopgaver og succeskriterier

I dette afsnit vil vi definere de opgaver, der skal stilles til testpersonerne i tænke-højt testen. Indledningsvist husker vi på, at den første opgave skal være let for testpersonen at løse. Det har været vigtigt at konstruere opgaverne således, at færrest mulige personfølsomme oplysninger kommer frem under forløbet af opgaveløsningen. Dette hensyn er selvfølgelig valgt udelukkende af hensyn til brugeren – det gør det mere trygt for testpersonen at indleve sig i testen, ligesom det er med til at styrke undersøgelsens professionalitet og validitet og gøre det ufarligt for brugeren at deltage. I tilfældet NemID kan det dog ikke fuldstændig undgås, at visse mere eller mindre personfølsomme oplysninger kommer til syne under opgaveløsningen – nøglekortet i sig selv er et eksempel på dette. I opgaverne kan der dog tages hensyn til dette ved i videst mulig omfang at holde sig fra bestemt typer af data, f.eks. oplysninger om konti og andet på netbanken, ligesom de fleste myndigheders selvbetjeningssystemer kan bruges til så komplekse ting, at brugeren først ledes ind på en (i forhold til personfølsomme oplysninger) forholdsvis harmløs forside efter at være logget ind. Det er desuden en væsentlig pointe, at vi ikke ønsker at teste brugervenligheden af de pågældende sider, men af NemID som autentifikationssystem. Opgaverne skal derfor ikke beskæftige sig indgående med navigation og lignende på de besøgte sider, ligesom der skal ydes tilstrækkelig hjælp i opgaveformuleringen til, at dette ikke bliver et problem.

Som nævnt i indledningen, har vi i dette projekt ikke mulighed for at se på samtlige aspekter af NemID. Vi vælger at fokusere på grundydelserne i systemet og ser bort fra de mere specielle og avancerede muligheder som medarbejdersignaturer, krypteret email, etc. De generelle koncepter ved NemID, som ønskes testet, er følgende:

- **Aktiveringsprocessen:** Mange medier har skrevet om, hvor besværligt det er overhovedet at komme i gang med NemID, og efter egne oplevede problemer med netop aktiveringsprocessen virker det naturligt at give denne et forholdsvist stort fokus i opgaverne. På denne måde sikrer vi også, at vi har at gøre med brugere, der aldrig har anvendt systemet.
- **Single-signon / single-signoff:** Dette koncept er blevet slået op som en af de helt store fordele ved NemID, selvom det i sig selv blot giver adgang til, at en mere overordnet tjeneste kan holde styr på logins til forskellige tjenester på én gang. Det er dog stadig interessant at undersøge, hvordan brugerne oplever og forstår disse to features, hvorvidt de er i stand til at anvende dem og føler sig trygge ved dem. Single-signon vil desuden efter vores mening uvægerligt blive koblet sammen med NemID.
- **NemIDs selvbetjening:** Denne er medtaget primært for at observere, om brugeren er i stand til at lokalisere og anvende muligheden for at personliggøre sit

Navn	Beskrivelse	Tid	Succes
1. Gennemlæsning af velkomstbrev	Brugeren gennemlæser velkomstbrevet, som han eller hun normalt ville gøre det med breve af denne type.	2-5	Velkomstbrevet er læst igennem.
2. Aktivering af NemID	Brugeren aktiverer først sit NemID og anvender det derefter til at logge ind på TastSelv på SKATs hjemmeside.	5-15	Brugeren kan aktivere sit NemID anvende det til at logge ind på SKAT.
3. Single-signon	Brugeren opretholder login på SKAT og logger i en ny fane ind på sundhed.dk for at finde sin læges adresse. Der kræves her *ikke* nyt logon.	2-3	Brugeren kan finde sin læges adresse uden af logge ind på sundhed.dk med sit NemID.
4. Single-signoff	Brugeren logger ud fra SKAT og skal herefter tilbage på "Min personlige forside" på sundhed.dk for at tilpasse den til egne behov. Her *kræves* nyt logon.	2-5	Brugeren kan foretage nyt logon på sundhed.dk.
5. NemID.nu selvbetjening	Brugeren bliver bedt om at bestille et ekstra nøglekort til en længere ferierejse (der gives ikke information om hvor og hvordan dette gøres).	2-10	Brugeren kan logge ind på NemID.nu's selvbetjening og lokalisere bestilling af ekstra nøglekort.

Tabel 2: Testopgaver udarbejdet til tænke-højt tests. Tid (i minutter) er forventet, og succes er simpelt defineret som at være i stand til at gennemføre opgaven.

NemID ved at foretage ændringer som at ændre adgangskode, spærre eller bestille ekstra nøglekort og lignende.

- **Learnability:** Det må forventes at tage forholdsvist længere tid, første gang testpersonerne skal bruge deres NemID. Vi ønsker derfor indirekte at måle *learnability* og *retention over time* ved at indlejre flere logins med systemet i løbet af testforløbet.
- **Generel tilfredshed med systemet:** Dette vil naturligt være en del af forløbet i enhver THT, men for NemIDs vedkommende har der på forhånd været en overvejende negativ indstilling til systemet, i hvert fald hvis man ser på mediernes og internettets dækning og kommentarer. Vi ønsker med andre ord at afdække brugerens *subjective satisfaction* med systemet.

Til at dække ovenstående områder er der udformet fem opgaver til testpersonerne i THT'en, som kan ses i Tabel 2. De kan ses i deres fulde længde i testplanen i Bilag A og i de til testpersonerne udleverede opgaver i Bilag D.

Opgave 1 og 2 dækker aktiveringsprocessen, opgave 3 og 4 dækker single-signon /

single-signoff, og opgave 5 dækker NemIDs selvbetjening. *Learnability* og *retention over time* er indkorporeret i opgave 4 og 5, og den generelle tilfredshed (*satisfaction*) dækkes løbende i forløbet gennem testpersonens kommentarer og ytringer, de indlagte CUT-sessions samt den efterfølgende debriefing og spørgeskema. Opgaverne er formuleret med så lille som mulig brug af tekniske termer og med tilstrækkelig hjælp til, at delopgaver uden tilknytning til NemID kan løses problemfrit. Eftersom der måles tider og registreres problemer undervejs i forløbet, vil disse opgaver eksplicit teste *efficiency* og *effectiveness* af systemet.

Hjemmesiderne tastselv.skat.dk og sundhed.dk er valgt, da indgangene hertil er forholdsvis direkte og uproblematiske, ligesom det er muligt at foretage simple handlinger, uden at større mængder personfølsomme oplysninger vises på skærmen. De er ligeledes begge med i det offentlige login-fællesskab, NemLog-in, hvilket muliggør single-signon.

Succeskriterierne er simple (kan testpersonen gennemføre opgaverne uden at anvende ekstern hjælp, f.eks. i form af DanIDs telefonsupport), og de er kombineret med forventede gennemførselstider på opgaverne. Disse må grundet THT'ens natur forventes at være en smule længere end for et virkeligt scenarie uden krav om at tænke højt. Testpersonerne er, inden opgaverne påbegyndtes, blevet bedt om at tilkendegive fuldførelse af den enkelte opgave ved at give opgaveteksten tilbage til testlederen. Herved kan aktuelle gennemførselstider lettere måles i den efterfølgende databehandling.

Testpersonerne er forud for testen blevet informeret om, at de undervejs vil få aktiveret og anvendt deres NemID. I løbet af opgave 2 vil de, afhængigt af hvordan de har modtaget deres NemID, desuden blive bedt om at tage stilling til, hvorvidt de vil have tilknyttet et OCES-certifikat til deres NemID, så dette kan anvendes hos offentlige myndigheder i tillæg til deres netbank. Testleder har gjort opmærksom på, at af hensyn til de videre opgaver er det nødvendigt at give dette samtykke. Forinden er testpersonerne dog blevet informeret om, at denne beslutning er reversibel, og testlederen kan efterfølgende give assistance hermed. I det hele taget er opgaverne udformet således, at der (udover selve aktiveringen) ingen reelle ændringer bliver foretaget, og opgaverne stoppes, så snart de er ét klik fra at udføre den ønskede handling.

4.2.2 Forløb, lokation, dataindsamling og testplan

Der er som støtte til testlederen og som generel dokumentation for forløbet udarbejdet en detaljeret testplan, som kan ses i Bilag A. Testplanen er lavet efter modellen i Rubin og Chisnell [31]. Vi vil her gengive de væsentligste pointer.

I forhold til lokation for testen har to placeringer været under overvejelse: Testpersonens eget hjem eller i testfaciliteter på Søndre Campus på KU. Valget faldt modsat DanIDs egne tests ud til den første mulighed, idet vi gerne vil eliminere flest mulige forstyrrelser udefra og lade testpersonen agere i vante og kendte omgivelser. Lige så vigtigt er det, at testpersonen anvender sin egen computer, som vedkommende er fortrolig med.

Under forløbet af testen er der en testleder samt en observatør til stede. Testlederens rolle er at guide og om nødvendigt hjælpe testpersonen, styre forløbet med introduktion, opgaver og tid, samt stå for debriefing, CUT-sessions og spørgeskema. Testlederen

er placeret skråt bagved testpersonen under forløbet og tager egne noter efter behov. Observatøren er placeret mere tilbagetrukket i forhold til testpersonen med frit udsyn til skærmen og tager undervejs fyldige og tidsstemplede noter til kommentarer, hændelser på skærmen, opgavers start- og sluttidspunkter og andre relevante observationer. Til notetagning er anvendt et Excel-ark med forprogrammeret makro til at registrere tidspunkt for indtastninger. Testleder og observatør har undervejs i det samlede forløb byttet roller. Udover Excel-arket med observatørens notater samt testlederens egne notater er seancerne blevet lydoptaget med en diktafon. Videooptagelse blevet fravalgt; dels af praktiske årsager, men primært af hensyn til testdeltagernes privatliv og hemmeligholdelse af personfølsomme oplysninger. Det er ligeledes blevet vurderet, at en videooptagelse i dette konkrete tilfælde ikke vil bidrage med større mængder relevant information og desuden vil forhøje efterbearbejdningstiden væsentligt.

Som hjælp til testlederen og for at sikre, at samtlige testpersoner gives præcis de samme informationer, er der udarbejdet et manuskript (igen efter modellen i Rubin og Chisnell [31]), som kan ses i Bilag B. Der er desuden efter inspiration fra Rubin og Chisnell [31] og Molich [22] udarbejdet en samtykkeerklæring, som underskrives af testpersonen, inden selve testen påbegyndes. Denne forklarer anvendelsen af de opsamlede data samt garanterer testdeltagerens anonymitet. Samtykkeerklæringen kan ses i Bilag C.

Som en del af debriefingen er der som nævnt i afsnit 4.1.3 brugt et seperat papirbaseret spørgeskema, der ligeledes agerer som screener og indsamling af demografiske data om testpersonen. På første side af spørgeskemaet findes spørgsmål om alder, køn, oprindelsesland, modersmål, beskæftigelse, om den gamle signatur er anvendt samt en subjektiv vurdering (fra "Meget uerfaren" til "Meget erfaren") af testpersonens egne IT-færdigheder. Desuden er der seks holdningsspørgsmål (fra "Meget uenig" til "Meget enig") til, hvorvidt NemID er let at oprette, let at bruge, sikkert, praktisk, sikrere end tidligere løsninger og lettere end tidligere løsninger. Side to består af de 10 SUS-spørgsmål, oversat til dansk og med 'NemID' indsat på de relevante pladser. Spørgeskemaet kan ses i Bilag E.

Som beskrevet i afsnit 4.1.2 er der efter opgave 3 og 4 indlagt korte CUT-seancer med det formål at fastholde og få uddybet testpersonens forståelse af single-signon og single-signoff. Her vil testlederen stille nogle få opklarende spørgsmål til den netop overståede opgave, f.eks. "Prøv at genfortælle, hvad der skete i denne opgave" og "Reagerede systemet, som du havde forventet?". Det er vores forventning, at disse seancer vil have indflydelse på testpersonens forventninger til udfaldet af opgave 4, men dette ses som tåleligt i forhold til de informationer, som seancerne kan tilvejebringe.

4.3 Planlægning af første spørgeskema

Som supplement til de primært kvalitative data, der indsamles i THT'erne udfærdiger vi to spørgeskemaer, som udsendes til den bredest mulige målgruppe. Det første spørgeskema skal tage sig af aktivering og første brug af NemID samt indsamle demografiske oplysninger om respondenterne. Vi spørger også til respondentens holdning til systemet, umiddelbart efter at aktiveringen er gennemført. Disse spørgsmål og andre statistiske opsamlinger gentages i det opfølgende spørgeskema, som vi beskriver nærmere i afsnit 4.4. Den måske største barriere for at anvende spørgeskemaer i dette projekt er, at

der ikke er afsat nogen form for ressourcer til at belønne respondenter for deres deltagelse, eller til at promovere (markedsføre/distrubuere) selve spørgeskemaet. På baggrund af dette er der gjort en del overvejelser omkring form og gennemførsel af undersøgelsen, som vi vil referere her.

Der er fra starten truffet beslutning om at anvende web-baserede elektroniske spørgeskemaer af flere årsager. Dels er de gratis og lette at distribuere og indsamle data fra, de kan udfyldes når som helst og hvor som helst af respondenten, og der findes et væld af elektroniske hjælpemidler og værktøjer til at udforme skemaerne. Efter at have afprøvet mange systemer, faldt valget til sidst på Google Docs[18], som er en gratis online-baseret kontorpakke, hvis regnearksdel indeholder en del funktionalitet til at håndtere spørgeskemaer. Google Docs blev valgt på grund af, at det er gratis, det indeholder de krævede spørgsmålstyper, det er utrolig let at bruge, det genererer automatisk et link til spørgeskemaet, som kan videresendes i f.eks. en email, og det eksporterer automatisk de indsamlede data til Excel til videre forarbejdning. Andre tjenester tager i omegnen af $99 om måneden og kan være umådeligt komplekse at anvende, ligesom nogle tjenester endda kræver delte rettigheder til de indsamlede resultater. Google Docs har ligeledes gode muligheder for tilpasning af design, brug af hjælpetekster og lignende, ligesom der er validering af manglende udfyldte felter og mulighed for betingede hop i skemaet. Desuden er interfacet på dansk, både for respondenten og spørgeren, som skal konstruere skemaet.

Selve indholdet af det første spørgeskema består som nævnt tidligere af demografiske og IT-relaterede spørgsmål for at kunne kategorisere respondenten, samt af efterfølgende spørgsmål til aktiveringsprocessen og NemID generelt. Disse spørgsmål besvares bedst samtidig med, at aktiveringen foregår, eksempelvis i en separat fane eller browservindue. Da processen, for hvordan man aktiverer sit NemID, afhænger af, hvordan man har modtaget det, og respondenter ikke skal tvinges til at svare på spørgsmål om trin i processen, de ikke har kunnet nå frem til pga. problemer undervejs, er det besluttet at lave betingede hop mange steder undervejs i skemaet. Vi gengiver i Tabel 3 herunder de konkrete spørgsmål, deres svartype og hvilket koncept eller område, der skal dække. Selve spørgeskemaet, som det er blevet udfyldt af respondenten, kan ses i Bilag K.[19] Vi vil ikke komme ind på begrundelser for hvert eneste spørgsmål, blot nævne at der har ligget mange timer og revisioner til grund for det endelige skema. Der er tilføjet korte forklarende hjælpetekster til langt de fleste spørgsmål. Disse kan ligeledes ses i Bilag K.

[18] Se: http://docs.google.com

[19] Spørgeskemaet kan ses og testes på adressen:
https://spreadsheets.google.com/viewform?formkey=dHkyLVM1Y2dSdlRJZVh4NHVWcWZnM0E6MQ

Tabel 3: Liste med spørgsmål i indledende spørgeskema. Efter svarmulighederne er angivet deres type og interne kodning. Kriterium er enten demografisk, IT-relateret eller falder normalt inden for de på side 41 opstillede koncepter. Obligatoriske spørgsmål er markeret med fedt nummer.

	Spørgsmål	**Svarmuligheder og -type**	**Kriterium**
		Demografiske spørgsmål	
1	Alder	(tekstfelt)	Demografi
2	Køn	Mand / Kvinde (multiple-choice)	Demografi
3	Oprindelsesland	Danmark / Andet (multiple-choice)	Demografi
4	Modersmål / førstesprog	Dansk / Andet (multiple-choice)	Demografi
5	Længste gennemførte uddannelse	Folkeskole / Gymnasial uddannelse / Kort videregående uddannelse / Mellemlang videregående uddannelse / Lang videregående uddannelse (multiple-choice, 1-5)	Demografi
6	Beskæftigelse	(tekstfelt)	Demografi
7	Eventuelle funktionsnedsættelser?	Ja / nej (multiple-choice, 1/0)	Demografi
8	Hvis ja, hvilke?	(langt tekstfelt)	Demografi
		Spørgsmål til brugerens IT-færdigheder	
9	IT-færdigheder	"Meget uerfaren" til "Meget erfaren"(skala, 0-4)	IT-færdigheder
10	Primære operativsystem	Windows / Mac OS X / Linux / Andet/Ved ikke (multiple-choice)	IT-færdigheder
11	Anvendelse af internettet, frekvens	Dagligt / Ugentligt / Månedligt / Sjældnere / Har aldrig anvendt (multiple-choice, 0-5)	IT-færdigheder
12	Anvendelse af internettet, steder	Hjemme / Arbejde / Uddannelsessted / Internetcaféer / Biblioteker / Andre steder (multiple-choice)	
13-17	Anvendelse af tjenester (Netbank/homebanking, TastSelv på www.skat.dk, www.borger.dk, www.sundhed.dk, www.eBoks.dk)	Dagligt / Ugentligt / Månedligt / Sjældnere / Har aldrig anvendt (multiple-choice i matrix, 0-4)	IT-færdigheder
18	Anvendt gammel digital signatur?	Ja / Nej / Ved ikke (multiple-choice, 1/0/2)	IT-færdigheder
		Velkomstbrevet	
19	Hvordan er NemID modtaget?	Automatisk tilsendt / Selv bestilt / Selv hentet i bank (multiple-choice, 0/1/2)	Aktivering
20	Læst velkomstbrevet?	Ja, helt / Ja, delvist / Nej (multiple-choice, 2/1/0)	Aktivering
21	Hvis nej, hvorfor ikke?	(langt tekstfelt)	Aktivering, *subjective satisfaction*

Fortsættes på næste side...

Fortsat fra forrige side...

	Spørgsmål	Svarmuligheder og -type	Kriterium
22-26	Enighed i udsagn om velkomstbrev (let at forstå, interessant, længde tilpas, indhold relevant, giver nok information)	"Meget uenig" til "Meget enig"(skalaer i matrix, 0-4)	*Subjective satisfaction*
27-30	Enighed i udsagn om nøglekort (overskueligt, praktisk, god størrelse, tilpas mængde koder)	"Meget uenig" til "Meget enig"(skalaer i matrix, 0-4)	*Subjective satisfaction*
31	Yderligere kommentarer til velkomstbrev eller nøglekort	(langt tekstfelt)	Aktivering, *subjective satisfaction*
	Aktivering (ens for netbank og selv bestilt)		
32	Lykkedes at oprette adgangskode	Ja / Nej (multiple-choice, 0/1)	Aktivering, *effectiveness*
33-37	Enighed i udsagn om oprettelse af adgangskode (let at oprette, regler lette at forstå, forstod opgaven, vinduet overskueligt, behov for mere hjælp)	"Meget uenig" til "Meget enig"(skalaer i matrix, 0-4)	*Subjective satisfaction*
38	Valgt selvvalgt bruger-id?	Ja / Nej (multiple-choice, 0/1)	Aktivering
39	Hvis nej, hvorfor ikke?	(langt tekstfelt)	Aktivering, *subjective satisfaction*
40	Fundet og indtastet korrekt kode fra nøglekort?	Ja / Nej (multiple-choice, 0/1)	Aktivering
41	Hvis nej, hvorfor ikke?	(langt tekstfelt)	Aktivering, *subjective satisfaction*
42-45	Enighed i udsagn om indtastning af nøgle (forstod opgaven, vinduet var overskueligt, symbolerne gav mening, behov for mere hjælp)	"Meget uenig" til "Meget enig"(skalaer i matrix, 0-4)	*Subjective satisfaction*
46	Lykkedes at aktivere NemID?	Ja / Nej (multiple-choice, 0/1)	Aktivering, *effectiveness*
47	Aktiveret gennem hvilken tjeneste?	Netbank/homebanking / TastSelv på www.skat.dk / www.borger.dk / www.sundhed.dk / www.eBoks.dk / Andet (multiple-choice)	Aktivering
	Grunde til problemer (kun hvis sådanne er opstået undervejs)		
48	Hvilke problemer opstod?	(langt tekstfelt)	Aktivering, *effectiveness*

Fortsættes på næste side...

Fortsat fra forrige side...

	Spørgsmål	Svarmuligheder og -type	Kriterium
49	Hvad kunne have hjulpet undervejs for at undgå problemerne?	(langt tekstfelt)	Aktivering, *effectiveness*
50	Afholde fra at anvende NemID i fremtiden?	Ja / Nej (multiple-choice, 0/1)	*Subjective satisfaction*
	NemID generelt og afrunding		
51-56	Enighed i udsagn om NemID generelt (let at oprette, let at bruge, sikkert, praktisk, sikrere end tidligere løsninger, lettere end tidligere løsninger).	"Meget uenig" til "Meget enig"(skalaer i matrix, 0-4)	*Subjective satisfaction*
57	Yderligere kommentarer til NemID generelt	(langt tekstfelt)	Aktivering, *subjective satisfaction*
58	Deltagelse i lodtrækning	Emailadresse eller telefonnummer (tekstfelt)	-
59	Opfølgende spørgsmål	Jeg må kontaktes (checkbox, 1/0, default nej)	-

Spørgeskemaet består af i alt 59 spørgmål, hvor langt størstedelen kan besvares med et enkelt klik. Ud af disse 59 spørgsmål er de 47 obligatoriske, og 3 spørgsmål skal kun besvares, hvis der opstår problemer undervejs. Alt efter, hvorvidt aktiveringsprocessen forløber problemfrit, bør dette spørgeskema kunne udfyldes sideløbende med processen på cirka 8-10 minutter.

Som det kan ses i Tabel 3 er der gjort stor brug af Likert-skalaer gående fra "Meget uenig" til "Meget enig". Dette er bevidst valgt, da man efterfølgende så vil kunne sætte tal på respondenternes holdninger, ligesom skalaerne kun kræver et enkelt klik at besvare. Som tidligere nævnt vil vi opfatte disse data som intervaldata. Man kunne i stedet have anvendt lange tekstfelter til et spørgsmål som f.eks. "Hvad synes du generelt om nøglekortet?", men dette ville have givet ikke-målbare svar og ville have drejet undersøgelsen i en mere kvalitativ retning. I stedet har vi valgt at lade brugeren kommentere i nogle få lange tekstfelter undervejs i undersøgelsen (spørgsmål 21, 31, 39, 41 og 57). Det er ligeledes blevet fravalgt at anvende *semantiske differentialer*[20], da formuleringen af mange af de ønskede spørgsmål i så fald ville blive mere eller mindre kunstig og bebyrde respondenten unødigt.

Mange af spørgsmålene reflekterer de mulige problemområder, som NemID kan tænkes at have, og som medier og fora har udsat for stor kritik: nøglekortet, velkomstbrevet, sikkerheden og aktiveringsprocessen. Nøglekortet og velkomstbrevet behandles særskilt, og aktiveringsprocessen følges detaljeret med spørgsmål undervejs i forløbet. Sikkerheden er mere subjektiv og behandles kun kort i de generelle spørgsmål til NemID. Spørgsmålene er lavet, så de kommer i præcis samme rækkefølge som trinene i aktiveringsprocessen,

[20] En semantisk differentiale består normalt i at bede brugeren om at sætte et kryds på en (måske trinløs) skala mellem bipolære adjektiver, f.eks. 'surt' og 'sødt'.

og respondenten opfordres til at gennemføre undersøgelse sideløbende med aktiveringen for at kunne svare så præcist og korrekt på spørgsmålene som muligt. Angiver respondenten, at der undervejs i processen er gået noget galt, ledes respondenten hen til en fejlside, hvor problemerne kan beskrives og vurderes, hvorefter spørgeskemaet afsluttes som normalt. Her overspringes altså de trin, som brugeren grundet problemerne ikke er i stand til at svare på. Spørgsmålene 51-56 er særlig væsentlige, idet de går igen i det opfølgende spørgeskema. Hermed har vi mulighed for at monitorere eventuelle ændringer i respondenternes holdning til produktet.

Til sidst i spørgeskemaet er indlagt et spørgsmål om, hvorvidt respondenten ønsker at deltage i konkurrencen om en iPod Shuffle. Denne udlodning er medtaget, da dette efter forfatterens egen mening vil kunne motivere flere til at deltage i undersøgelsen. Bosnjak og Tuten [1] konkluderer i deres undersøgelse af forskellige mulige incitamentsmetoder for spørgeskemaundersøgelser følgende:

> *"On a practical note, the results do add credence to the popular method of using prize draws as incentives in web surveys. Based on this study, this method is the most sound in terms of generating a strong response rate and high data quality. It is possible, though, that this effect is somehow based on the current 'culture' of surveys. In other words, culture studies emphasize that individuals learn a culture through socialization practices. Because many survey studies are not funded at levels that provide the opportunity for individual monetary incentives for each participant, many studies do rely heavily on prize drawings to encourage response. Therefore, it is possible that individuals have been socialized by this practice to anticipate the use of a prize drawing rather than an individual, guaranteed reward. This, too, could help to heighten the effect of trust — or distrust — as the case may be."*
>
> Bosnjak og Tuten [1], s216

Ozok [26] er ligeledes enig i, at der for længere spørgeskemaer bør tilbydes en form for kompensation, hvilket kan medvirke til højere return rates, ligesom en maksimum på 15 minutter foreslås for gennemførsel af et spørgeskema, hvortil der ikke ydes kompensation.

Endelig findes her også koblingen til det opfølgende spørgeskema, eftersom respondenter her har mulighed for at give aktivt samtykke til at måtte blive kontaktet per email med opfølgende spørgsmål. På denne måde kan vi i det opfølgende spørgeskema holde os inden for den samme population. Spørgeskemaet er testet på en repræsentativ respondent inden udsendelse, hvilket kun gav anledning til små ændringer og omformuleringer. Testpersonen tog cirka 8 minutter om at gennemføre skemaet. Antallet af spørgsmål per side er for overskuelighedens skyld og for at højne fornemmelsen af fremdrift holdt så lavt som muligt.

4.4 Planlægning af opfølgende spørgeskema

Som nævnt i forrige afsnit er det efterfølgende spørgeskema målrettet mod de respondenter fra det første, som aktivt har sagt ja tak til at måtte blive kontaktet. Det opfølgende spørgeskema er (ligesom det første) lavet i Google Docs og løbende udsendt som kort-

link[21] per mail til respondenterne cirka 3-4 uger efter, de har aktiveret deres NemID. Dette er muligt at gøre, da Google Docs automatisk registrerer dato og tid for, hvornår de enkelte respondenter har udfyldt spørgeskemaerne.

Dette spørgeskema er meget kort og har primært som formål at spørge til de generelle holdninger til NemID igen (spørgsmål 51-56 i Tabel 3), men indsamler samtidig også information om eventuelle problemer under den daglige brug af NemID samt antal anvendelser og tid siden aktivering. Spørgsmålene er gengivet i Tabel 4. Igen er der tilføjet forklarende hjælpetekster til alle spørgsmål, og respondentens udgave af spørgeskemaet kan ses i Bilag N[22]. Dette spørgeskema tager cirka 2-3 minutter at gennemføre og er

	Spørgsmål	**Svarmuligheder og -type**	**Kriterium**
1	Uger siden aktivering	(tekstfelt)	Statistik
2	Antal anvendelser siden aktivering	(tekstfelt)	Statistik, *learnability*
3	Tjenester, som NemID er anvendt på	Netbank/homebanking, TastSelv på www.skat.dk, www.borger.dk, www.sundhed.dk, www.eBoks.dk, www.su.dk, Andre (multipel multiple-choice)	*Efficiency (learnability)*
4	Oplevet problemer siden aktivering?	Ja / nej (multiple-choice, 1/0)	*Effectiveness*
5	Hvis ja, hvilke?	(langt tekstfelt)	*Effectiveness*
6	NemID er let at bruge	"Meget uenig" til "Meget enig" (skala i matrix, 0-4)	*Subjective satisfaction*
7	NemID er sikkert	"Meget uenig" til "Meget enig" (skala i matrix, 0-4)	*Subjective satisfaction*
8	NemID er praktisk	"Meget uenig" til "Meget enig" (skala i matrix, 0-4)	*Subjective satisfaction*
9	NemID er sikrere end tidligere metoder	"Meget uenig" til "Meget enig" (skala i matrix, 0-4)	*Subjective satisfaction*
10	NemID er lettere end tidligere metoder	"Meget uenig" til "Meget enig" (skala i matrix, 0-4)	*Subjective satisfaction*

Tabel 4: Liste med spørgsmål i opfølgende spørgeskema. Efter svarmulighederne er angivet deres type og interne kodning. Det testede koncept er enten statistisk eller falder inden for de generelle usability-kriterier. Obligatoriske spørgsmål er markeret med fedt nummer.

ligeledes blevet præ-testet inden udsendelse. Prætesten gav ikke anledning til yderligere ændringer. Det mest tidskrævende for respondenten i denne delundersøgelse kan være at

[21] Det blev valgt at konvertere det meget lange Google Docs-link til et kortlink for begge spørgeskemaers vedkommende for lettere at kunne sprede undersøgelsen via sociale medier og fora på nettet. Se mere herom i afsnit 4.5.

[22] Eller prøves på adressen: https://spreadsheets.google.com/viewform?formkey=dGZPeDNTOFMxSm42SVk3cUU3d1htU1E6MQ

finde ud af, hvor mange gange han eller hun har anvendt sit NemID siden aktiveringen. Her er der risiko for, at respondenterne kommer med estimater eller skriver tekst i felterne, der så skal efterbehandles manuelt, hvilket også gør sig gældende for spørgsmålet med antal uger siden aktivering.

4.5 Udvalg og rekruttering af målgrupper

I dette afsnit beskriver vi, hvordan målgrupperne for de forskellige undersøgelser er udvalgt og rekrutteret. Da NemID henvender sig til samtlige personer i den danske befolkning over 15 år, burde man rent principielt selvfølgelig teste et repræsentativt udvalg af hele denne målgruppe. Dette er dog inden for rammerne af projektet utopisk, og vi indsnævrer derfor udvalget en smule. For spørgeskemaernes vedkommende er disse konstrueret, så hele populationen har mulighed for at svare. Dette er også at foretrække, da vi ønsker at nå ud til en respondentgruppe, der er så repræsentativ for befolkningen som muligt. Vi er dog på forhånd realistiske og forventer ikke, at dette vil kunne opfyldes fuldt ud med et budget på nul kroner. I de nedenstående afsnit beskriver vi kort målgruppen og rekrutteringen af denne for hver af de tre undersøgelser.

4.5.1 Tænke-højt test

Til tænke-højt ønskes rekrutteret otte testpersoner plus en ekstra til præ-test af forløbet. Det er blevet besluttet at rekruttere personer fra de fire grupper vist i Tabel 5. Det tilstræbes desuden, at mænd og kvinder er ligeligt repræsenteret, samt at aldersfordelingen er så bred som muligt. En særlig målgruppe, som ikke er medtaget i Tabel 5, er de blinde og svagtseende samt personer med funktionsnedsættelser eller handicap. Fra starten havde vi en intention om at inkludere en eller flere af disse i testen, pga. mediernes og diverse organisationers kritik af NemIDs anvendelighed for de pågældende grupper. Det blev dog vurderet, at rekruttering og gennemførsel af en sådan test dels ville afvige markant fra de resterende tests, dels ville forberedelse og research til en sådan test være en meget stor opgave, da det særskilte system udviklet til især blinde mennesker afviger en hel del fra det normale system, ligesom personer med funktionsnedsættelser eller handicap kan kræve særlige faciliteter til rådighed. Mere om de konkret rekrutterede personer i afsnit 5.1.

For at rekruttere fra de ovenstående målgrupper er vi gået gennem forskellige kanaler:

- **Eget netværk:** Rekrutteringsopslag er lagt ud på Facebook og formidlet gennem forfatterens eget netværk af kolleger, familie og venner. Denne rekrutteringsmetode har leveret hovedparten af testpersonerne.
- **ÆldreSagen:** ÆldreSagen har været behjælpelig med kontakt til respondenter i gruppe B.
- **Ungdomsorganisationer:** Der er taget kontakt til Dansk Ungdoms Fællesråd (DUF) og videre til Danske Gymnasieelevers Sammenslutning (DGS) med henblik på rekruttering til gruppe A. Begge havde dog ikke umiddelbart mulighed for at skaffe testpersoner, og derfor blev gruppe A rekrutteret gennem eget netværk.
- **Muslimernes Fællesråd:** Muslimernes Fællesråd er blevet kontaktet med henblik på rekruttering til gruppe C, men nåede dog ikke at svare på henvendelsen, før

Gruppe	Navn	Beskrivelse	Antal
A	Unge	Unge mellem 15 og 18 år. Denne gruppe har oftest et stort forbrug af og føler sig hjemmevante med digitale tjenester og udstyr og må formodes at "klare sig godt" i denne test. De er dog ikke nødvendigvis rutinerede netbankkunder eller fortrolige med offentlige IT-systemer eller muligheder med disse. DanID har efter vores oplysninger ikke udført tests med denne målgruppe.	2
B	Ældre	Personer over 65 år, som ikke længere er på arbejdsmarkedet. Denne gruppe er medtaget, da bl.a. ÆldreSagen har været ude med kritik af NemID for at være for svært at aktivere og anvende for ældre mennesker.	2
C	Anden etnisk baggrund	Personer i vilkårlig alder med en anden etnisk baggrund end dansk. Denne gruppe er medtaget for at teste, om begreber, koncepter og processer opfattes anderledes end hos etniske danskere. Med anden etnisk baggrund menes at være født og (måske kun delvist) opvokset i et andet land and Danmark.	2
D	Den brede befolkning	Personer i alderen 25-60 år, som fortrinsvist er i arbejde. Denne population udgør den 'almindelige' bruger af systemet, og her vil vi tilstræbe, at der rekrutteres personer med forskellig alder, køn, beskæftigelse eller lignende.	2 (3)

Tabel 5: Målgrupper og antal for tænke-højt tests.

testen var gennemført. Rekruttering til gruppe C foregik derfor ligeledes gennem eget netværk.

4.5.2 Første spørgeskema

Rekrutteringen til det første spørgeskema er som ovenfor tiltænkt at ramme bredest muligt. Indledningsvis blev der taget kontakt til 2-3 mellemstore almene fagforeninger, da disse har et bredt kundegrundlag og sandsynligvis (i modsætning til de helt store og fagspecifikke fagforeninger) vil være mere tilbøjelige til at hjælpe med distribution af spørgeskemaet. Ingen af de kontaktede fagforeninger var dog interesserede i at hjælpe, selvom en enkelt fagforening måtte sige nej, da den for øjeblikket selv var ved at udbrede awareness om NemID til dens kunder.

Efter grundigt at have overvejet alternative foreninger, organisationer og virksomheder blev det besluttet at sprede spørgeskemaet udelukkende via internettet. Udbredelsen af sociale medier, chatfora og lignende blandt de fleste befolkningsgrupper gør dette til en potentiel guldgrube med hensyn til rekruttering af respondenter til undersøgelser. Der er ligeledes også fora om næsten alle temaer på internettet, hvilket potentielt mu-

liggør en mere specifik rekruttering fra bestemte aldersgrupper, faglig baggrund eller politisk/religiøs overbevisning. I Tabel 6 er gengivet de sociale medier og internetfora, hvor der er blevet postet en invitation til at deltage i undersøgelsen, samt deres antal registrerede brugere (og dermed potentielle respondenter). Invitationen indeholdt en kort information om undersøgelsen samt et kortlink til startsiden på forfatterens egen server. Fordele, ulemper og forslag til best practice ved denne fremgangsmåde analyseres nærmere i afsnit 6.2.2.

Navn	Adresse	Registrerede brugere
Facebook	www.facebook.com (oprettet åben begivenhed)	> 2.5 mio.[(a)]
Jubii Debatforum	www.dindebat.dk > Computere og Internet	143.111
DelDig.dk	www.deldig.dk/forum > Livsstil og relationer > Job & økonomi	4.228
MorsVerden.dk	www.morsverden.dk/debat > Andre debatter > Åbent forum	2.944
Gul & Gratis	www.guloggratis.dk/forum > Diverse	814.224
MobilDebat.dk	www.mobildebat.dk > Diverse > Fri debat	35.287
Lav-Det-Selv.dk	www.lav-det-selv.dk/Forum > Blandet > Over hækken	24.444
Opslagstavle.com	www.opslagstavle.com	0[(b)]

Tabel 6: Tabel over sociale medier og internetfora, som første spørgeskema er blevet spredt på.
[(a)] Kilde: http://www.facebakers.com/facebook-statistics/denmark. Tallet er fra 29. november 2010 og angiver antallet af registrerede danske Facebook-brugere.
[(b)] Opslagstavle.com er en elektronisk opslagstavle, hvor der kan slås alskens opslag op. Dets registrerede brugere er selve opslåerne, ikke de besøgende.

4.5.3 Opfølgende spørgeskema

Rekrutteringen til det opfølgende spørgeskema er ligetil, da målgruppen består af de respondenter, som i det første spørgeskema har sagt ja tak til at måtte blive kontaktet med opfølgende spørgsmål. Disse respondenter har, cirka 3-4 uger efter de har svaret på det første spørgeskema, fået tilsendt en mail med et kortlink til det opfølgende spørgeskema. Det har desværre ikke været umiddelbart praktisk muligt at sammenholde besvarelser fra de to spørgeskemaer. Dette ville enten kræve uhyggelig meget arbejde fra forfatterens side eller unødigt bøvl for respondenten – med en efterfølgende potentiel stigning i drop-off rate.[23] Der er efterfølgende blevet udsendt en email med en reminder om at svare på det opfølgende spørgeskema.

Opsamling Vi har i dette afsnit set nærmere på tre forskellige metoder til brugerundersøgelser: teænke-højt, *cooperative usability testing* og spørgeskemaer, som hver på

[23] Ratio af respondenter, der hopper fra mellem de to spørgeskemaer.

deres egen måde kan bidrage til en usability-evaluering af NemID. Vi har undervejs udlagt og beskrevet den metodiske fremgangsmåde samt sat fordele og ulemper overfor hinanden, ligesom vi har set nærmere på de data, som metoderne kunne tænkes eller ønskes at opsamle. Desuden har vi gjort en del ud af målgrupper og rekruttering for de enkelte undersøgelser, som er meget forskellig og kan være en begrænsende faktor for undersøgelsernes succes.

5 Analyse af evalueringsresultater

I dette afsnit bearbejder og analyserer vi de data, der er indsamlet i de forskellige undersøgelser. Vi vil ikke diskutere resultaterne og drage overordnede konklusioner her – dette sker i stedet i diskussionen i afsnit 6. For hver undersøgelse beskriver vi først de deltagende respondenter, hvorefter vi bearbejder og analyserer de indsamlede data.

5.1 Tænke-højt test

Vi vil først karakterisere de ni testpersoner, hvorefter vi analyserer den forholdsvis store mængde data, der er opsamlet. Undervejs vil vi identificere, klassificere og gruppere de fundne usabilityproblemer. En meget stor del af dette bygger på det store arbejde, som Kasper Hansen har lavet for respondenterne P01-P08 i sit projekt om samme emne. Han har ligeledes fungeret som skiftevis observatør og testleder på de ni udførte THT'er og efterfølgende været behjælpelig med bearbejdningen af den omfattende mængde opsamlede data. Se eventuelt hans resultater i Hansen [7].

Der er som ovenfor nævnt blevet udført i alt ni THT'er. De otte heraf er blevet udført i testpersonens egne vante omgivelser (seks i hjemmet og to på arbejdspladsen), og en enkelt er blevet udført i et til lejligheden lånt lokale på Datalogisk Institut. Samtlige testpersoner anvendte deres egne computere til testen, og testene blev udført i perioden fra den 4. november til den 2. december 2010. Samtlige respondenter underskrev uden problemer samtykkeerklæringen og udtalte efterfølgende, at det havde været spændende at deltage i undersøgelsen. Testpersonernes identitet er forfatteren bekendt, men de er af hensyn til eventuelle personfølsomme oplysninger i teksten benævnt P01-P09.

P01 blev oprindeligt rekrutteret til indledningsvist at teste selve forløbet og opgaverne. Denne test gav dog ikke anledning til større ændringer, og resultaterne af testen er derfor medtaget som en ekstra besvarelse i gruppe D.

5.1.1 Karakteristik af testpersoner

I Tabel 7 kan ses karakteristika for de ni testpersoner. Der er rekrutteret to testpersoner til gruppe A, én til gruppe B, to til gruppe C og fire til gruppe D. Fordelingen er ikke helt som ønsket, men der er dog en repræsentation af de fleste aldersgrupper, flere typer etnicitet og modersmål, og selvom kønsfordelingen er skæv til kvindernes fordel, er begge køn stadig repræsenteret. Det havde været godt med en mandlig testperson i gruppe B, men dette var desværre ikke muligt at skaffe i tide. Ligeledes havde det været godt med en testperson i alderen 50-60 år, samt en bedre alders- og kønsspredning i gruppe C. Overordnet er vi dog under de givne forhold tilfredse med rekrutteringen og fordelingen af testpersoner til denne undersøgelse, da forskelle i både alder, køn, etnicitet og uddannelsesniveau er repræsenteret. Med hensyn til testpersonernes IT-erfaring er denne subjektiv fra deres egen side, og vi er ud fra testens forløb ikke nødvendigvis enige i disse vurderinger – fordelingen er sandsynligvis bredere, end Tabel 7 indikerer. Vi vil nu for hver testperson give en kort beskrivelse af personen og det pågældende testforløb.

P01, gruppe D
Yngre kvinde. Testen blev udført i eget hjem og på egen computer. Havde ikke umid-

	Alder	Køn	Modersmål	Beskæftigelse	DigSig	IT-færdigheder
P01	40	K	Dansk	Akkrediteringskonsulent	Nej	Erfaren
P02	48	K	Dansk	Kontorsekretær	Ja	Erfaren
P03	34	K	Dansk	Fuldmægtig i bogholderi	Ja	Neutral
P04	31	M	Dansk	Radiovært	Ja	Neutral
P05	24	K	Arabisk	Studerende	Ja	Neutral
P06	16	K	Dansk	Gymnasieelev	Nej	Neutral
P07	24	K	Arabisk	Studerende	Ja	Erfaren
P08	86	K	Dansk	Pensionist	Ja	Neutral
P09	16	M	Dansk	Gymnasieelev	Ja	Erfaren

Tabel 7: Liste med karakteristika for THT-testpersoner. DigSig indikerer, hvorvidt testpersonen har anvendt den gamle digitale signatur. IT-færdigheder går på femtrinsskalaen fra "Meget uerfaren" til "Meget erfaren".

delbart nogen særlig holdning hverken for eller imod NemID, men i slutningen af testen nævnte hun, at en indledende skepsis overfor produktet havde vendt sig til at "være ok". Opgaverne blev efter lidt teknisk bøvl løst hurtigt og uden større problemer.

P02, gruppe D
Midaldrende kvinde, der gennemførte testen i arbejdstiden på sin arbejdsplads og ved brug af sin arbejdscomputer. Havde indledningsvist en meget påpasselig indstilling til sikkerhed ved færden online, og egne erfaringer blev detaljeret præsenteret. Til selve NemID var holdningen dog neutral – noget som generelt forblev uændret, dog viste der sig en stor usikkerhed omkring single-signon/-signoff konceptet i forbindelse med deling af og adgang til personfølsomme oplysninger. Opgaverne blev løst uden større problemer.

P03, gruppe D
Yngre kvinde, som gennemførte testen i eget hjem og på egen computer. Opgaverne blev løst særdeles hurtigt, og der kom ikke ytringer hverken for eller imod NemIDs virkemåde eller opbygning. Under forløbet kiggede samleveren med over skulderen og kom nogle gange til at ytre kritik at systemet. Det blev henstillet af testleder at undgå dette, og det vurderes ikke at have haft indflydelse på testpersonens holdning eller opførsel under forsøget.

P04, gruppe D
Yngre mand, som gennemførte testen i eget hjem og på egen computer. Var på forhånd kritisk indstillet overfor NemID og det offentliges adgang til og opbevaring af personfølsomme oplysninger, hvilket også kom til udtryk i kommentarer og ytringer undervejs i testforløbet. Accepterede modvilligt at tilknytte et offentligt certifikat og fik efterfølgende også hjælp til at fjerne dette tilsagn igen. Løste opgaverne uden større problemer.

P05, gruppe C
Ung kvinde med anden etnisk baggrund end dansk, som gennemførte testen i et til lejligheden lånt lokale, men dog med egen computer. Testpersonen virkede meget fortrolig med IT og udviste stor villighed til at forsøge sig med opgaverne. Nåede dog ikke længere end til opgave 2, som ikke kunne løses, hvilket bekymrede testpersonen, som troede, at

det var hendes fejl. Hun udviste dog stor vilje til at forsøge igen og ville ikke give op, da NemID jo var "vigtigt" og "vejen frem".

P06, gruppe A
Ung kvinde, som gennemførte testen i eget hjem og på egen computer. Går til daglig i en gymnasial IT-klasse og virker desuden også meget fortrolig med IT, uden dog at kalde sig erfaren bruger. Testpersonen udviste ikke nogen større interesse for NemID, og opgaverne blev løst pånær en task failure på opgave 2, som testleder var nødsaget til at hjælpe testpersonen videre med (efter opkald til DanIDs telefonsupport).

P07, gruppe C
Yngre kvinde med anden etnisk baggrund end dansk, som gennemførte testen på egen computer i et lånt lokale i en forening, som hun selv var medlem af og udførte IT-relateret arbejde for. Virkede meget fortrolig med IT generelt og udtrykte fra starten bekymring over sikkerheden i NemID. Undervejs i opgave 2 virkede det som om, testpersonen allerede havde aktiveret sit NemID, selvom hun forsikrede os om, at dette ikke var tilfældet. Hvis dette er korrekt, udgør det et særdeles interessant tilfælde i forhold til NemID, og ellers må opgaven siges ikke at kunne bidrage med særlig relevant information.

P08, gruppe B
Ældre pensioneret kvinde, som udførte testen i eget hjem og på egen computer. Trods personens alder virkede hun umiddelbart fortrolig med brugen af en computer, internettet og sin netbank. NemID blev anset som noget, man "bare skulle med på", og der sporedes hverken konkret modstand eller glæde ved processen. Undervejs måtte testleder intervenere en del gange på grund af en smule forvirring fra testpersonens side om det videre forløb, og derfor nåede testperson også kun at gennemføre opgave 1 og 2. Testpersonen var villig til at gennemføre de resterende opgaver på et senere tidspunkt også, men dette vurderede vi ikke at være nødvendigt, ligesom testpersonen efter egen mening "nok skulle finde ud af det der NemID".

P09, gruppe A
Ung mand, som gennemførte testen i eget hjem og med egen computer. Gik i samme IT-klasse som P06 og virkede særdeles hjemmevant inden for IT. Anvendte knap så ofte netbank, men andre online tjenester dog en del. Kunne ikke umiddelbart se det store personlige behov for NemID og oplevede desuden præcis samme task failure i opgave 2 som P06 og måtte derfor assisteres af testleder for at kunne gennemføre resten af opgaverne, som blev løst uden større problemer.

Samtlige testpersoner anvendte Microsoft Windows som styresystem (som minimum Windows XP) og gennemførte alle testen ved brug af Internet Explorer som browser. Alle computere havde ligeledes præ-installeret Java.[24] Yderligere forhold omkring testpersonernes IT-faciliteter blev ikke undersøgt nærmere.

[24] En faktor, som ellers kunne have været interessant at inkludere i undersøgelsen. I det hele taget må alle de anvendte computere og operativsystemer siges at have opfyldt NemIDs tekniske krav: https://www.nemid.nu/support/tekniske_krav/understoettet_software/.

5.1.2 Dataanalyse

Under forløbet af THT'erne blev der opsamlet forskellige typer data: en elektronisk log i Excel med tidsstempler på indføringer, en lydoptagelse af forløbet samt et afsluttende spørgeskema. Vi vil nu kort beskrive, hvordan data fra de forskellige typer er blevet efterbehandlet samt diskutere kvaliteten af de opsamlede data.

De elektroniske logs, som er den primære kilde til identifikation af usabilityproblemer i et senere afsnit, er efterfølgende blevet gennemlæst for de værste stavefejl og misforståelser. Der er desuden blevet beregnet task times ud fra tidsstemplerne i loggen, ligesom task failures er blevet skemasat for de forskellige respondenter. Der er efterfølgende lavet deskriptiv statistik på relevante faktorer. De elektroniske logs kan ses i Bilag F, og data og deskriptiv statistik kan ses i Bilag G. Gennemførselstider og task completion rates er gengivet i Tabel 8 herunder.

	P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**Gns.**
Opg. 1 løst	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	-
Opg. 1 tid	3.75	0.75	1.50	2.25	4.75	1.00	1.00	1.50	1.75	2.02
Opg. 2 løst	Ja	Ja	Ja	Ja	Nej	Nej	Ja	Nej	Nej	-
Opg. 2 tid	16.75	13.00	6.25	17.75	22.00	33.50	7.00	64.00	34.00	23.81
Opg. 3 løst	Ja	Ja	Ja	Ja	-	Ja	Ja	-	Ja	-
Opg. 3 tid	5.00	7.00	2.25	1.75	-	1.25	1.25	-	1.25	2.82
Opg. 4 løst	Ja	Ja	Ja	Ja	-	Ja	Ja	-	Ja	-
Opg. 4 tid	3.00	2.75	1.75	2.75	-	1.25	2.25	-	2.00	2.24
Opg. 5 løst	Ja	Ja	Ja	Ja	-	Ja	Ja	-	Ja	-
Opg. 5 tid	9.75	5.00	5.00	5.50	-	1.75	5.00	-	1.75	4.82
Sum	38.25	28.50	16.75	30.00	26.75	38.75	16.50	65.50	40.75	35.71

Tabel 8: Task completion rates og task completion times (afrundet til nærmeste kvarte minut) fra THT-forløb.

Som det kan ses af Tabel 8 er langt hovedparten af tiden gået med opgave 2, som omhandler den konkrete aktivering og første brug af testpersonens NemID. Der kan ligeledes observeres en stor spredning i tidsforbruget hos testpersonerne, hvor meget dog kan forklares ved fejl undervejs eller manglende forståelse for de anvendte fagtermer og koncepter. Opgave 5 har også vist en større spredning i task completion times – primært grundet respondenternes forskelligartede adfærd under fremsøgning af NemIDs selvbetjening, hvor nogle googler med det samme, nogle kigger i velkomstbrevet, nogle kigger på netbanken, etc. Generelt skal det dog også nævnes, at THT i sig selv vil give anledning til højere task completion times end normalt, da selve tænke-højt konceptet for det første pålægger testpersonen et større *mental workload*, ligesom testleder ikke fuldt ud kan facilitere, at testpersonen går 'den direkte vej' gennem opgaverne uden for meget udenomssnak. Det er dog vores vurdering, at denne udenomssnak udgør en forholdsmæssigt lige stor del hos samtlige testpersoner.

Vi kan ligeledes se, at fire ud af de ni respondenter ikke selvstændigt kunne have gennemført aktiveringen (og dermed heller ikke de efterfølgende opgaver) – der er med andre ord observeret task failure på opgave 2 i 44% af tilfældene. Grundene hertil vender vi tilbage til i afsnit 5.1.3. Samtlige respondenter, som nåede så langt, har med succes

gennemført de resterende opgaver.

Lydoptagelsen af de enkelte seancer er blevet semi-transskriberet i den forstand, at kun relevante udtalelser og svar på spørgsmål er blevet fuldt transskriberet, med henvisning til tidspunkt og kontekst/opgave. Formålet hermed er at kunne referere til udvalgte citater i analyse- og diskussionsfasen, og disse udtalelser har ligeledes tjent som afklaring i forbindelse med uklarheder i loggen. Lydoptagelsen har generelt været af en sådan kvalitet, at al tale fra samtlige tilstedeværende har kunnet aflyttes fuldt ud. De semi-transskriberede seancer kan ses i Bilag H.

Det afsluttende spørgeskema er blevet indtastet i Excel, hvorefter der er blevet lavet deskriptiv statistik på udvalgte faktorer. Herudover er SUS-scoren for hver respondent blevet beregnet som anvist i Brooke [2]. Resultaterne af spørgeskemaet kan ses i Bilag G. Udvalgte faktorer fra spørgeskemaet kan ses i Tabel 9 herunder. Middelværdien for SUS-scoren givet af de 9 testpersoner er beregnet til 54.4 (SD=15.25) ud af et maksimum på 100.0.

	P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**Gns.**
Let at oprette	4	5	4	4	2	4	5	3	3	3.78
Let at bruge	3	5	4	3	2	4	5	3	3	3.56
Sikkert	2	4	3	2	3	4	1	3	2	2.67
Praktisk	2	4	3	4	5	1	1	4	4	3.11
Sikrere end tidligere metoder	3	4	3	1	3	3	1	4	2	2.67
Lettere end tidligere metoder	2	3	3	3	2	2	1	5	4	2.78
SUS-score (0-100)	75.0	67.5	62.5	32.5	35.0	42.5	57.5	67.5	50.0	54.4

Tabel 9: Svar på generelle spørgsmål om NemID samt SUS-score givet af THT-testpersoner. Svarene på de generelle spørgsmål går på en skala fra "Meget uenig" (1) til "Meget enig" (5) med "Neutral" (3) som neutralt punkt.

5.1.3 Problemidentifikation

Vi vil i dette afsnit kort beskrive proceduren for identifikation af usabilityproblemer[25] og -observationer ud fra den førte log for hver testperson. Det skal bemærkes, at identifikationen udelukkende er baseret på loggen – de semitransskriberede optagelser af seancerne er af tidsmæssige årsager ikke gennemgået for usabilityproblemer. Her ville muligvis kunne identificeres flere observationer, men de fleste af disse ville efter vores vurdering ikke være nye. Transskriberingerne tjener derfor mest som supplement til loggen samt ytringsbaseret illustration af pointer omkring de i loggen identificerede problemer. Identifikationen for testpersonerne P01-P08 er udført af Kasper Hansen, og P09 er udført af forfatteren selv. For at gøre identifikationen konsistent, er de samme kriterier for udvælgelse benyttet. Disse er følgende (her frit efter Hornbæk og Frøkjær [9], s507):

1. Brugeren udtrykker et mål og kan ikke nå det inden for tre minutter.
2. Brugeren giver eksplicit op.

[25] Vi skal på forhånd undskylde, at vi uvægerligt vil komme til at bruge betegnelserne 'observation' og 'problem' i flæng – også når der er tale om *positive findings*.

3. Brugeren udtrykker et mål og er nødt til at afprøve tre eller flere handlinger for at finde en løsning.
4. Brugeren fremkommer med et resultat, der afviger fra den stillede opgaves.
5. Brugeren udtrykker overraskelse.
6. Brugeren udtrykker en negativ holdning eller karakteriserer noget som et problem.
7. Brugeren kommer med et designforslag (til ændring).
8. Der opstår et systemnedbrud.
9. Evaluator generaliserer en gruppe allerede fundne problemer til et nyt problem.

Valget af ovenstående identifikationsnøgle medfører, at samtlige *potentielle* usabilityproblemer kommer i betragtning – også dem, som er meget små og måske ubetydelige, men observeres disse hos flere brugere, er der belæg for at bibeholde dem i det senere arbejde med klassifikation og gruppering. Desuden vil vi også identificere positive ytringer fra brugerne som 'problemer', nærmere defineret som *positive findings* (se næste afsnit). Disse skal selvfølgelig ikke ses som en negativ opdagelse, men kan være relevante at få med for at afkode brugernes positive erfaringer med systemet.

Under udpegningen er der i loggen tilføjet kolonnen 'Label', som angiver placering og reference til det fundne problem. Samtlige udpegede usabilityproblemer tilføjet en kort beskrivelse kan ses i Bilag I. Logreferencen er desuden efterfulgt af en reference til hvilken kontekst, problemet er identificeret i. Her betyder CUT3 og CUT4 hhv. CUT-seancen efter opgave 3 og opgave 4, og DEBR er den afsluttende debriefing. Eksempelvis vil referencen 'P - P07-L03 (CUT3)' henvise til et problem i kategori P (se næste afsnit), opdaget hos P07 som problem nummer 3 i dennes log, i forløbet under CUT-seancen efter opgave 3.

Der er direkte ud fra loggen i alt udpeget 288 usability-relaterede observationer. Disse observationer er herefter gennemgået for hver enkelt testperson med henblik på at identificere ens problemer – eksempelvis udtryk fra brugeren på forskellige linjer i loggen, der drejer sig om samme grundlæggende problem. Denne gruppering bragte antallet af observationer ned på 187. Der er her ikke taget højde for problemdubletter på tværs af testpersoner, da dette undersøges i et senere afsnit.

Bemærk, at vi her ikke gentager processen med identifikation af problemer i P01-P08, da denne allerede er udført i Hansen [7]. En grundig gennemgang af de forskellige logs har dog resulteret i mindre yderligere identificerede problemer hos et par testpersoner.

5.1.4 Problemklassifikation

Det næste skridt i processen efter identifikation af usabilityproblemerne er at klassificere dem, efter hvor alvorlige de er. Igen gør vi her brug af en nøgle, som er gengivet i Tabel 2 i Lindgaard og Chattratichart [18] (udvidet efter Molich [23]). Denne klassifikationsnøgle

kan ses i Tabel 10. Nøglen anvendes til klassifikation af samtlige identificerede observationer, uanset type og måden de er fundet på.

Klassifikation	Navn	Beskrivelse
C	Positiv oplevelse	Testpersonen udtrykker tilfredshed med en feature eller et designelement, som det derfor bør overvejes at fastholde.
A	God idé	Et forslag til forbedring af brugeroplevelsen, som er stillet af testpersonen undervejs i forløbet.
P	Mindre problem	Får testbrugeren til at tøve (afbryde opgaveløsningen) i et mindre antal sekunder. Herefter kan opgaveløsningen fortsættes.
Q	Alvorligt problem	Denne type problem forårsager en gang imellem situationer, hvor brugeren ikke er i stand til at udføre sin opgave, men kan også blot forsinke brugerens opgaveløsning i 1-5 minutter.
R	Kritisk problem	Denne type problem forårsager ofte situationer, hvor produktet ikke sætter brugeren i stand til at løse en rimelig opgave, eller hvor brugeren udtrykker kraftig irritation over produktets virkemåde.
T	Bug	Produktet opfører sig på en måde, som klart ikke er i overensstemmelse med den bagvedliggende specifikation. Heri er også inkluderet stavefejl, døde links, scripting-fejl og lignende. For NemIDs vedkommende kender vi ikke den præcise forventede opførsel af systemet, men almindelig sund fornuft vil dog kunne fange de fleste af denne type fejl.
N/A	Uden for kategori	Denne kategori indeholder problemer, der ikke umiddelbart finder indpas i ovenstående kategorier, eller som må karakteriseres som følgevirkninger eller -udtalelser af eller om tidligere oplevede problemer.

Tabel 10: Definitioner af problemklassifikationer (frit efter Tabel 2 i Lindgaard og Chattratichart [18]). Klassifikationen N/A er tilføjet.

Klassifikationen er udelukkende foretaget på baggrund af, hvordan brugeren har oplevet eller italesat problemet, eller hvordan testleder og/eller observatør har bemærket eventuelle forsinkelser eller fejl under opgaveløsningen. Der er ikke taget hensyn til, hvor stor en arbejdsbyrde det ville kræve at rette problemet, ligesom der heller ikke er indregnet en sandsynlighed for, hvor ofte det pågældende problem vil kunne forventes at optræde (som eksempelvis *criticality* beskrevet i Rubin og Chisnell [31], s261f). Sådanne faktorer har vi umiddelbart ingen information om, men i problemgrupperingen i afsnit 5.1.5 vil vi for hver identificeret gruppe angive, hvor mange problemer der findes af hver kategori i gruppen. Dette giver efter vores mening et mere sigende billede end at konstruere gæt på hyppigheden af problemgruppen.

Igen har Kasper Hansen foretaget klassifikation af P01-P08, og forfatteren har selv gjort det samme for P09. Udover brugen af den samme klassifikationsnøgle, er der for at sikre konsistens yderligere foretaget en inter-rater reliability test på de identificerede observationer i P02 og P03 (tilfældigt udvalgt). Vi har her selvstændigt klassificeret de for P02 og P03 i alt 39 identificerede observationer efter nøglen i Tabel 10 og efterfølgende

beregnet Cohens vægtede kappa for det resulterende skema, som kan ses i Tabel 11. I diagonalen er her angivet antallet af klassifikationer, som begge ratere er enige om. Tal udenfor diagonalen indikerer uenighed mellem de to ratere. Cohens vægtede kappa (oprindeligt fra Cohen [3]) er en værdi mellem 0 og 1, der angiver, hvor god overensstemmelse der er mellem to ratere: 0.0 er total uenighed, og 1.0 er total enighed. Cohens kappa er beregnet i R ved hjælp af `wkappa()`-funktionen fra `psych`-biblioteket.

	C	**A**	**P**	**Q**	**R**	**T**	**N/A**	**Sum**
C	8	-	-	-	-	-	-	8
A	-	-	-	-	-	-	-	-
P	-	-	11	-	-	-	-	11
Q	-	-	4	9	-	-	-	13
R	-	-	-	-	-	-	-	-
T	-	-	-	-	-	1	-	1
N/A	-	-	1	-	-	-	5	6
Sum	8	-	16	9	-	1	5	39

Tabel 11: Inter-rater reliability test af P02 og P03. Kasper Hansens klassifikation er i kolonner og forfatterens egen i rækker. Skraverede felter i diagonalen indikerer enighed mellem de to ratere.

Vi får $\kappa = 0.923$ (ikke-vægtet: $\kappa = 0.828$), hvilket indikerer meget god overensstemmelse mellem de to ratere.[26] Ud fra resultatet af denne stikprøvetest vælger vi ikke at klassificere samtlige observationer én gang til, men kan med ro i sindet genbruge Kaspers, idet der vil være god overensstemmelse mellem P01-P08 og P09. Samtlige observationer er dog selvfølgelig gennemgået for korrektur og fejl af forfatteren.

I Tabel 12 er vist de 187 identificerede observationer fordelt på kategorier og testpersoner, og i Tabel 13 er gengivet klassifikationen af de under CUT-seancerne identificerede problemer.

	P01	**P02**	**P03**	**P04**	**P05**	**P06**	**P07**	**P08**	**P09**	**Sum**
C	3	5	3	1	-	-	-	-	1	13
A	-	-	-	-	-	-	-	-	-	-
P	16	8	8	20	3	2	3	8	10	78
Q	2	7	2	18	6	-	6	16	11	68
R	-	-	-	-	1	1	-	7	1	10
T	1	-	1	-	-	2	-	-	2	6
N/A	1	4	1	-	1	-	-	2	3	12
Sum	23	24	15	39	11	5	9	33	28	187

Tabel 12: Fordelingen af usabilityproblemer på klassifikationer og testpersoner.

Der kan knyttes en del bemærkninger til Tabel 12 og 13:

[26] En teknisk detalje, der bør nævnes, er placeringen af kategorien N/A i skemaet. Cohens *vægtede* kappa-værdi medtager værdier uden for diagonalen i skemaet, og disse vægtes efter, hvor langt de befinder sig fra diagonalen. Her kan N/A ikke siges at være en 'værre' klassifikation end T (den vil sandsynligvis skulle placeres omkring A eller P), men den er alligevel medtaget i beregningen. En udelukkelse ville sandsynligvis kun forbedre kappa-værdien.

	C	A	P	Q	R	T	N/A	Sum
CUT3	1	-	11	7	-	-	-	19
CUT4	-	-	9	4	-	-	1	14
Sum	1	-	20	11	-	-	1	33

Tabel 13: Fordelingen af usabilityproblemer på klassifikationer og CUT-seancer.

- Testpersonerne har generelt ikke ytret sig særlig positivt om det testede produkt undervejs i testforløbet (13 positive ytringer). Disse ytringer er ligeledes primært fordelt på de tre første testpersoner. Det kan hertil bemærkes, at der er en forholdsvis god sammenhæng mellem de testpersoner, der har ytret sig positivt, og dem som ingen kritiske problemer (kategori R) har oplevet.
- Ingen testpersoner er kommet med forslag til forbedringer af systemet (kategori A).
- Der er fundet ganske få deciderede bugs (seks i kategori T, forbehold for dubletter).
- Der er identificeret hele 10 kritiske problemer i det færdige og lancerede produkt, ligesom antallet af alvorlige problemer (68) efter vores mening er ganske højt. Q-klassifikationen indeholder en del problemer, der er baseret på testpersonernes utilfredshed med aspekter af systemet – noget der dog bestemt ikke bør ses mindre seriøst på.
- De to CUT-seancer har givet anledning til 33 observationer, hvor 11 er klassificeret som alvorlige. Disse observationer er unikt identificeret i CUT-seancerne.

Vi kommer nærmere ind på ovenstående samt flere observationer i diskussionsafsnittet (afsnit 6). I det følgende afsnit vil vi gruppere de fundne usabilityproblemer i mere forklarende grupper.

5.1.5 Problemgruppering

Den sidste fase i analysen af resultaterne af THT'erne er problemgrupperingen. De foregående to afsnit har fokuseret på identifikation og klassifikation af problemerne – i dette afsnit vil vi inddele dem i overordnede og sub-ordnede grupper for at give et bedre overblik over, hvad problemerne omhandler samt få et bedre overblik på tværs af de enkelte testpersoner. På denne måde undgår vi også problemdubletter. Igen tager vi i stort omfang udgangspunkt i den allerede foretagne gruppering, som vi gennemgår og udvider med problemerne fra P09.

Den udførte problemgruppering, eller *matching*, er baseret på fremgangsmåden i Hornbæk og Frøkjær [9]. Til de identificerede usabilityproblemer er der knyttet en beskrivelse, klassifikation og kontekst for hvert problem. Det er af omfangs- og tidsmæssige årsager fravalgt at give hvert problem en overskrift samt en mulig løsning. Vi er ikke i dette projekt interesseret i at se på mulige forbedringsforslag til produktet – vi ønsker her blot at evaluere det. Det er listen med disse problembeskrivelser, der ligger til grund for den følgende gruppering, som er udført med *affinitetsdiagrammering*[27]. Her kan man eksempelvis udskrive listen med problemer, klippe hver enkelt ud og placere dem én for én i

[27] Se f.eks. Gerry Gaffneys udgave: http://www.infodesign.com.au/usabilityresources/affinitydiagramming

bunker efter ensartethed. Denne 'ensartethed' kan, som beskrevet i Hornbæk og Frøkjær [9], defineres på mange forskellige måder. Kasper Hansen har i problemgrupperingen for P01-P08 valgt teknikken *similar changes*. Denne teknik bygger på at identificere problemer, der ville kræve de samme ændringer i produktet for at imødekomme. Konkret identificeres disse ved at se på *identiske kommentarer* givet af brugerne og ligeså kommentarer logget af observatøren til fejl eller hændelser, der ikke kommenteres af brugeren:

> *"Identical comments: Two problem comments are identical if fixing one of the comments would most likely also fix the other and vice versa. The fix does not have to be in accordance with the solution suggested in either of the comments. If one comment is a generalization of another, they are not identical."*
>
> Molich og Dumas [24], s267

Similar changes er valgt som metode på grund af dens simple tilgang samt mulighed for at favne alle de fundne usabilityproblemer uden at blive tvunget til at se på løsningsforslag til dem. *Practical prioritization*, *The model of Lavery et al (1997)* og *User Action Framework* er af samme årsager fravalgt. Vi anvender derfor også teknikken til P09.

Efter en grundig gennemgang af grupperingerne vil vi revurdere navne og klassifikationer af de fundne grupper, ligesom vi tilføjer problemerne fra P09 til de eksisterende grupper. Denne proces har ikke givet anledning til omordning af eksisterende eller definition af nye grupper. De 18 fundne problemgrupper, som igen er overinddelt i 6 overordnede problemgrupper, kan ses i Tabel 14. Her er ligeledes angivet, hvilken overordnet klassifikation gruppen er tildelt, samt i hvilke kontekster de fundne problemer er opdaget (OPG eller CUT). Grunden, til at der kun er 185 problemer i stedet for de observerede 187, er, at en bestemt bug (kategori T) blev observeret på præcis samme sted og tidspunkt i forløbet af tre forskellige testpersoner – disse tre problemer blev derfor slået sammen. Gruppens overordnede klassifikation er baseret på et skøn fra forfatterens side, vurderet efter antallet af problemer i de forskellige kategorier, antal brugere der har observeret problemerne, etc. Ligesom for P01-P08 er der for P09 ikke opstået situationer, hvor et problem kunne indpasses i flere grupper.

Som det kan ses i Tabel 14, er 33 af de 185 problemer identificeret under de eksperimentelle CUT-seancer efter opgave 3 og 4. Mange af disse problemer er unikke i den forstand, at de ikke er observeret på andre tidspunkter i forløbet. Selvom langt de fleste af disse 33 problemer selvfølgelig er karakteriseret ved at være holdninger eller ytringer fra brugeren fremfor deciderede fejl eller uhensigtsmæssigheder, giver dette alligevel belæg for overhovedet at anvende CUT-metoden som supplement til den almindelige procedure for THT'er.

Vi vil nu kort beskrive hver problemkategori, og hvorfor den er blevet tildelt dens tilhørende overordnede klassifikation. Vi vil desuden tilknytte relevante skærmbilleder af den pågældende situation, hvor det er muligt.

Aktivering af NemID – Tilvalg af OCES-certifikat (R)

I denne gruppe er der store problemer for særligt den unge del af testpersonerne (gruppe

Problemområde og gruppe	Kat.	OPG	CUT3	CUT4	Deb	Sum
Aktivering af NemID						
Tilvalg af OCES-certifikat	R	11	0	0	1	12
Generelt	Q-R	13	0	0	2	15
Valg af bruger-id	Q-R	14	0	0	3	17
Første pålogning	Q	3	0	0	0	3
Valg af adgangskode	P	4	0	0	0	4
Login med NemID						
Det offentlige log-in-fællesskab	Q	10	0	0	0	10
Indtastning af bruger-id	P	3	0	0	1	4
Indtastning af nøgle fra nøglekort	P	9	0	0	1	10
Bestilling af nøglekort						
Lokalisering af webside og bestilling	Q	11	0	0	2	13
Konceptet						
Single-signon med NemID	Q-R	5	14	14	3	36
Samme kode til flere tjenester / Sikkerhed	Q	2	2	0	12	16
Nøglekortet	P	4	1	0	7	12
Information						
Velkomstbrevet	P-N/A	3	0	0	0	3
Andet						
Skat.dk	Q	7	0	0	0	7
Netbank	Q	6	0	0	0	6
Sundhed.dk	P	1	1	0	0	2
NemID generelt	P-C	2	1	0	9	12
Proces	N/A	0	0	0	3	3
Sum	-	108	18	14	44	**185**

Tabel 14: Identificerede problemområder med tilhørende klassifikation (kategori) og antal fordelt på opgaver og CUT-seancer. 'Deb' er den afsluttende debriefing.

A). Det er for dem umuligt at finde muligheden for at tilknytte et OCES-certifikat til deres NemID – et trin som af uransaglige årsager ikke var at finde i aktiveringsprocessen på netbanken. Da de endelig finder muligheden på selvbetjeningen på nemid.nu, forvirrer formuleringerne så meget, at de begge ikke selvstændigt kan gennemføre opgaven.

> *"Bortset fra at det gik galt, så gik det egentlig meget godt. Men det gik meget galt med det certifikat der."*
>
> P06, Opgave 2

> *"Testleder: Det, som man ikke lige ved, er, at du faktisk skulle ind og bestille NemID igen for at få adgang til offentlige tjenester. P09: Eeeej, hvor er det langt ude! [...] Det er virkelig, virkelig smart lavet det her system... Testleder: Det tager jeg som en ironisk bemærkning? P09: Ja, meget sarkastisk."*
>
> P09, Opgave 2, efter konstateret task failure

> *"Testleder: Hvad var det, der var svært i opgave to? P09: Der var ikke noget direkte 'HER kan du bestille til det offentlige', det var sådan 'Bestil ID' - det var meget kryptisk at få hjælp. Jeg sidder med det i hånden, jeg har det...altså...hvorfor skal jeg bestille det en gang til? Hvis ikke du havde fortalt mig det, det havde jeg aldrig fundet selv. Testleder: Så der havde du også givet op på et eller andet tidspunkt? P09: Ja."*
>
> P09, debriefing

Det er ligeledes heller ikke tydeligt beskrevet, om og hvordan dette tilsagn kan trækkes tilbage igen – noget som ligger meget på sinde for én af testpersonerne. Selve skærmbilledet er desuden meget højt og teksttungt, og der skal på mange skærme scrolles for at kunne se begge valgmuligheder samtidig.

Det er vurderet som kritisk, at flere testpersoner ikke selvstændigt kan gennemføre tilknytningen af et OCES-certifikat til deres NemID. Ligeledes er der ikke megen hjælp at få til dette område, formuleringerne i hjælpen er misvisende, og der opstår tvivl om konsekvenserne af det valg, man bedes træffe. Derfor opnår gruppen kategoriseringen R. Det har desværre ikke været muligt at skaffe et skærmbillede af situationen.

Aktivering af NemID – Generelt (Q-R)

Denne gruppe indeholder en del forskellige problemer, som dog ved brug af teknikken *similar changes* stadig kan tilskrives samme overordnede kategorisering. Blandt de alvorlige problemer er for eksempel manglende fejlmeddelelser om, at den midlertidige adgangskode er blevet anvendt for mange gange (kommer først efter adskillige forsøg), samt flere tvivl om den security warning, som dukker op, når DanIDs Java-applet skal godkendes (i et enkelt tilfælde ville testpersonen have 'krydset' (annulleret) godkendelsen og dermed ikke kunnet være kommet videre i forløbet.)

"Den vil jo godt have, at jeg kører noget Java... så jeg siger "Run". Men spørgsmålet er, om jeg overhovedet har lov til det på en arbejdsmaskine? Lad os se, hvad der sker."

P02, Opgave 2

"Jeg siger jo nok bare 'Run' jo..."

P03, Opgave 2

"Nogle gange læser jeg ikke engang, hvad der står...men det er ikke godt."

P05, Opgave 2

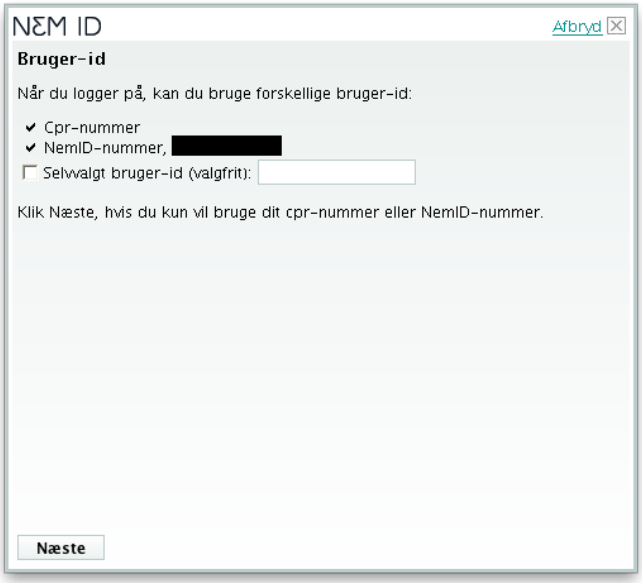
"Her er jeg ikke klar over, hvad jeg skal. Skal jeg trykke der? Testleder: Har du set sådan en før? P08: Nej. Det har jeg ikke. Men jeg skal derop og trykke, ikke? [krydset øverst til højre i vinduet, red.]"

P08, Opgave 2

Disse security warnings er ikke beskrevet i velkomstbrevet eller den umiddelbart tilgængelige dokumentation. Der er ligeledes problemer for flere brugere med at få startet selve aktiveringsprocessen – flere brugere forsøger at logge direkte ind med NemID i stedet. For en enkelt bruger lykkedes dette af uvisse årsager. En anden bruger præsenteres ved login på netbanken slet ikke for aktiveringsprocessen og må selvstændigt opsøge den.

Klassifikationen Q-R er givet, da flere brugere oplever kritiske eller alvorlige problemer i denne gruppe, ligesom disse problemer skaber stor forvirring og flere task failures – aktiveringen kan ikke gennemføres og opkald til support med efterfølgende re-aktivering er nødvendig.

Aktivering af NemID – Valg af bruger-id (Q-R)
Denne kategori drejer sig primært om følgende skærmbillede fra aktiveringsprocessen:



Brugerne er her meget i tvivl om betydningen af det valg, de bliver bedt om at træffe. Nogle brugere ser tekstfeltet som et sted, hvor de kan indtaste den af de ovenstående to muligheder, de vil bruge. Andre tror, at ved at vælge et selvvalgt bruger-id mister de muligheden for at bruge CPR-nummer eller NemID-nummer til login. En bruger ved ikke, hvor man skal finde det pågældende bruger-id henne, og to brugere kan ikke afkode, hvad der ligger i begrebet 'bruger-id' (modsat en formulering som f.eks. 'brugernavn'). Flere brugere kan ikke se nødvendigheden af at oprette et selvvalgt bruger-id, da f.eks. CPR-nummer stadig vil kunne anvendes. Det giver med andre ord ikke øget sikkerhed men blot endnu en 'kode' at skulle huske på. Desuden er brugernavnet ikke skjult, mens det indtastes.

> *"Jeg er tvivl om, om der skal flueben i 'Selvvalgt bruger-id'. Jeg forstår ikke, hvad dialogen gerne vil have mig til. Det ser ud som om, den har mit CPR-nummer, den har det her NemID-nummer, som jeg har fra brevet, og hvis jeg klikker 'Næste', så tror jeg, at så vil den bede om mit nøglekortsnummer."*
>
> P02, Opgave 2

> *"Det var vagt formuleret, om man kunne bruge det ene eller alle tre og fandt så ud af senere, simpelthen ved bare at teste det, at man kunne bruge alle tre typer logon. Og det synes jeg egentlig også er lidt overdrevet eller lidt overkill, at man har tre forskellige bruger-id, når man nu har valgt et selvstændigt bruger-id. Hvorfor skal jeg så have muligheden for at logge ind med mit CPR-nummer eller mit NemID-nummer, som står skrevet her, som alle kan gå ind og læse. [...] Altså jeg kan ikke se pointen i at oprette et bruger-id, hvis man kan bruge alle de andre. Det synes jeg, er åndssvagt."*
>
> P04, debriefing

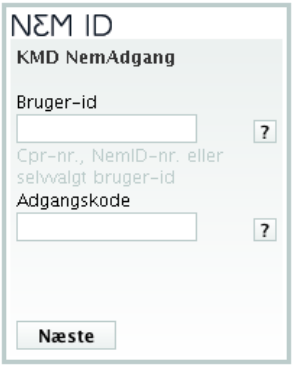
> *"Så jeg satsede på at lave et bruger-id, der var så langt og besværligt, som jeg kunne huske for at logge på, og så var jeg bare lidt overrasket over, at det ikke var krypteret. Fordi jeg blottede det jo fuldstændig overfor Gud og hvermand. Testleder: Altså at det ikke stod med prikker? P04: Ja, præcis."*
>
> P04, debriefing

> *"Skal jeg bruge mit CPR-nummer her? Det står der deroppe. [...] Selvvalgt bruger-id...valgfrit. Der kan jeg bruge mit CPR-nummer, ikke? Jeg kan også bruge det samme som før måske? For at huske det."*
>
> P08, Opgave 2

Der er altså overordnet store problemer med forståelsen af begreberne og formuleringerne i dialogen i sig selv for et stort antal brugere. Klassifikationen havner derfor på Q-R.

Aktivering af NemID – Første pålogning (Q)

Denne gruppe består af problemer under første brug af NemID efter aktiveringen. To brugere er i tvivl om, hvad der skal stå i feltet 'Bruger-id', og hjælpeteksten er for en enkelt brugers vedkommende uhensigtsmæssigt placeret under feltet 'Adgangskode' i stedet for under feltet 'Bruger-id'. Der opstår også tvivl om, hvorvidt bindestregerne i NemID-nummeret skal medtages, når dette anvendes som bruger-id.

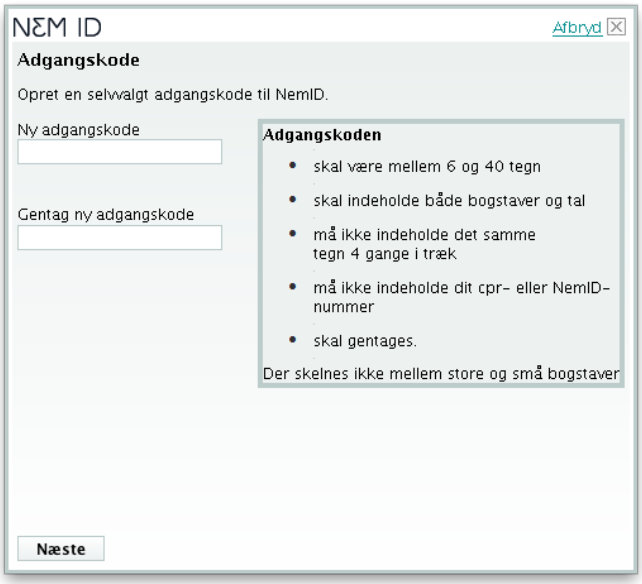


> *"Det er jo lidt forvirrende det her, når der står 'Log på med NemID-nummer', så tror jeg, at det er her, de vil, og så vil de have mit CPR-nummer eller det nummer der [NemID-nummer, red.]. Så det er som om, at man kan vælge, men alligevel kan man ikke. Det er sådan, jeg føler nu."*
>
> P05, Opgave 2

Gruppen er udelukkende på baggrund af de enkelte problemers klassifikation tildelt klassifikationen Q.

Aktivering af NemID – Valg af adgangskode (P)

Denne kategori drejer sig primært om følgende skærmbillede fra aktiveringsprocessen:



Denne dialog har ikke voldt de store problemer for de fleste brugere. Nogle få er dog en smule forvirrede over, om de bliver bedt om at lave en ny adgangskode eller indtaste en tidligere, f.eks. fra netbanken (hvilket også kan lade sig gøre). En enkelt bruger overser den interaktive hjælp, som afkrydser kravene, så snart de opfyldes, og en anden bruger kommer først igennem efter to forsøg. Dialogen fortæller heller ikke, hvorvidt adgangskoden kan ændres igen efterfølgende.

> *"Nu håber jeg det må være den [adgangskode, red.], som jeg brugte før."*
>
> P01, Opgave 2

> *"Det er vel en form for login-kode til netbanken. Men jeg ved det ikke. Jeg ved ikke, om det er en form for login-password?"*
>
> P02, Opgave 2

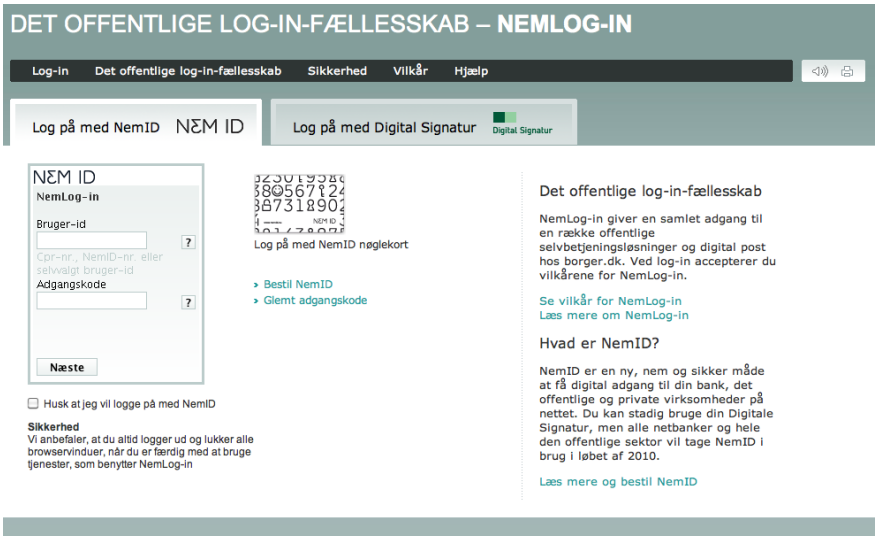
> *"Nu skal jeg lave en adgangskode. Kan man ændre den eller hvad? Nej, det kan man så ikke..."*
>
> P03, Opgave 2

> *"Ny adgangskode...er det den, jeg har brugt i netbanken? Testleder: Hvad ville du tro? P08: Jamen det ville jeg tro selv. At jeg kan bruge den, som jeg har brugt til at komme ind i netbanken."*
>
> P08, Opgave 2

Da de fleste problemer i gruppen er af mindre karakter, og kun en enkelt bruger har haft større problemer, klassificerer vi gruppen til P.

Login med NemID – Det offentlige login-fællesskab (Q)

Denne gruppe drejer sig om det skærmbillede, som brugeren bliver præsenteret for, når vedkommende skal logge ind på en offentlig tjeneste:



Bemærk, at der på ovenstående skærmbillede er klikket på NemID-fanebladet. Her er der en enkelt bruger, der fuldstændig overser, at der er to forskellige faneblade og efterfølgende giver op. Andre brugere overser det indledningsvist og leder lidt efter det, inden de indser, at der er tale om en fanebladsstruktur. Flere brugere er ligeledes irriterede over, at det er fanen med digital signatur, der kommer frem som standard, og enkelte brugere opdager heller ikke muligheden for, at systemet selv kan huske den foretrukne loginmetode (checkboksen under NemID-appletten):

> *"Jeg tænker, når de nu kører så storstilet, hvor mange bruger digital signatur fremfor NemID, og hvad der så skulle ligge forrest."*
>
> P01, Opgave 4

To brugere er varsomme overfor denne mulighed, idet det frygtes, at siden så også 'husker' både brugernavn og adgangskode:

> *"Dem der, som man klikker af, er jo normalt sådan nogle 'husk min adgangskode', og den ville jeg jo ikke klikke af. Det ville også være rimelig dårlig sikkerhed, hvis man kunne det."*
>
> P01, Opgave 4

Flere brugere kan ikke huske at have set siden før, da de kommer tilbage til den for anden gang:

> *"Observatør: Opdagede du på noget tidspunkt nogen tegn på, at SKAT og sundhed.dk 'hang sammen'? P06: Ja altså, det var jo samme system til at logge ind med, samme koder og sådan. Det må vel være den eneste sammenhæng sådan lige. Observatør: Var det den samme hjemmeside, du loggede ind på også? P06: Ja, det...nej det var det ikke...[...] Jo, det var vel egentlig den samme, var det ikke...? Jo, det... Det var ikke noget, jeg tænkte meget over - det kunne det meget vel have været. Observatør: Jeg kan afsløre, at det var faktisk den samme. P06: Jeg kunne forestille mig, at det var den der nemid.nu."*
>
> P06, debriefing

Gruppen har grundet flere brugeres oplevelser af forsinkelse og utilfredshed med virkemåden fået klassifikationen Q.

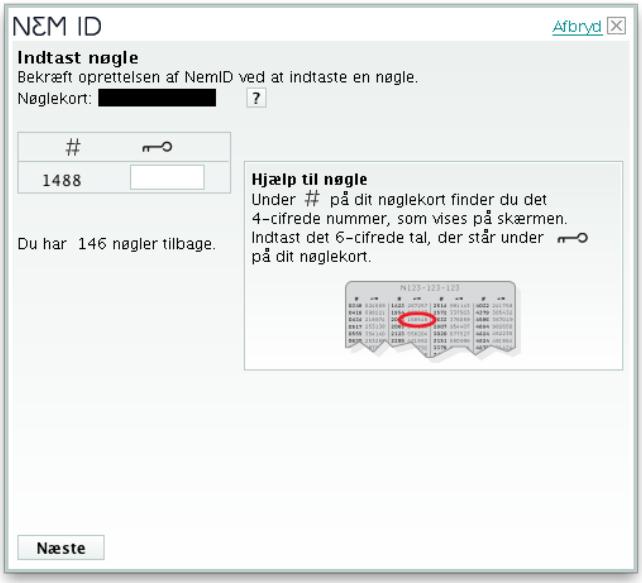
Login med NemID – Indtastning af bruger-id (P)
Denne gruppe består mestendels af undren fra brugernes side over, hvorfor bruger-id'et ikke skrives skjult, altså med prikker. Dette ses som en sikkerhedsbrist i systemet, men giver ikke anledning til yderligere forsinkelser eller kommentarer:

> *"Jeg er lidt overrasket over, at sådan noget (CPR-nummer, red.) ikke er 'krypteret'."*
>
> P04, Opgave 4

En enkelt bruger forstår ikke konceptet med et bruger-id. Her havde betegnelsen "brugernavn" måske afhjulpet problemet.

Gruppen er klassificeret til P.

Login med NemID – Indtastning af nøgle fra nøglekort (P)
Denne gruppe indeholder både indtastning af nøgle under aktiveringen samt under den efterfølgende brug af det aktiverede NemID. Vi gengiver her skærmbilledet af dialogen fra aktiveringen:



En enkelt bruger kan ikke gennemskue konceptet med at finde den rigtige nøgle på kortet – forventer at det er den første nøgle, der skal indtastes første gang og så fremdeles. Opgiver til sidst og ville ringe til support:

> *"Hvor skal jeg så starte? Er det ligegyldigt, hvor jeg starter, og hvor jeg ender? Næh...jeg skal starte fra oven af. Og så er der hjælp til nøgler. Skal jeg indtaste den, der står her? Der står ikke de samme tal."*
>
> P08, Opgave 2

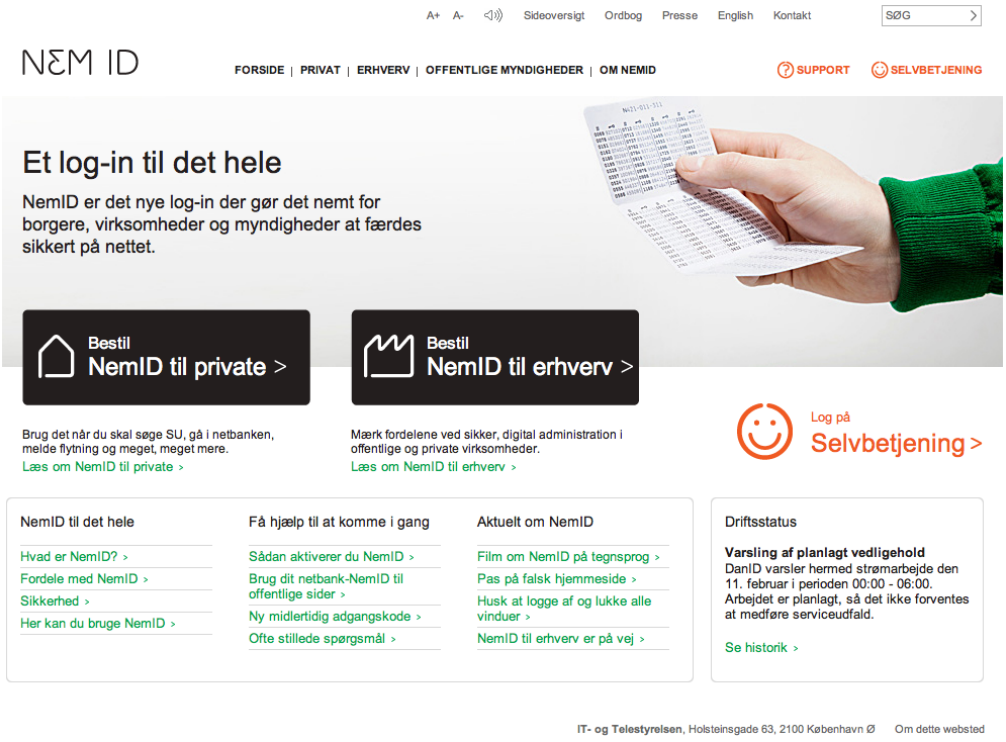
Ligeledes er tooltippet (den lille forklarende tekst, som fremkommer, når musen føres over) på Næste-knappen sat til "Log på med netbank", hvilket må siges at være særdeles uhensigtsmæssigt under aktiveringsprocessen:

> *"Nå, så står der, jeg skal logge på med netbank. Så står der ikke noget om, hvor jeg skal logge på henne. Testleder: Her er du i tvivl? P08: Ja, der er jo ingen steder, jeg kan logge på."*
>
> P08, Opgave 2

En anden bruger udtrykker undren over, at den indtastede nøgle ikke er skjult med prikker, ligesom han er irriteret over at skulle finde nøglekortet frem hele tiden.

Gruppen er klassificeret P, da kun en enkelt bruger har oplevet alvorlige eller kritiske problemer med indtastning af nøgler fra nøglekortet.

Bestilling af nøglekort – Lokalisering af webside og bestilling (Q)
Hjemmesiden nemid.nu, hvor brugeren kan logge ind på selvbetjeningen og blandt andet bestille ekstra nøglekort, ser således ud:



Et stort antal brugere har problemer med at finde ud af, på hvilken hjemmeside bestilling af et ekstra nøglekort kan foretages. Nogle Googler med ret forskelligt resultat forskellige udgaver af 'NemID', nogle prøver danid.dk og endda nemid.dk,[28] før de ad omveje eller ved at finde adressen på nøglekortet finder hen til nemid.nu:

[28] Denne adresse har intet med NemID at gøre. Der var på et tidspunkt stor furore i medierne, da det kom frem, at flere brugere havde forsøgt at logge på denne side med deres NemID. Det viser sig, at firmaet Assemble (som driver sitet nemid.dk) også har en løsning ved navn NemID, som blev introduceret i 2006. DanID har altså ikke været i stand til at få dette domæne inden lanceringen.

> *"Jamen det er NemIDs side. [nemid.info, red.] [...] Men jeg tænker, det er mere en informationsside, end et sted, hvor man kan gøre så meget. Overskrifterne er meget sådan 'Prøv det her'. Virker som om, det er et sted til nogen, som ikke har NemID endnu, men som gerne vil have det. [...] Jeg så et sted, hvor der stod, at man kunne bestille dem, men jeg kan ikke huske hvor."*
>
> P01, Opgave 5

> *"Mit første gæt er, at jeg skal jo nok på DanIDs hjemmeside så. Uden at vide, om det er rigtigt. [...] Jeg tager et kig på bagsiden af nøglekortet, for jeg tænker, det måske kan stå der. [efter at have besøgt danid.dk, red.]"*
>
> P02, Opgave 5

> *"Men de har simpelthen ikke fået domænet nemid.dk? [...] Det er jo helt til hest, altså...! Så kunne man have fundet på et andet navn i stedet for, altså."*
>
> P03, debriefing, kommentar fra manden ved siden af

Da brugerne endelig finder ind på sitet, overser de fleste det forholdsvis store orange link med smiley'en, der findes midt på siden. Mange leder dog efter "et sted at logge ind", som åbenbart burde findes et andet sted på siden. Brugerne er dog ret sikre på, at de er havnet det rigtige sted, og de fleste anvender sidens søgefunktionalitet til at finde bestillingen af ekstra nøglekort.

> *"Nu er jeg kommet ind på nemid.dk's forside. Og der vil jeg tro, at jeg skulle kunne bestille et nyt nøglekort. Men det kan godt være, jeg ikke kan det! Det kan selvfølgelig også være, man skal gøre det over banken?"*
>
> P03, Opgave 5

En enkelt bruger går dog i gang med at bestille NemID påny, idet dette forveksles med et ekstra nøglekort – brugeren stoppes dog af testlederen, inden bestillingen gennemføres grundet uvisse konsekvenser af dette forløb.

Flere brugere udtrykker også tvivl om konceptet med automatisk at modtage nye nøglekort:

> *"Hvad så, hvis jeg allerede var færdig med alle mine nøglekort?"*
>
> P07, Opgave 5

Grundet den store mængde brugere, der brugte forholdsvis lang tid på at finde den korrekte side, samt at denne side åbenbart ikke giver et godt nok overblik, klassificeres gruppen til Q.

Konceptet – Single-signon med NemID (Q-R)

Denne gruppe indeholder primært problemer identificeret under CUT3, CUT4 og debriefingen og drejer sig om den forventede opførsel af single-signon konceptet samt brugernes holdninger hertil. Her har CUT-seancerne vist sig at være særdeles givtige, og det er også her, at langt de fleste problemer identificeret ved denne metode befinder sig. De fleste brugere havde ikke forventet, at de automatisk var logget på en helt anden side, efter at være logget på SKATs selvbetjening. Ligeledes gav samtlige brugere (som vel at mærke

nåede videre end opgave 2) udtryk for usikkerhed ved dette koncept og dets virkemåde. Til denne gruppe er der et meget stort antal relevante citater – vi gengiver her nogle af dem og henviser så til Bilag H for resten:

> *"Jeg havde klart forventet, at jeg skulle logge mig på igen. Det kan godt være, det også er det offentlige, men det er trods alt et 'andet sted'."*
>
> P01, CUT3

> *"Den må jo umiddelbart kunne registrere, at jeg ikke har logget mig af. Altså at jeg har ikke har lukket det andet vindue, jeg har åbent. Maskinen har vel et id-nummer, som den kan læse og sige: 'Hov, den bruger er logget på'. Fordi hvor skulle de ellers vide det fra? [...] Jeg synes egentlig, det er et dårligt vindue. Det giver én det indtryk, at man er logget på, når man ikke er logget på. (sundhed.dk, red.) At det er et andet vindue, der er årsagen til, at du kan være logget på. Hvorfor skal man ikke logge på hver sit vindue med hver sin kode? Det har jeg lidt svært ved at forstå. [...] Man bliver sådan lidt i vildrede, når man lige pludselig ikke har adgang til siden, du havde adgang til, fordi jeg har jo ikke logget af fysisk på den side. Jeg forventede, at jeg stadig var logget på sund, efter jeg havde logget af SKAT."*
>
> P02, CUT3, Opgave 4 og CUT4

> *"Det må jo være fordi at...pas! Noget med...noget med at det er den her computer, jeg sidder på eller et eller andet...?"*
>
> P03, CUT3

> *"Okay...jeg er så åbenbart allerede logget på... Jeg vil gerne proklamere min utilfredshed allerede. [...] Jeg ville hellere have haft en selvstændig løsning til hver hjemmeside. Og det var faktisk også det, jeg troede, det var. Jeg var meget overrasket over, at det blev centralt styret. Jeg var faktisk også lidt overrasket over, at jeg stadig var logget ind, selvom jeg gik væk fra siden. [...] Jeg kunne måske godt have tænkt mig en løsning, hvor man blev logget ud fra siden, når man gik væk fra den. Det går jeg ud fra, at man godt kan lave, hvis man vil."*
>
> P04, Opgave 3, debriefing

> *"Den gjorde det af sig selv. Jeg ved ikke, hvordan den gjorde det, men det gjorde den. Jeg trykkede bare på 'Log på', så pludselig var jeg logget på. Det må vel være noget med, at jeg var logget ind et andet sted, så kunne den ligesom have ramt at det var samme bruger, der var logget på. Noget i den retning, kunne jeg forestille mig. [...] Testleder: Hvordan med netbank, ville du være logget på den også? P06: Det kunne jeg forestille mig, ja. Det ville virke naturligt. [...] Observatør: Så du ville tro, at NemID-selvbetjeningen også var med i det der, at når du loggede ind på den, så havde du også adgang til de andre? P06: Det ville da virke logisk."*
>
> P06, Opgave 3, debriefing

> *"Jeg har oplevet, at min sikkerhed er meget i fare, kan jeg se. Fordi når jeg logger på ét sted, så er jeg automatisk logget på et andet sted. Og det synes jeg ikke om. Nummer et. Nummer to, det her betyder, fordi jeg har ikke brugt*

> *mit nøglekort på min computer. Det betyder, at jeg kan logge ind på min netbank fra hvilket som helst internet. Og det synes jeg heller ikke om! Jeg kan godt forstå, at det er meget nemmere nu, at have adgang til tingene, men...altså...jeg synes ikke, at man skal have samme facilitet til alle tingene. Altså man skal lige kunne skelne mellem, hvad der er for privat og som skal være besværligt at logge ind på, på f.eks. netbank, end hvad der er på f.eks. sundhed.dk's hjemmeside. Men jeg har lige så nem adgang til begge to."*
>
> P07, CUT3

> *"Så ryger jeg direkte ind. Så står der, hvem jeg er, og så videre. Det er jo egentlig dumt, fordi hvis jeg glemmer at logge ud, er der nogen folk, der kan logge på som mig."*
>
> P09, Opgave 3

Enkelte brugere kom dog også med positive ytringer omkring single-signon:

> *"Ah hvor smart! Lige som jeg havde forventet, så da jeg loggede ud af SKAT, så loggede den mig også ud af sundhed.dk. Det vil sige, at jeg skal faktisk logge på igen. Så jeg nåede faktisk ikke til den der side, jeg blev faktisk smidt over til det offentlige login-fællesskab igen, hvor at jeg skal logge på pøny."*
>
> P04, Opgave 4

> *"Nej, jeg havde ikke tænkt på det på forhånd, men det kommer nu heller ikke som nogen stor overraskelse, vil jeg sige. Men det er da meget godt - det er smart, synes jeg."*
>
> P06, CUT3

Grundet det store antal brugeres utryghed og manglende forståelse for konceptet og accept af virkemåden, klassificeres gruppen som Q-R. Her er taget in mente, at NemID i sig selv ikke en formel del af det offentliges single-signon faciliteter, men det vil uvægerligt blive knyttet sammen med det og er i markedsføringen også en central del af det offentliges IT-selvbetjening. Havde NemID været en direkte del af single-signon, havde vi hævet klassifikationen til R.

Konceptet – Samme kode til flere tjenester / Sikkerhed (Q)

Denne gruppe er meget lig den foregående, dog er der her ikke fokus på selve single-signon i den forstand, at man kan være logget på flere steder samtidig, men derimod konceptet med at have ét login til det hele – et af NemIDs største slogans i markedsføringen. Ikke overraskende er der her blandede fornemmelser fra brugerne, som dels er bekymrede over "at putte alle sine æg i én kurv" og at lade et privat firma varetage adgangen til ens personfølsomme oplysninger.[29] En del brugere udtrykker dog også deres tilfredshed med (hele eller dele af) konceptet:

> *"Jeg havde egentlig ikke tænkt så meget over, at jeg også kunne bruge det til andre steder, men jeg bruger jo også SKATs hjemmeside et par gange om året eller sådan noget, og der er det i virkeligheden igen det samme. Der har man*

[29] Noget, som dog også var tilfældet med den gamle digitale signatur. Her var både TDC og DanID inde over, i to omgange.

> *den der tastselv-kode, og hvad var det nu den var, og så skal man bestille en ny igen, og der er det jo sådan set meget nemmere, at man har den samme kode, man kan huske. [...] Ellers skulle man ind på den der nemid.nu og se, hvad for et fællesskab, der er. Der kunne man jo sikkert få at vide, hvor præcist man kan bruge det. [...] Jeg er helt sikkert mere nysgerrig nu for at se, om der er nogle andre steder, hvor jeg kan have fordel af at bruge det."*
>
> P01, debriefing

I forhold til sikkerheden er der også blandede følelser:

> *"Selve konceptet [med ét bruger-id, adgangskode, etc., red.] er rigtig godt, fordi netop personfølsomme oplysninger. Hackere som aflæser tastekoder, kender ikke ens nummer. Så på den måde synes jeg, det forøger sikkerheden i forhold til, hvad jeg plejer at bruge."*
>
> P02, debriefing

> *"Negativt overrasket. Fordi at jeg føler lidt, at mit logon hos SKAT går ind og styrer mit login hos mit sundhed. To ting, der ikke har noget med hinanden at gøre, går ind og påvirker hinanden. Så derfor føler jeg lidt, at mine digitale æg er samlet i samme kurv. Og det er jeg ikke tilfreds med. Det føler jeg ikke, er sikkert. Testleder: Hvorfor ikke? P04: Man kan sige, hvis det bliver kompromitteret, f.eks. hvis jeg taber pungen, så går alle mine æg i stykker. Så kan folk jo komme ind på alle mine sider, og som vi snakkede om før, så hvis jeg nu bare går tilbage til min forside og stadig er logget på, så kan folk faktisk gå ind lige nu og logge sig på alle mine offentlige sider, går jeg ud fra. Jeg ved ikke, om det virker på netbank, men det kunne jeg faktisk godt tænke mig at få at vide. Det føler jeg mig ikke tryg ved."*
>
> P04, CUT4

> *"Ideen er god nok. Jeg ved ikke, hvor sikkert det, når det kommer til stykket, fordi at det er jo både og. Egentlig har du en mur, men når først den er blevet væltet, så har du slet ikke nogen. Så du er fuldstændig sårbar overfor tingene. Testleder: Så du er lidt bekymret for sikkerheden? P09: Ja, men det er også bare... al min paranoia, når man er på nettet."*
>
> P09, debriefing

Selvom en del brugere udtrykker sig positivt om konceptet med ét login til flere steder, er der dog samtidig så stor usikkerhed om sikkerheden i konceptet, at vi klassificerer gruppen som Q.

Konceptet – Nøglekortet (P)

Denne gruppe indeholder mestendels ytringer fra brugerne om selve nøglekortet – ikke dets brug, som er dækket i en tidligere kategori. Der er igen mange delte holdninger til kortet fra mange forskellige brugere. Nogle er bekymrede for de mange tal og disses størrelse på kortet, andre er bekymrede for at miste det, og nogle er positivt indstillede overfor den øgede sikkerhed, som kortet giver. Vi gengiver her nogle af de mest interessante citater. Først det positive:

> *"Man kan sige, NemID forøger sikkerheden, på grund af at du har en ny adgangskode hver gang. Den er meget lig den digitale signatur."*
>
> P02, debriefing

Og så de mere negativt ladede ytringer:

> *"Jeg synes, det [kodegeneratorer, red.] er lidt nemmere end at skulle sidde og finde de der tal..."*
>
> P01, efter Opgave 5

> *"Jesus Christ, det virker fuldstændig oldnordisk, at jeg skal rende rundt med det her kort!"*
>
> P04, Opgave 2

> *"Jeg kommer sikkert også til at smide det væk, glemme det og...jeg er nødt til at skulle have det på mig hele tiden. Man kan så sige, at det var jeg også nødt til med mit ActiveCard, men jeg synes, jeg kunne styre det bedre selv ved at have egen adgang til Danske Bank og til Nordea og så have mine offentlige instanser med nogle tastselv-koder og hvad ved jeg."*
>
> P04, debriefing

> *"Det er ikke meget store tal, der står. Hvis man nu var gammel eller blind, så var det jo nok ikke så smart."*
>
> P06, Opgave 2

En bruger syntes ligefrem, det var skræmmende med alle de tal:

> *"Det er lidt skræmmende at kigge på det her, synes jeg! Man har alle de der tal."*
>
> P05, Opgave 2

En enkelt bruger bliver forvirret over de "dekorationstegn", som er på bagsiden af nøglekortet:

> *"Alle de tegn der, hvad betyder de? Jeg er godt klar over, at det er nøgler, og alle de forskellige ting, der står hernede, de der koder og...men det er nøgler, og alt det der, jeg tænker på, hvad betyder det? Testleder: Hvad tror du? P08: Det er nogle tegn fra brevene, det er jeg godt klar over, det er. Ja, og så en smiley, der er på - det er det, jeg ikke er klar over."*
>
> P08, Opgave 1

Da ingen af de identificerede problemer har voldt brugerne større vanskeligheder, må de klassificeres som mindre betydende – P.

Information – Velkomstbrevet (P-N/A)

Velkomstbrevet er blevet meget lidt kommenteret af brugerne, selvom det i sig selv udgjorde den ene af de fem opgaver. En enkelt bruger har ikke modtaget sin midlertidige adgangskode sammen med brevet, og en anden bruger udtaler under gennemlæsningen, at:

> *"Det virker ikke umiddelbart nemt, vil jeg sige. Men det kan være, det er det, når man starter."*
>
> P01, Opgave 1

De fleste brugere læser forholdsvis hurtigt brevet igennem og springer reglerne over, og en enkelt har ligefrem læst brevet på forhånd. Dette kan ses mere tydeligt i Tabel 8 på side 58. Brevet bliver kun en enkelt gang brugt som reference senere i forløbet, da brugeren søger information om bestilling af ekstra nøglekort.

Gruppen klassificeres som P-N/A – primært på grund af den manglende modtagne kode samt den indledende nervøsitet omkring kompleksiteten.

Andet – Skat.dk (Q)
Denne gruppe har ikke direkte tilknytning til NemID, men er medtaget, da det efter vores mening er relevant også at se på synligheden af NemID på forskellige hjemmesider. Kan brugerne finde systemet på en hjemmeside, der tilbyder login med NemID? Primært bygger problemerne på, at brugerne har svært ved at lokalisere muligheden for at logge på. Logoet for NemID vises meget småt ved siden af teksten 'Log på med digital signatur' – noget som af flere brugere ikke forbindes umiddelbart med NemID. En enkelt bruger ville helt have givet op, efter at have forsøgt flere muligheder på siden, og en anden bruger udtrykker mild irritation over, at NemID-logoet er forskudt en smule i forhold til resten af teksten.

Desuden viste der sig hos flere brugere en fejl, da de forsøgte at logge ud fra SKAT. Brugerne selv bemærkede den knap nok, men den kan forårsage usikkerhed om, hvorvidt brugeren er logget ud eller ej – igen noget som har fået en del opmærksomhed i medierne og kan være et potentielt sikkerhedshul. Det viste sig dog i alle tilfælde, at brugeren *var* logget ud.

Gruppen er klassificeret Q grundet de mange brugere, der brugte forholdsvist lang tid på at finde login-muligheden for NemID.

Andet – Netbank (Q)
Denne gruppe består mestendel af nogle identificerede bugs, der dog ikke gav anledning til større frustration hos brugerne, samt en frustration fra en enkelt bruger over ikke at kunne tilmelde sit NemID til offentlige tjenester på netbanken. Noget som de resterende brugere ellers er blevet tilbudt undervejs i processen. Det samme scenarie udspillede sig hos den anden testperson i samme målgruppe, men her var frustrationen knap så tydelig.

Gruppen er klassificeret til Q. Dette bygger primært på de bugs og manglende præsentation af valg undervejs i processen, som blev identificeret. Det skal dog bemærkes, at de forskellige netbanker implementerer aktiveringsprocessen meget forskelligt, hvilket i sig selv kan være en hæmsko for produktet samlet set. Derfor er det ikke sikkert, at disse fejl vil kunne genskabes af andre brugere. Fejlene må desuden menes at være relateret til midlertidige performance- eller konfigurationsfejl og bør altså ikke være generelle.

Andet – Sundhed.dk (P)
For sundhed.dk's vedkommende er der meget lidt at komme efter. Kun en enkelt bruger

bruger lidt over et minut på at lokalisere loginknappen, selvom dennes placering er beskrevet i opgaveteksten. Gruppen får derfor klassifikationen P, selvom de to identificerede problemer begge er af typen Q.

Andet – NemID generelt (P/C)
Denne gruppe indeholder de ytringer og holdninger fra brugerne til NemID generelt, som ikke kunne indpasses i de andre grupper. Her finder vi en god blanding af positive og negative kommentarer:

> *"Når man først har installeret det, så er det sådan set nemt nok at bruge. Og på en måde er det måske meget rart, at man bruger den samme kode til flere steder. F.eks. kan jeg aldrig huske min kode til den der email-konsultation, fordi jeg bruger den simpelthen så sjældent, så de skal altid sende en ny, hvis jeg skal sende en mail til dem. Så hvis jeg kan bruge den her, så er det jo meget nemmere. Så det er jo sådan set en fordel."*
>
> P01, debriefing

> *"Jeg tror generelt, at det er sådan noget, du skal prøve nogle flere gange, før du begynder at blive fortrolig med det."*
>
> P02, debriefing

> *"Jeg synes egentlig, det har været okay. Altså det var jo rimelig simpelt at gå ind og få sit NemID jo. Der blev du bare ledt igennem. Det der med at bestille et ekstra nøglekort, der må jeg så sige, der var det lidt mere besværligt. Ellers synes jeg egentlig ikke, at det var sådan, at der var nogle store problemer."*
>
> P03, debriefing

> *"Jeg kan bare godt forestille mig, at jeg får smidt det der [nøglekort, red.] væk. Det må jeg indrømme. Og hvis man nu ikke ser så godt - jeg er f.eks. nærsynet - man kan jo ikke læse det, altså. Der er mange koder, man skal skrive hele tiden, og man skal sidde og lede hele tiden, så praktisk det er det da godt nok ikke. Det kan godt være, det er sikkert, det skal jeg ikke kunne udtale mig om. Det er ikke smart, nej. Men hvis jeg skal logge ind på nogle af de der ting, så vil jeg jo bruge det. Der er jo ikke rigtig andre muligheder. Der er simpelthen for mange små tal, man skal kunne. Så siger man, at man kan bruge det på mange forskellige computere, ja ja, men så skal man slæbe rundt med det hver gang, det gør man jo alligevel ikke. Så ved jeg, jeg ville smide det væk i hvert fald, så det vil jeg prøve at undgå. [...] Det kommer nok lidt an på, hvordan man definerer 'de fleste personer'. For jeg tror, at i min aldersgruppe vil der nok ikke være så mange problemer, men når man bliver lidt ældre, kunne jeg forestille mig, at der er mange, der vil have lidt svært ved det. Specielt det med de små tal. [...] Når først man har forstået det, så er det jo brugervenligt som sådan."*
>
> P07, debriefing

Desuden er der for en enkelt brugers vedkommende en del irritation over uinteressante introduktionsvideoer og en generel mangel på "en lige vej" hen til det, man ønsker at opnå.

Ingen af kommentarerne givet af brugerne om NemID *generelt* har givet anledning til at klassificere denne gruppe højere end P.

Andet – Proces (N/A) Med processen menes her selve forløbet, hvor opgaverne er blevet løst, og aktiveringen er blevet gennemført. Flere brugere udtalte, at de sandsynligvis ikke var kommet lige så langt uden vores 'hjælp', ligesom processen også har været en øjenåbner i forhold til at anvende NemID til andre tjenester end netbanken.

Da disse kommentarer ikke direkte vedrører produktet eller er af specielt positiv karakter, er gruppen klassificeret som N/A.

5.1.6 Sammenfatning

Vi har i de foregående afsnit analyseret de fra THT'erne indsamlede resultater. Der blev i alt identificeret og klassificeret 185 unikke usabilityproblemer. Heraf var 13 positive findings, og 10 var klassificeret som kritiske problemer. De 185 identificerede problemer blev dernæst grupperet i 18 grupper, som igen blev overinddelt i seks grupper, som vist i Tabel 14 på side 65. Det blev på baggrund af en inter-rater reliability-test ikke fundet nødvendigt at gentage identifikation eller klassifikation af usabilityproblemerne, men navne, beskrivelser og overordnede klassifikationer af de 18 grupper af problemer blev let revideret.

Vi skal i de næste to afsnit se nærmere på de resultater, som spørgeskemaerne har givet og til sidst sammenholde disse med ovenstående for bedre at kunne illustrere tendenser samt forhåbentlig anvende spørgeskemaerne til at bekræfte og generalisere nogle af THT-resultaterne.

5.2 Første spørgeskema

I dette afsnit vil vi se nærmere på de resultater, som første spørgeskema har givet. For opbygning af spørgeskemaet og en liste med de stillede spørgsmål, se afsnit 4.3. Selve spørgeskemaet kan ses i Bilag K. Vi vil ikke gå i detaljer med diskussionen af resultaterne, men her blot gengive de vigtigste i tekstuel og grafisk form. Resten behandler vi i diskussionen i afsnit 6.

I alt 152 personer valgte at svare på det første spørgeskema, som var åbent for besvarelser i perioden 15. november til 31. december. Det er desværre ikke blevet undersøgt, *hvordan* respondenterne har fået kendskab til undersøgelsen, men det er vores vurdering, at en stor del har fået det gennem Facebook (58 har tilkendegivet, at de deltager i den oprettede begivenhed, 12 har sagt måske) og resten gennem opslag på elektroniske fora på nettet og mund-til-mund. På den indledende startside på forfatterens egen hjemmeside er der desuden lavet en tæller for, hvornår og hvor mange gange nogen har startet selve undersøgelsen, samt registreret besøgendes IP-adresse (for at kunne identificere spammere). Ud af de 436 besøgende, som har set startsiden, har 269 (61.7%) startet undersøgelsen, og heraf har 152 (56.5%) gennemført den. Den forholdsvis store *drop-off rate* på 43.5% kan skyldes flere ting, blandt andet at man undervejs har fundet ud af, at man alligevel ikke var i målgruppen eller simpelthen blot nysgerrighed omkring

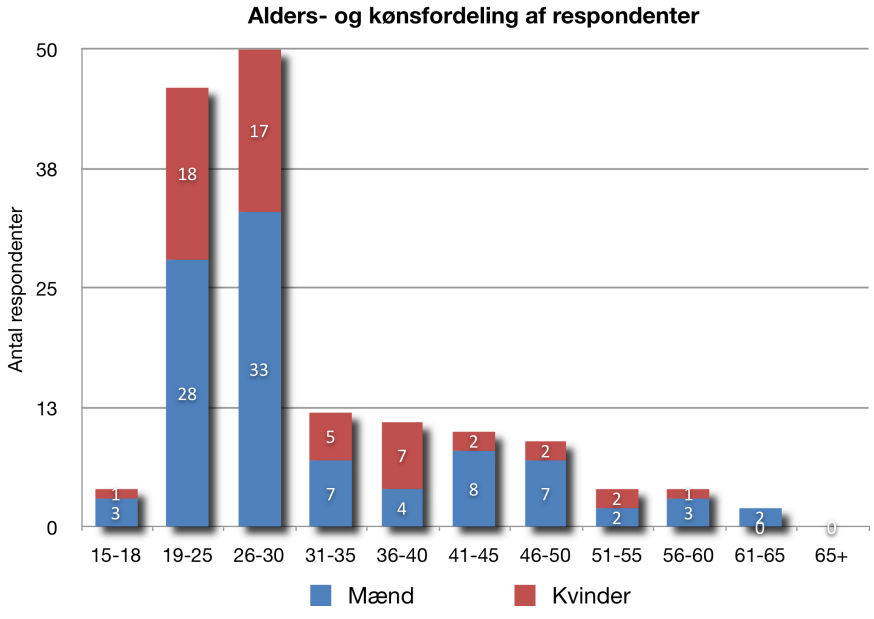
undersøgelsens indhold.

Med hensyn til antallet af besvarelser er vi overordnet set tilfredse. Antallet er stort nok til, at man kan udføre statistiske tests, samtidig med at gruppen er bred nok til at kunne relatere til den generelle population, som i vores tilfælde er de ca. 3 millioner danskere, der har modtaget et NemID med posten eller alternativt bestilt et selv. Det havde dog selvfølgelig være rart med langt flere og bredere besvarelser for med mere sikkerhed at kunne sige noget statistisk om resultaterne, men igen skal man her tage undersøgelsens forudsætninger og ressourcer i betragtning.

5.2.1 Karakteristik af respondentgruppen

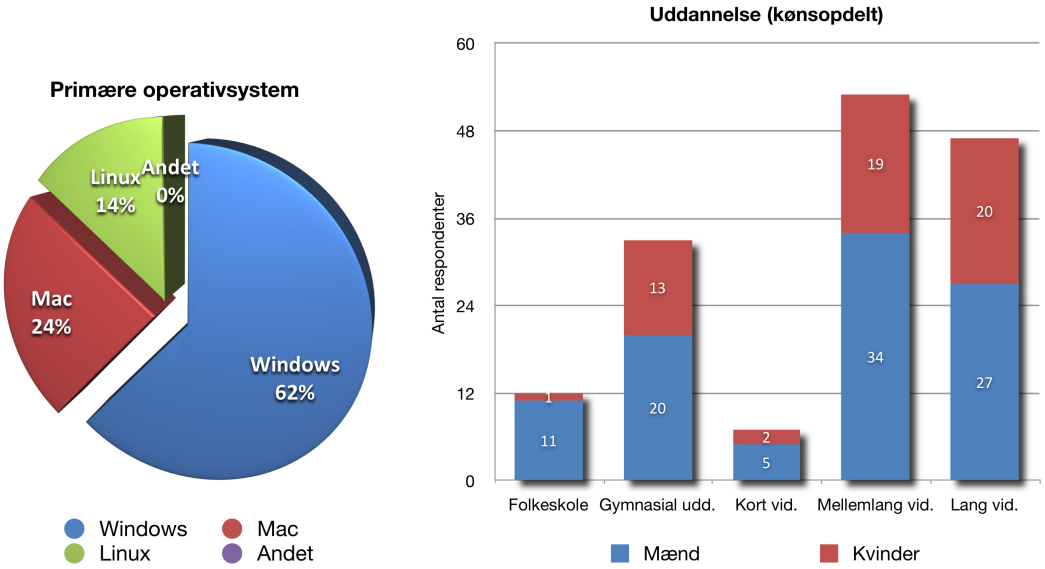
Vi vil nu ud fra de stillede spørgsmål omhandlende demografi og IT-færdigheder komme med en karakteristik af respondentgruppen. Vi bygger altså vores beskrivelse på spørgmålene 1-18 (se Tabel 3 på side 46).

Alder og køn Fordelingen af respondenterne på alder og køn kan ses i Figur 7. Som det ses af grafen, ligger størstedelen af massen i intervallet 19-30 år. Dette er selvfølgelig ikke helt tilfældigt, i og med at Facebook har leveret så stor en del af besvarelserne. Den oprettede begivenhed blev indledningsvist udsendt til forfatterens egen vennegruppe på Facebook, som i sagens natur indeholder flest i denne aldersgruppe. Man kan dog også forestille sig, at den pågældende aldersgruppe er mere tilbøjelig til at have lyst til at deltage i en semi-teknisk undersøgelse som denne. Blandt respondenterne udgjorde de 55 (36.2%) kvinder, hvilket må siges at være en nogenlunde tilfredsstillende fordeling. Aldersfordelingen spænder fra 16 til 61 år.



Figur 7: Fordeling af besvarelser på alder og køn.

Etnicitet og uddannelsesniveau Fire respondenter (2.6%) angav, at de havde et andet oprindelsesland end Danmark, og seks respondenter (4.6%) angav, at de havde et andet modersmål end dansk. Fordelingen af respondenternes uddannelsesmæssige baggrund kan ses i Figur 8. Der er i vores øjne tale om en bred repræsentation af uddannelsesniveau, hvor der i hver kategori forefindes et rimeligt antal besvarelser. Kønsfordelingen over de forskellige uddannelsestrin harmonerer ligeledes nogenlunde med den generelle kønsfordeling.



(a) Primære styresystem

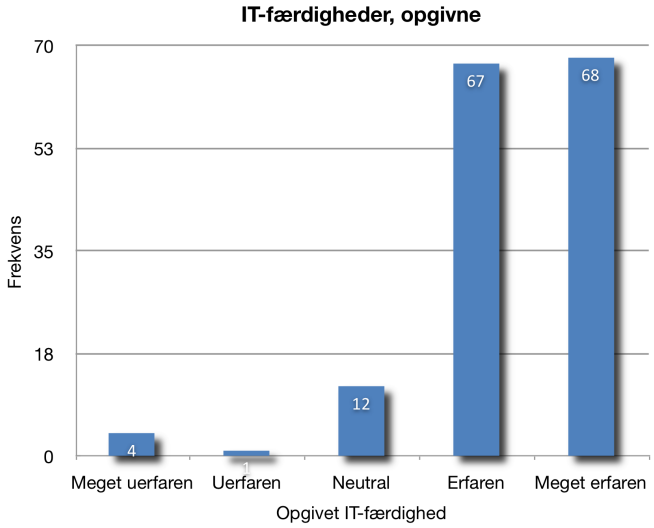
(b) Uddannelsesniveau fordelt på køn

Figur 8: Primære styresystem og uddannelsesniveau for respondenter

IT-færdigheder Respondenternes IT-færdigheder er målt på to måder: objektivt og subjektivt. Den objektive del baserer sig på, hvor og hvor ofte de anvender internet, om de har brugt den gamle digitale signatur, samt hvor ofte de anvender forskellige relevante tjenester på nettet (f.eks. netbank, borger.dk og eBoks). Den subjektive vurdering er foretaget ved at lade respondenten vurdere sine egne evner inden for IT på en skala fra "Meget uerfaren" til "Meget erfaren".

De objektive data kan ses i deres fulde omfang i Bilag L. Næsten samtlige respondenter (151, 99.3%) har angivet, at de anvender internettet dagligt og i hjemmet. 107 respondenter (70.4%) anvender ligeledes internettet på deres arbejdsplads. Fordelingen af respondenternes primære operativsystem kan ses i Figur 8a. Denne fordeling må forventes at være noget skæv i forhold til den generelle population, da antallet af Linux-brugere sandsynligvis er væsentlig højere her. Dette vil igen kunne forklares med det store antal respondenter i alderen 19-30 år samt udbredelsen af undersøgelsen på forfatterens eget studie. 111 respondenter (73.0%) har angivet, at de har anvendt den gamle digitale signatur. Dette tal er igen en smule højt i forhold til den generelle population.

Med hensyn til respondenternes subjektive vurdering af deres egne IT-færdigheder, kan fordelingen ses i Figur 9. Her springer det selvfølgelig hurtigt i øjnene, at respon-



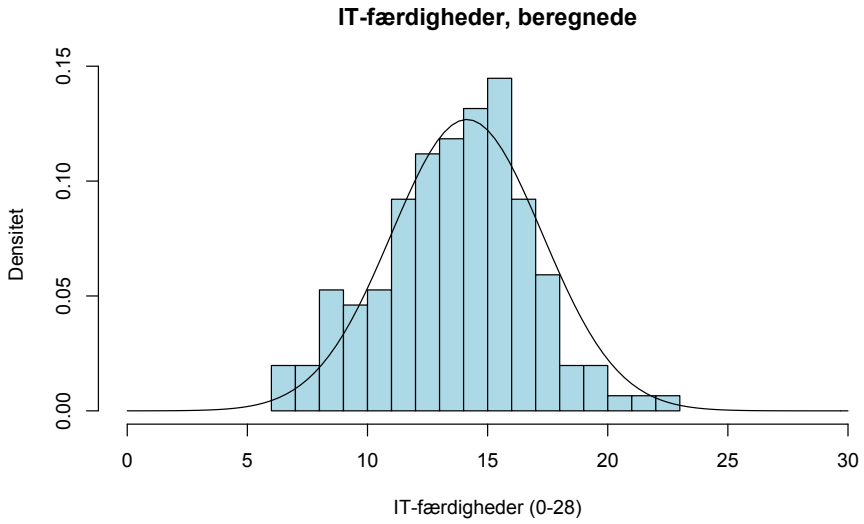
Figur 9: Respondenternes subjektive vurdering af egne IT-færdigheder

denterne enten er særdeles habile IT-brugere, eller at de måske har en mere subtil måde at vurdere deres egne evner på. Ligesom vi også observerede i THT'erne, kan folks subjektive opfattelse af egne evner variere meget på vurderingsgrundlaget. Der er dog heller ikke i spørgsmålet sat specifikke grænser på, hvordan de enkelte punkter på skalaen relaterer til praktisk erfaring eller konkret kendskab til IT. Det blev dog vurderet i planlægningsfasen, at en sådan udpensling ville være for omfattende at medtage i et allerede forholdsvist omfattende spørgeskema.

Vi kan dog stadig komme frem til en mere objektiv vurdering af respondenternes overordnede IT-færdigheder. Ser vi på spørgsmålene i kategorien IT, kan vi ud fra disse konstruere en score for hver enkelt respondent. Scoren er beregnet på følgende måde:

$$\begin{aligned} it\_beregnet \quad = \quad & netbank\_frekvens + skat\_frekvens + borger\_frekvens + sundhed\_frekvens \\ & + eboks\_frekvens + internet\_frekvens + anvendt\_gammel * 4 \end{aligned}$$

Denne udregning giver en score mellem 0 og 28 (hvor 28 er bedste score), som baserer sig på, hvor ofte respondenten anvender sites, der kræver login med digital signatur eller NemID (spørgsmål 13-17), hvor ofte de anvender internettet (spørgsmål 11), samt hvorvidt de har anvendt den gamle digitale signatur (spørgsmål 18). Selvom frekvensen af de anvendte sider her tillægges en væsentlig del af scoren, er dette efter vores mening en god indikator for respondentens erfaring med lignende systemer og onlinetjenester generelt og et udmærket supplement til den subjektive vurdering. Som man også kan se på frekvenserne for de forskellige tjenester, er det kun netbank, der har daglige brugere blandt respondenterne, ligesom mange respondenter aldrig har anvendt tjenester som f.eks. eBoks eller borger.dk. De beregnede IT-færdigheder kan ses i Figur 10.



Figur 10: Respondenternes beregnede IT-score. Den indlagte kurve er en standard-normalfordeling med middelværdi 14 og samme spredning som for besvarelserne.

Som det kan ses af Figur 10, fordeler de beregnede IT-færdigheder sig væsentlig anderledes end de subjektive – de kan med god tilnærmelse endda siges at være normalfordelte, hvilket Shapiro-Wilks test for normalitet heller ikke kan afkræfte ($p = 0.08978$).[30] Hvis vi dermed ser på den beregnede score som mål for fordelingen af respondenternes IT-færdigheder, har vi en fordeling, som er mere retvisende for deres erfaring med lignende systemer, online selvbetjening og internettet generelt. Scoren skal dog selvfølgelig tages med en del forbehold, som også nævnt ovenfor.

Hvorvidt fordelingen af respondenternes IT-færdigheder er repræsentativ for befolkningen generelt, er lidt sværere at svare på. IT- og Telestyrelsens It-barometer[31] angiver f.eks. for 2009, at 19% af befolkningen aldrig har anvendt en computer (niveau 0), mens 29% har "meget stærke it-færdigheder" (niveau 3, hvilket desværre ikke er defineret nærmere). Så set i forhold til den subjektive vurdering har vi ikke ramt særlig godt – den objektive er tættere på. Vi har altså i undersøgelsen fat i IT- og Telestyrelsens niveau 1-3, men med et elektronisk spørgeskema ville vi i sagens natur også have svært ved at få fat i niveau 0-brugere, særligt når markedsføringen er foregået rent elektronisk.

5.2.2 Validitet og pålidelighed

Ligegyldigt hvilken type af undersøgelse, man foretager, er man interesseret i at finde ud af, hvor 'gode' (valide) data er, samt hvor pålidelig undersøgelsen generelt er. Ozok [26] skelner her mellem *validity* og *reliability*. Validity vil normalt skulle bedømmes, inden undersøgelsen skydes igang, og reliability efterfølgende. Da det ikke har været muligt at

[30] Se: http://en.wikipedia.org/wiki/Shapiro–Wilk_test
[31] Se: http://borger.itst.dk/it-barometer

tage sig af validity inden undersøgelsens start, behandles begrebet derfor nu. Validity kan opdeles i to underpunkter:

- **Construct validity** udtaler sig om, hvorvidt undersøgelsen bygger på tidligere resultater og teori for at bekræfte, at *"the survey didn't come out of the imagination of the designer, but rather relies on a number of different research studies conducted by a number of different researchers"* (Ozok [26], s1161). På baggrund af det udførlige teoretiske grundlag i afsnit 4.1.3 samt den grundige og konkrete planlægning i afsnit 4.3 er der efter vores mening basis for høj *construct validity* i dette spørgeskema.

- **Predictive validity** udtaler sig om, hvorvidt undersøgelsen har *"the ability and power [...] to predict correct results in repetitive use"* (Ozok [26], s1161). Predictive validity er lidt ulden at definere og bestemme for en konkret undersøgelse, men Ozok nævner *pilot testing* og *peer review* af andre eksperter inden for området som muligheder for at bedømme predictive validity. Begge dele er for nærværende undersøgelse foretaget: Erik Frøkjær (som også er vejleder på dette projekt) har grundigt gennemgået spørgeskemaet inden udsendelse, og desuden er spørgeskemaet blevet pilottestet på en repræsentativ respondent. På baggrund af dette anser vi spørgeskemaet for at have god predictive validity.

Med hensyn til *reliability* er der her tale om, hvorvidt undersøgelsens besvarelser er konsistente, samt om de er pålidelige. En stor del heraf bliver afgjort af, om respondenter forstår spørgsmålene på den måde, som spørgeren havde i tankerne, jævnfør kriteriet om entydighed fra Stone [34]. En ofte anvendt metode til at efterprøve *internal reliability* i spørgeskemaer er at indlægge to ens spørgsmål, men formuleret omvendt. Et eksempel er spørgsmålene "Jeg havde let ved at udfylde spørgeskemaet" og "Jeg havde svært ved at udfylde spørgeskemaet", hvor begge spørgsmål besvares på en enighedsskala. Man vil i dette tilfælde forvente, at respondenten svarer modsat på de to spørgsmål – andre udfald kan f.eks. indikere, at respondenten har sat tilfældige krydser og dermed ikke kan ses som en seriøs besvarelse. Ozok ser denne indlejring af ens spørgsmål som helt uundværlig i kvantitative undersøgelser inden for HCI.

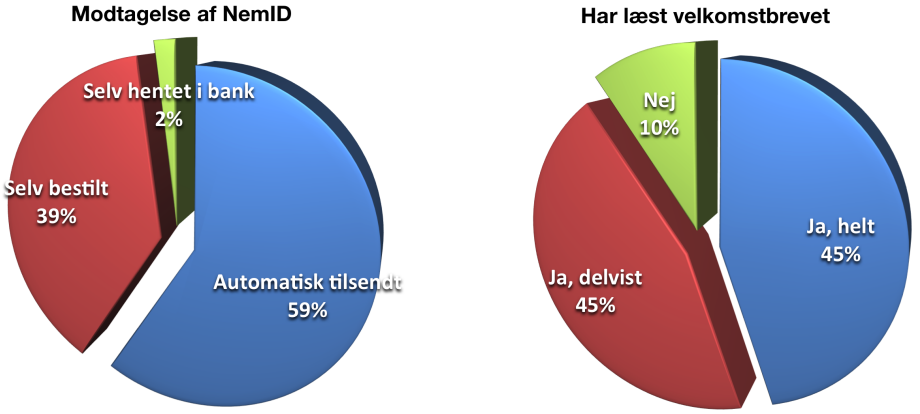
I denne undersøgelse er der med vilje *ikke* indlagt sådanne omvendte spørgsmål. Dette er fravalgt for at undgå at genere respondenterne mere end højst nødvendigt eller nedgøre deres intelligens. Der er dog to steder, hvor der er indlagt *næsten* ens, men omvendte spørgsmål: under oprettelsen af den selvvalgte adgangskode ("Adgangskoden var let at oprette" og "Jeg havde brug for mere hjælp") samt under indtastningen af en nøgle fra nøglekortet ("Jeg forstod min opgave" og "Jeg havde brug for mere hjælp"). Disse spørgsmål er ikke direkte modsætninger i den klassiske forstand, men til at teste reliability er de i vores øjne tilstrækkelige. Den konkrete test består i at udregne Cronbachs α-koefficient for de to sæt af ens spørgsmål. Cronbachs α er et tal mellem 0 og 1, hvor 0 indikerer dårligst og 1 bedst reliability. α-værdier større end eller lig 0.7 anses for at indikere god internal reliability (Ozok [26], s1162). Cronbachs α er beregnet i R med `cronbach()`-funktionen fra `multilevel`-biblioteket (se Bilag M). For oprettelse af adgangskode får vi $\alpha = 0.7337$ ($n = 152$), og for indtastning af nøgle får vi $\alpha = 0.7375$ ($n = 139$). Begge indikerer altså god internal reliability af spørgeskemaet.

5.2.3 Udvalgte resultater og tests

Efter at have karakteriseret respondentgruppen og argumenteret for, at undersøgelsens resultater er tilstrækkeligt valide og pålidelige, vil vi i dette afsnit gengive udvalgte resultater af spørgeskemaundersøgelsen samt se på eventuelle statistisk signifikante faktorer, der kunne have indflydelse på besvarelserne. De fulde resultater samt deskriptiv statistik kan ses i Bilag L. Der er af respondenterne afgivet en stor mængde subjektive kommentarer, som vi vil inddrage, hvor det er relevant. Disse kan ligeledes findes i Bilag L.

Modtagelse, velkomstbrev og nøglekort

Første del af spørgeskemaets mere interessante del omhandler modtagelsen af NemID, velkomstbrevet og nøglekortet generelt. Resultaterne af disse spørgsmål kan ses i Figur 11, 12 og 14. I Figur 13 og 15 kan ses de kønsopdelte holdninger til velkomstbrevet og nøglekortet.



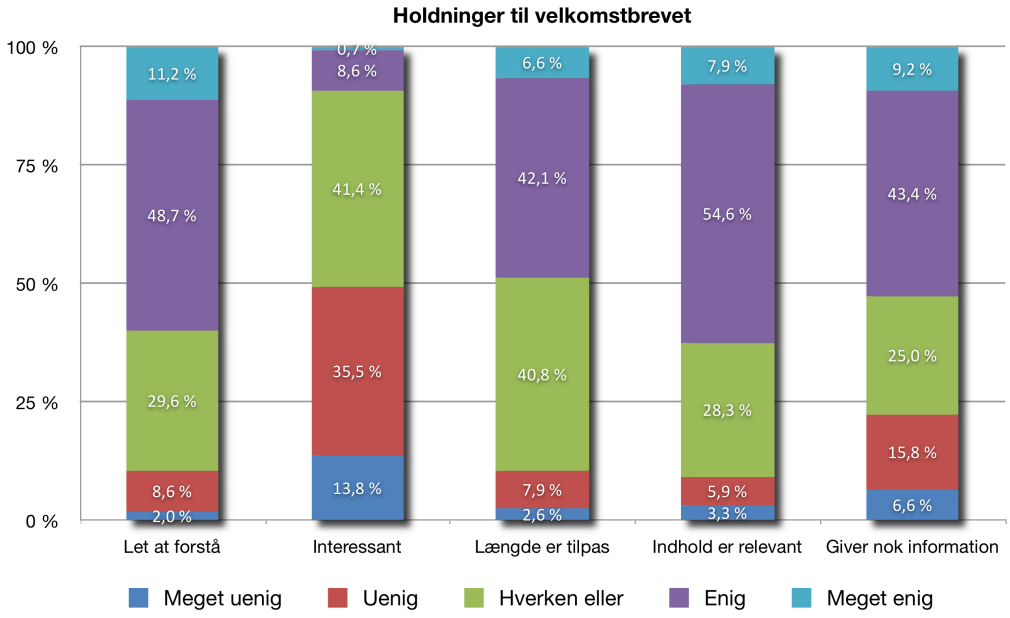
(a) Hvordan respondenterne har modtaget deres NemID.

(b) Andel af respondenter, der har læst velkomstbrevet.

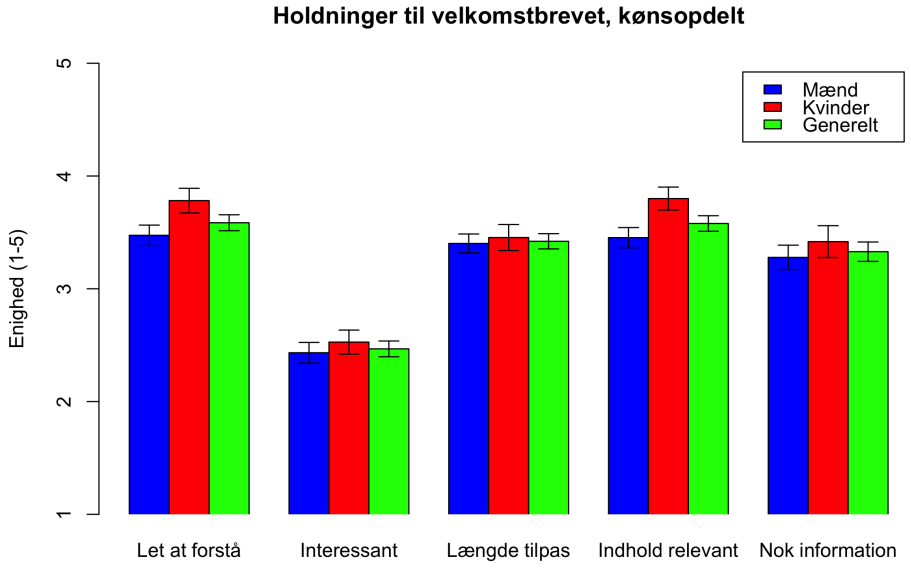
Figur 11: Respondenternes modtagelse af NemID og læsning af velkomstbrev.

Som det kan ses af Figur 11a, har en overraskende stor del af respondenterne selv bestilt deres NemID (39%), mens den overvejende del (59%) har modtaget det automatisk med posten gennem deres netbank. Dette harmonerer ikke specielt godt med de virkelige tal (hvor langt hovedparten modtager det automatisk gennem deres netbank), men for selve undersøgelsen er dette irrelevant. Det kan endog være en fordel, i og med at der så potentielt vil være flere, der benytter forskellige hjemmesider til aktiveringen.

Som det kan ses af Figur 12, synes næsten halvdelen af respondenterne, at velkomstbrevet er uinteressant, og ca. 10% synes ikke, at brevet har en tilpas længde eller indeholder relevant information. 22.4% synes ikke, brevet giver nok information. Ser vi på de kønsopdelte tal i Figur 13, ser vi, at kvinderne har en tendens til at være en smule mere positivt stemt overfor velkomstbrevet. Bortset fra spørgsmålet om, hvorvidt brevet er interessant, ligger gennemsnittet for vurderingerne en smule over middel, dog uden at være prangende.



Figur 12: Holdninger til velkomstbrevet, generelt.



Figur 13: Holdninger til velkomstbrevet (gennemsnit), kønsopdelt. Der er indlagt standardfejl på 5% i grafen for at illustrere eventuelle statistiske forskelle.

Overordnet har 90% læst velkomstbrevet helt eller delvist. De 10%, som ikke har, angiver, at de enten ikke gider eller ikke mener, der står noget vigtigt eller relevant i brevet:

"Der var så mange papirer, og jeg tænkte at der nok ikke stod noget vigtigt i

nogen af dem. Hvis det viser sig at der er noget jeg skulle have læst, så bliver jeg nok opmærksom på det, når jeg forsøger at tage NemID'et i brug, og så kan jeg læse det når det bliver nødvendigt."

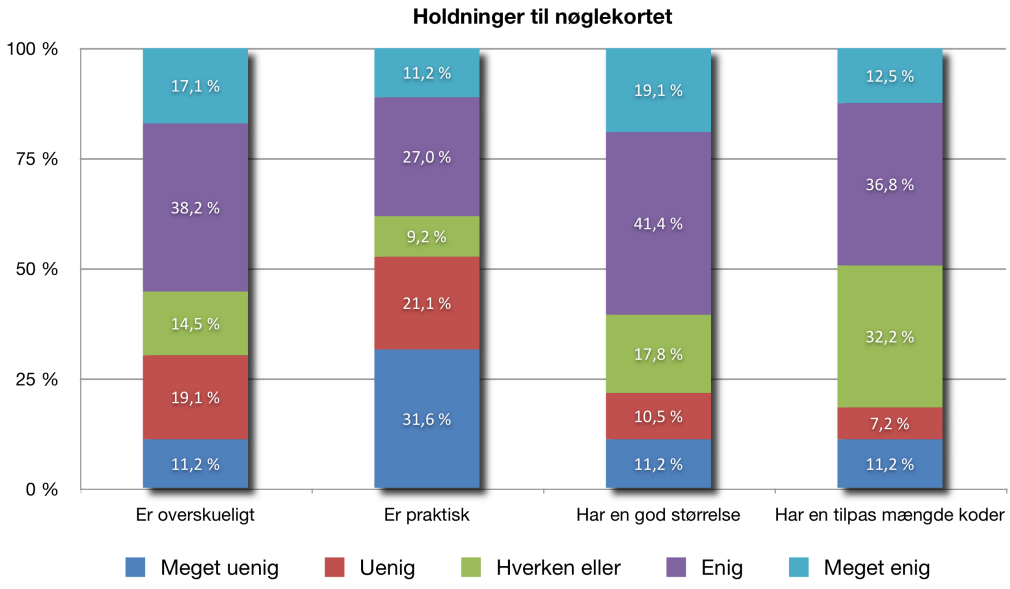
"Allerede efter overskriften lagde jeg brevet fra mig, da jeg ikke synes at den lød indbydende til at læse videre."

Der er ligeledes kommet en del kommentarer til velkomstbrevet generelt:

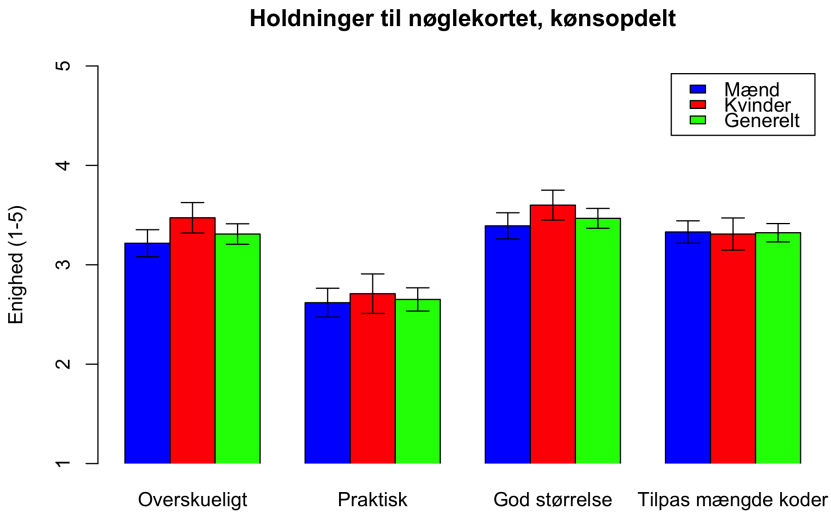
"Det værste ved velkomstbrevet er, at der ikke gives ordentlig information om, hvor man skal logge på første gang."

"Velkomstbrevet var langt og kedeligt. Mange unødvendige informationer blandet ind i en masse tekst. Man skulle lede efter de centrale informationer mellem en masse tekst med oplysninger som virkede selvindlysende."

Nøglekortet i sig selv er ikke blevet så positivt modtaget. Som det kan ses i Figur 14, synes 30.3%, at kortet er uoverskueligt, og lidt over halvdelen (52.7%) synes, at det er upraktisk. Dette må siges at være en meget stor del af respondentgruppen. Ser vi på den kønsopdelte udgave af de samme spørgsmål, ligger kvinderne igen generelt lidt i den højere ende af tilfredshedsskalaen. Pånær spørgsmålet, om hvorvidt kortet er praktisk, ligger de resterende gennemsnit omkring middel.



Figur 14: Holdninger til nøglekortet, generelt.



Figur 15: Holdninger til nøglekortet (gennemsnit), kønsopdelt. Der er indlagt standardfejl på 5% i grafen for at illustrere eventuelle statistiske forskelle.

Der er kommet særdeles mange kommentarer til nøglekortet generelt, hvoraf langt de fleste er negative. Kommentarerne går primært på, at kortet er "upraktisk", "papirspild" og "oldnordisk", og at det er irriterende at skulle have med sig overalt. De følgende kommentarer er endog meget blide i deres sprogbrug, i forhold til andre i samme bunke:

> *"Nøglekortet er en komplet hjernedød opfindelse. Det er papirspild, det økonomisk helt gak at skulle snail-maile nye nøglekort ud til folk hele tid og det er voldsomt irriterende nu at skulle holde styr på en åndssvag papirlap."*

> *"Problemet er at man intet kan foretage sig uden dette kort. Man skal altid have det med sig - det må ikke blive væk. Hvis man får stjålet sin taske (som jeg gjorde sidst jeg var i udlandet) skal man have tilsendt et nyt nøglekort før man igen kan bruge netbank osv. Det vil ikke fungere hvis man rejser meget, og er afhængig af netbank. Min søn er talblind, og uden min hjælp har han svært ved at anvende nøglekortet. (altså når vi efter lang tids søgen endelig får det fundet)"*

> *"Mht. nøglekortet synes jeg det er virkelig besværligt at skulle være afhængig af noget fysisk for at komme på netbank. Desuden synes jeg der er for få koder. Efter jeg har fået NemID er jeg begyndt at tænke over hvor ofte jeg går på netbank og vurdere om det er nødvendigt frem for at gøre det når jeg har behov for at checke mine konti eller lignende."*

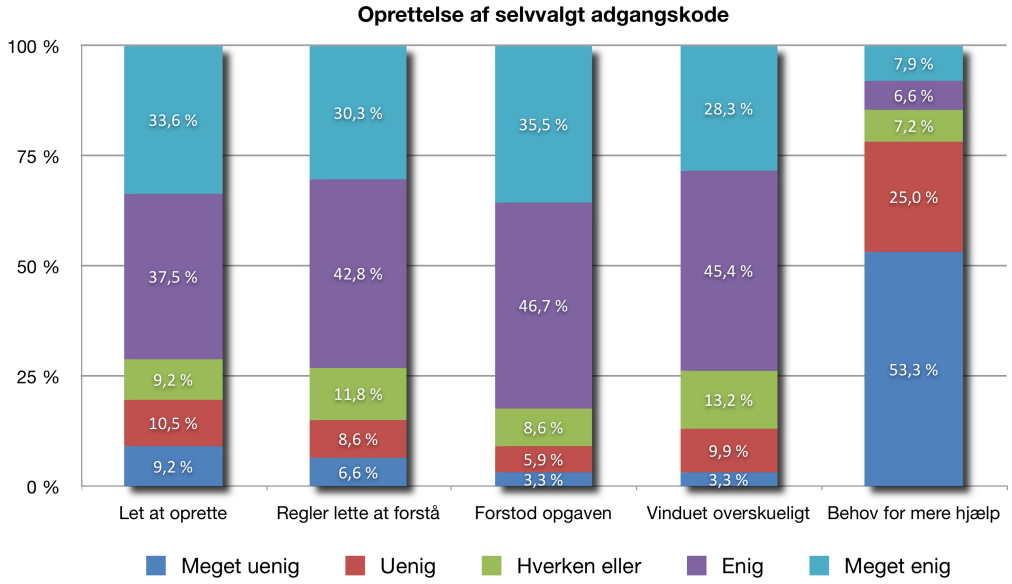
Oprettelse af selvvalgt bruger-id og adgangskode

58% af respondenterne valgte at oprette et selvvalgt bruger-id. Langt de fleste af dem,

som ikke oprettede et, vil hellere bare bruge deres CPR-nummer, som er nemmere at huske i forhold til at skulle finde på endnu et nyt brugernavn:

> *"Jeg kan bruge mit CPR nummer, og det kan jeg jo alligevel huske. Jeg kan ikke forstå hvorfor jeg skulle have lyst til at oprette flere brugernavne?"*

Processen, hvor brugeren bliver bedt om at oprette en selvvalgt adgangskode, er umiddelbart blevet vurderet positivt af de fleste respondenter (se Figur 16), dog synes 19.7% ikke, at opgaven var let, og 14.5% har haft behov for hjælp på den ene eller anden måde.



Figur 16: Holdninger til dialogen om oprettelse af adgangskode.

Selvom vurderingerne af dialogen ligger over middel, er det for 13 af respondenterne (9%) ikke lykkedes at oprette en selvvalgt adgangskode og dermed heller ikke at aktivere deres NemID. Disse respondenter er derfor blevet ledt videre til beskrivelsen af de eventuelle fejl og har ikke svaret på de resterende spørgsmål i spørgeskemaet. Hertil er der forskellige årsager:

> *"Timeout pga. jeg brugte specialtegn i eget password. Dette bevirkede at jeg skulle finde et nyt, gerne ligeså sikkert. Timeout kom dog først efter jeg havde valgt brugernavn ved næste skærmbillede."*

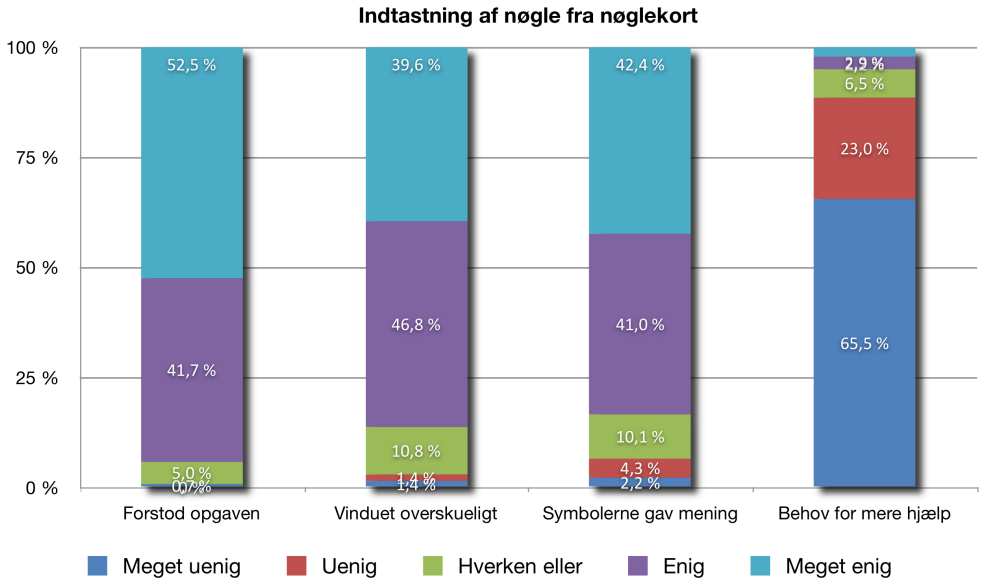
> *"Systemet skrev at der opstod en fejl. Jeg forsøgte herefter igen. Så skrev systemet at jeg havde brugt min aktiveringskode for mange gange. Nu skal jeg have tilsendt en ny. Det er uacceptabelt."*

Flere af fejlene hænger dog på den midlertidige adgangskode, som de fleste har fået tilsendt, og altså ikke direkte på oprettelsen af den selvvalgte adgangskode. De 13 respondenter, som ikke er lykkedes med aktiveringen, er altså et udtryk for de personer, der har forsøgt eller ikke har nået at forsøge at oprette en selvvalgt adgangskode. En respondent kommer med et forslag til en forbedring af dialogen på dette punkt:

> *"NemID anbefaler at man bruger specialtegn, men siger dog at nogle* ikke *er tilladt. Der står dog ingen steder hvilke tegn der tillades. Så en udførlig liste med tilladte tegn havde været god. Endvidere ville det have været rart hvis det blev oplyst at der var timeout på."*

Indtastning af nøgle fra nøglekort

Sidste trin i processen, inden oprettelsen er gennemført, er indtastningen af en nøgle fra nøglekortet for at bekræfte identiteten af brugeren. Vurderingen af dialogen og processen kan ses i Figur 17.

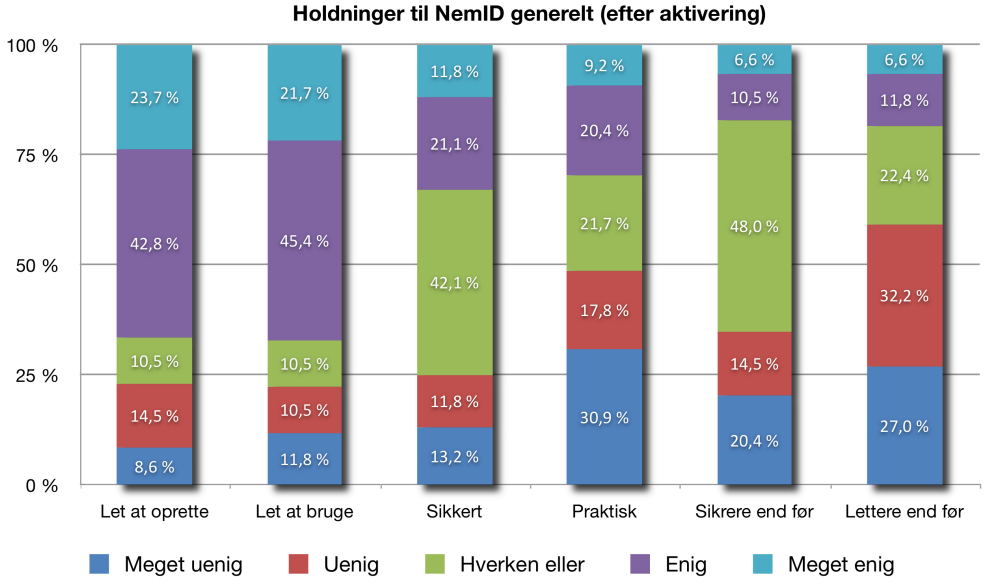


Figur 17: Holdninger til indtastning af nøgle, generelt.

Som det kan ses af Figur 17, har denne dialog virket overraskende positivt og har ikke givet anledning til problemer for langt de fleste brugere. Her er der da også kun to respondenter (1.4% af de tilbageværende), som dette er mislykkedes for, bl.a. på grund af sammenblanding af linierne på kortet.

NemID generelt

De sidste spørgsmål til NemID i spørgeskemaet går på respondenternes overordnede holdning til NemID efter endt aktivering. Resultaterne hørende til de seks spørgsmål kan ses i Figur 18 og i en kønsopdelt udgave i Figur 19.



Figur 18: Holdninger til NemID generelt.

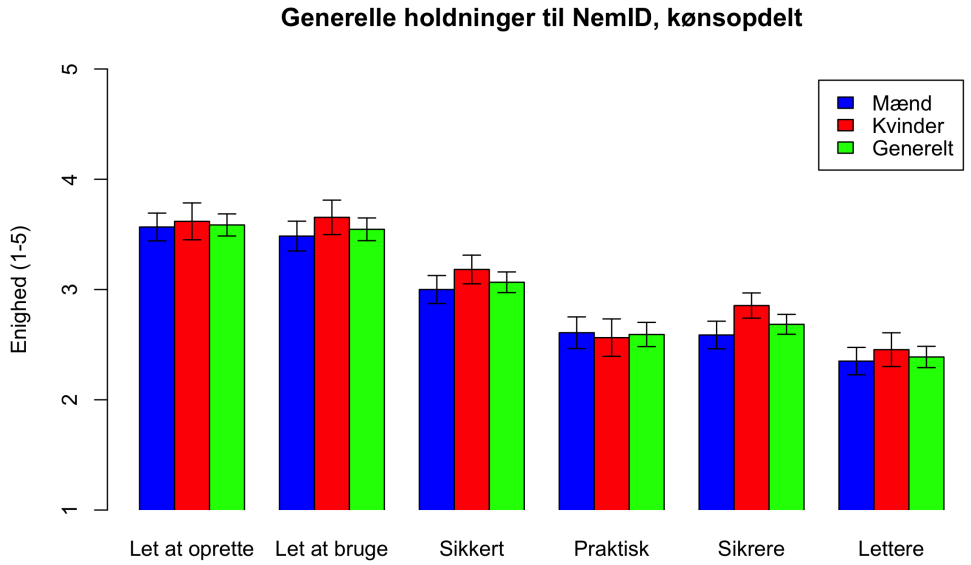
Trods en noget lunken vurdering af de enkelte trin i aktiveringsprocessen, vurderes NemID som værende overvejende let at oprette (76.9%) og anvende (77.7%). En fjerdedel mener dog ikke, det er sikkert, og næsten halvdelen (48.7%) synes, det er upraktisk. I forhold til tidligere løsninger (f.eks. netbank eller den gamle digitale signatur) mener 34.9%, at det er mindre sikkert, ligesom hele 59.2% mener, det er blevet *sværere* med NemID.

Ser vi på den kønsopdelte udgave i Figur 19, ligger kvinderne igen generelt en smule højere i vurdering end mændene, og vurderingerne ligger generelt lige over eller under middel. Der er dog ikke de store udsving i gennemsnittet mellem mænd og kvinder.

En interessant observation er dog, at mænd (pånær spørgsmål G1) har en tendens til at svare mere ekstremt på disse seks spørgsmål. Dette kan ses ved at lave en såkaldt *proportion test* i R (se Bilag M) for at se andelen af mænd, der anvender de forskellige svarkategorier. Resultatet af proportion-testen kan ses i Tabel 15. Her kan man observere, at i yderpunkterne er andelen af mandlige besvarelser pånær for G1 generelt større. Der er med andre ord flere kvinder, der anvender de midterste, mere neutrale kategorier. Som vi skal se senere, er flere af disse forskelle også statistisk signifikante.

Statistiske tests

Der er på baggrund af de modtagne besvarelser lavet en del statistiske tests for afhængighed mellem variable. Disse indbefatter blandt andet test for afhængighed mellem køn og generelle holdninger til NemID, alder og anvendelse af tidligere digital signatur. De udførte tests er lavet som χ^2-tests med 5%-niveauer i R (se Bilag M) og resultaterne kan ses i Tabel 16, hvor p-værdierne er gengivet for test mellem de to angivne faktorer. Da antallet af besvarelser ikke er voldsomt stort, har det givet god mening også at udføre



Figur 19: Holdninger til NemID generelt, kønsopdelt. Der er indlagt standardfejl på 5% i grafen for at illustrere eventuelle statistiske forskelle.

Spørgsmål	Meget uenig	Uenig	Hverken eller	Enig	Meget enig
G1	0.615	0.682	0.688	0.615	0.639
G2	0.722	0.750	0.563	0.594	0.667
G3	0.800	0.778	0.531	0.625	0.722
G4	0.702	0.481	0.606	0.677	0.714
G5	0.839	0.682	0.507	0.688	0.800
G6	0.732	0.551	0.676	0.556	0.700

Tabel 15: Andel af mandlige besvarelser for de generelle spørgsmål til NemID (spørgsmål 47-52 eller G1-G6) i det første spørgeskema. Bemærk, at andele ikke summerer til 1.0, da der er flere mandlige end kvindelige respondenter.

tests med kun tre enighedsgrader i relevante spørgsmål. Her er besvarelserne "Meget uenig" og "Uenig" slået sammen, ligesom "Enig" og "Meget enig" er det. Ellers risikerer vi, at de sammenlignede 'bunker' simpelthen bliver for små til at kunne sige noget statistisk meningsfuldt ud fra.

En overraskende generel bemærkning til de udførte tests er, at nærmest ingen faktorer er afhængige af hinanden, meget imod vores generelle forventning. Vi vil her knytte nogle korte kommentarer til de fundne resultater:

- Flere kvinder (98.2%) end mænd (85.6%) læser velkomstbrevet.
- Der er statistisk signifikant forskel på mænds og kvinders svar på spørgmålene omkring sikkerheden af NemID. Kvinderne er her mere neutrale (eller usikre, alt efter hvordan man tolker resultaterne) end mændene.
- Med tre enighedsgrader synes folk, der har anvendt den gamle digitale signatur, at

NemID er mere sikkert (gns. enighed = 3.20), end folk som ikke har anvendt den gamle signatur (gns. enighed = 2.74).

- Alder, køn, uddannelse, IT-færdigheder, hvorvidt man har læst velkomstbrevet, eller hvorvidt man har anvendt den gamle signatur har umiddelbart ingen indflydelse på, om man har held med at aktivere sit NemID. Her skal det dog bemærkes, at antallet af respondenter i denne undersøgelse, som ikke gennemførte, er forholdsvis lavt (15 ud af 152). Derfor er det svært at sige noget solidt om emnet.
- Køn, uddannelse eller IT-færdigheder har ingen indflydelse på, hvad man synes om NemID generelt.

Faktor 1	**Faktor 2**	p**-værdi**	**Tre grader**
Køn	G1 (Let at oprette)	$p = 0.9710$	$p = 0.8547$
Køn	G2 (Let at bruge)	$p = 0.6433$	$p = 0.3731$
Køn	G3 (Sikkert)	$p = 0.1104$	$p = 0.02961$
Køn	G4 (Praktisk)	$p = 0.3535$	$p = 0.6920$
Køn	G5 (Sikrere end før)	$p = 0.01503$	$p = 0.004927$
Køn	G6 (Lettere end før)	$p = 0.3955$	$p = 0.8429$
Køn	Succesfuld aktivering	$p = 0.5439$	-
Køn	Læst velkomstbrev	$p = 0.02622$	-
Alder (over/under 30)	Succesfuld aktivering	$p = 0.7322$	-
Alder (over/under 40)	Succesfuld aktivering	$p = 0.7660$	-
Anvendt gammel DS	G1 (Let at oprette)	$p = 0.4655$	$p = 0.3460$
Anvendt gammel DS	G2 (Let at bruge)	$p = 0.7493$	$p = 0.7176$
Anvendt gammel DS	G3 (Sikkert)	$p = 0.06614$	$p = 0.01686$
Anvendt gammel DS	G4 (Praktisk)	$p = 0.8212$	$p = 0.6694$
Anvendt gammel DS	G5 (Sikrere end før)	$p = 0.2805$	$p = 0.1952$
Anvendt gammel DS	G6 (Lettere end før)	$p = 0.2384$	$p = 0.2324$
Anvendt gammel DS	Succesfuld aktivering	$p = 0.1326$	-
Læst velkomstbrev	Succesfuld aktivering	$p = 0.9856$	-
Uddannelse (1-3)	G1 (Let at oprette)	$p = 0.7600$	$p = 0.9584$
Uddannelse (1/2/3)	G2 (Let at bruge)	$p = 0.6536$	$p = 0.2424$
Uddannelse (1/2/3)	G3 (Sikkert)	$p = 0.3827$	$p = 0.5195$
Uddannelse (1/2/3)	G4 (Praktisk)	$p = 0.1082$	$p = 0.1599$
Uddannelse (1/2/3)	G5 (Sikrere end før)	$p = 0.3618$	$p = 0.3599$
Uddannelse (1/2/3)	G6 (Lettere end før)	$p = 0.08897$	$p = 0.5420$
Uddannelse (1/2/3)	Succesfuld aktivering	$p = 0.7221$	-
IT-færdigheder (1/2)	Succesfuld aktivering	$p = 0.4779$	-
IT-færdigheder (1/2)	G1 (Let at oprette)	$p = 0.6062$	$p = 0.6908$
IT-færdigheder (1/2)	G2 (Let at bruge)	$p = 0.8830$	$p = 0.5686$
IT-færdigheder (1/2)	G3 (Sikkert)	$p = 0.9990$	$p = 0.9734$
IT-færdigheder (1/2)	G4 (Praktisk)	$p = 0.2193$	$p = 0.2001$
IT-færdigheder (1/2)	G5 (Sikrere end før)	$p = 0.3632$	$p = 0.6227$
IT-færdigheder (1/2)	G6 (Lettere end før)	$p = 0.8922$	$p = 0.7936$

Tabel 16: Liste over udførte korrelationstest (χ^2-tests). Ved tre enighedsgrader er "Meget uenig" og "Uenig" slået sammen, og "Meget enig" og "Enig" ligeså. Uddannelsesniveau er samlet i 3 grupper, ligesom IT-færdigheder (subjektive) er inddelt i to grupper. Statistisk signifikante sammenhænge er gengivet med grå baggrund.

Vi vil komme nærmere ind på disse observationer i diskussionen i afsnit 6.

5.3 Opfølgende spørgeskema

I dette afsnit vil vi se nærmere på de resultater, som det opfølgende spørgeskema har givet. For opbygning af spørgeskemaet, se afsnit 4.4. Selve spørgeskemaet kan ses i Bilag N.

Ud af de 92 respondenter, der i det første spørgeskema sagde ja til at blive kontaktet igen, har i alt 66 (71.7%) svaret på det opfølgende spørgeskema, som var åbent for besvarelser i perioden 10. december til 10. februar. Selvom der af praktiske årsager i emailen med linket til spørgeskemaet var angivet en tidsfrist for besvarelse, har respondenterne fået lov til at svare, når de lystede. Tidsfristen var indført for at forsøge at sikre konsistens i antallet af uger, der gik efter aktiveringen.

Igen kunne vi godt have tænkt os et større antal besvarelser, men en svarprocent på 71.7 er ganske godt sammenlignet med lignende undersøgelser beskrevet i litteraturen (som eksempelvis i Ozok [26]). På grund af spørgsmålenes opbygning er der dog i dette spørgeskema ikke mulighed for at lave meningsfyldte statistiske tests på resultaterne (vores samples er afhængige), ligesom vi heller ikke har mulighed for at sammenholde resultaterne for hver enkelt respondent med resultaterne fra det første spørgeskema. Vi vil dog stadig tillade os at sammenligne de to spørgeskemaer, særligt de generelle spørgsmål til NemID, da respondentgruppen er stor nok til at komme med kvalificerede bud.

Vi vil for det opfølgende spørgeskema ikke gå i deltaljer med validitet og pålidelighed, da spørgeskemaet er forholdsvis kort, samt at halvdelen af spørgsmålene er gengangere fra det første spørgeskema, hvis validitet og pålidelighed vi allerede har begrundet. Konstruktionen af skemaet bygger selvfølgelig på samme grundlag som det første, og er igen blevet reviewet og pilottestet inden udsendelse. Vi vil på baggrund af dette mene, at også dette spørgeskema har god validitet og pålidelighed.

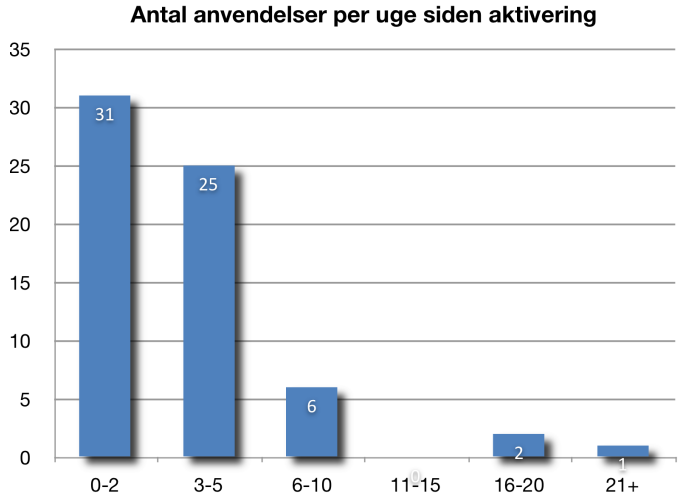
Brug af NemID siden aktivering

Til spørgsmålet, om hvor mange gange siden aktiveringen respondenterne har anvendt deres NemID, er hovedvægten i det lave område, som det kan ses i Figur 20. Der er dog et par "power users", som anvender deres NemID indtil flere gange om dagen. Her skal det dog bemærkes, at udformningen af det første spørgsmål i skemaet har givet misforståelser i et par tilfælde. Den supplerende tekst til spørgsmålet har givet et hint til, hvordan man finder antallet af *resterende* koder på nøglekortet, men selve spørgsmålet beder om antallet af *anvendte* koder. En enkelt respondent angav en dato for aktivering, som lå før, NemID overhovedet blev lanceret – denne besvarelse er ikke medtaget i grafen i Figur 20.

Problemer opstået undervejs

22.7% af respondenterne angiver, at de siden aktiveringen har oplevet problemer med deres NemID. Disse problemer er af forskellig karakter, mange primært tekniske men dog også helt basale fejl:

> *"NemID blev spærret fordi de have sendt kode og brugernavn samme dag. Dernæst blev min digitale signatur spærret."*



Figur 20: Frekvens af antal anvendelser af NemID per uge (gns.).

> *"Jeg aktiverede NemID med min netbank og brugte ikke engangskoden, der kom med brev. Da jeg dagen efter forsøgte at logge på Netbank igen fik jeg fejlbeskeden (sådan cirka): Din engangskode er udløbet, dit NemID er spærret. To timer senere kunne jeg godt logge på med NemID."*

> *"Det er ikke altid, den registrerer et login. Eller - den tæller ned på nøgleantallet, men logger mig ikke ind..."*

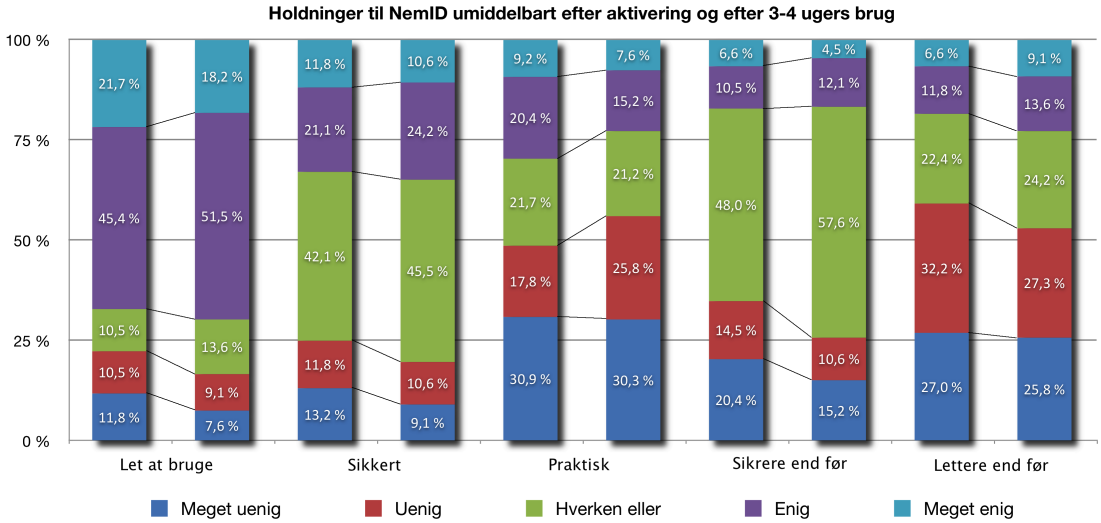
Det første citat ovenfor er udtryk for de store problemer, som DanID på et tidspunkt havde med mange tusinde brugere, der modtog deres nøglekort og midlertidige adgangskode med posten samme dag – noget som Datatilsynet også har kritiseret.[32]

Opfølgende holdninger til NemID

Den måske nok mest interessante del af det opfølgende spørgeskema er at se på udviklingen i respondenternes generelle holdninger til NemID. Der er her tale om spørgsmålene G2-G6 fra det første spørgeskema (G1 om oprettelsen er udeladt, da respondenterne i så fald vil skulle huske tilbage på processen). En samlet graf for de oprindelige og opfølgende holdninger kan ses i Figur 21, og en gennemsnitlig score for spørgsmålene kan ses i Figur 22.

Som det kan ses af Figur 21, er der ikke nogen større forskel i, hvordan respondenterne har evalueret NemID. En interessant observation er dog, at der for alle faktorer *pånær* det praktiske aspekt er en svagt positiv tendens. Respondenterne er altså umiddelbart blevet en smule gladere for deres NemID, både i forhold til produktet selv og set i forhold til deres tidligere anvendte løsninger, men de synes i højere grad, at det er upraktisk.

[32] Se f.eks. http://www.datatilsynet.dk/afgoerelser/seneste-afgoerelser/artikel/vedroerende-krav-om-spaerring-af-27611-certifikater



Figur 21: Holdninger til NemID generelt, umiddelbart efter aktivering og efter nogle ugers brug.

Den samme tendens kan observeres i Figur 22, hvor vi også kan se, at samtlige faktorer stadig ligger lige over eller lige under middel.

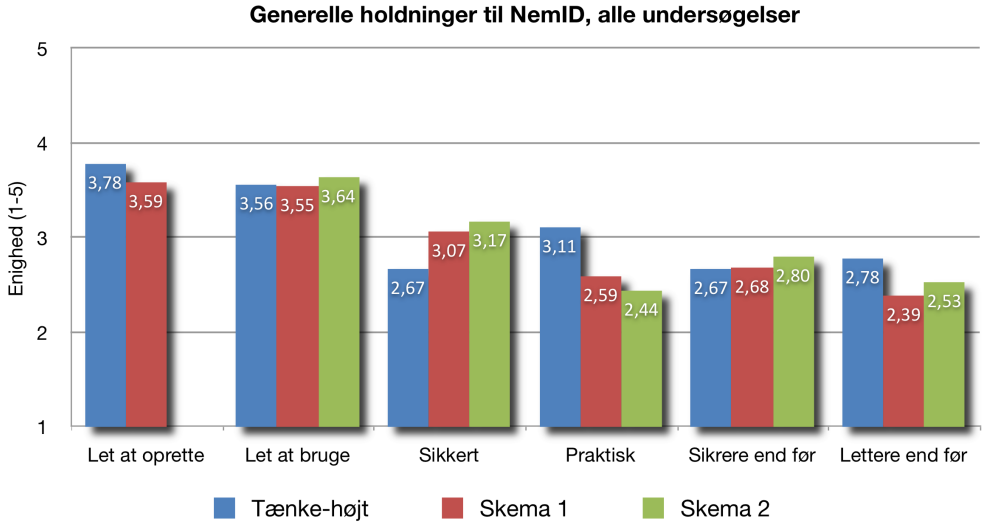
5.4 Sammenfatning

I dette afsnit vil vi kort sammenfatte de udledte resultater fra både THT'erne og spørgeskemaerne. Undervejs vil vi forsøge at sammenholde resultater fra de forskellige metoder, så tendenserne træder tydeligere frem.

For THT'ernes vedkommende har vi set, at der er observeret i alt 185 unikke usabilityproblemer, hvoraf de 13 er positive findings, og de 10 er karakteriseret som kritiske. Der er altså i alt identificeret 172 deciderede problemer af større eller mindre grad. Med hensyn til selve aktiveringen er der observeret task failure i 4 ud af 9 tilfælde, hvilket er væsentlig højere end for spørgeskemaerne (9.9%). Det kan ikke udelukkes, at det har haft en indflydelse, at den i THT stillede opgave udover aktiveringen inkluderede at logge på SKATs selvbetjening, men i og med at denne opgave er særdeles realistisk, er processen med at tilknytte et OCES-certifikat set som en naturlig del af aktiveringsprocessen.

I forhold til de generelle holdninger til NemID, harmonerer resultaterne fra THT'erne og spørgeskemaerne godt (se Figur 22), dog er der hos THT-testpersonerne større tvivl om sikkerheden men en mere positiv indstilling til det praktiske aspekt. NemID som produkt er dog i alle tre undersøgelser og på de undersøgte aspekter vurderet til lige over eller lige under middel.

Med hensyn til de identificerede problemområder er det umiddelbart svært at drage paralleller mellem de forskellige undersøgelser, da der er væsentlig forskel på beskrivelser observeret af observatøren under en THT og rapporteret af en respondent i et spørgeskema. Detaljeringsgraden er væsentlig højere for THT, men det kan alligevel



Figur 22: Holdninger til NemID generelt (gns.) for THT og spørgeskemaer.

observeres, at brugerne af NemID generelt er enige om, at indtastning af en nøgle fra nøglekortet pånær få undtagelser er en smertefri proces, ligesom velkomstbrevet (udover ikke at være interessant for brugerne) scorer forholdsvist neutralt. De umiddelbart største problemer ved NemID er selve nøglekortet og særligt det upraktiske i denne løsning, tiltroen til sikkerheden samt det forholdsvist store antal task failures under aktiveringen. Det kunne i denne forbindelse have været interessant at se på, hvor mange respondenter der havde valgt at bruge deres NemID til offentlige tjenester også, da det er her, to af de mest kritiske task failures i THT'erne forekom. En anden af årsagerne til de mange task failures kan hænge sammen med den kommentar til velkomstbrevet, der påpeger, at der mangler information om, hvor man skal logge på første gang.

Et stort antal testpersoner i THT'erne og respondenter i spørgeskemaerne har udtrykt meget stærke holdninger til det generelle koncept bag NemID, og særligt single-signon har hos THT-testpersonerne givet anledning til stor kritik og dermed mange identificerede problemer. Grundet THT'ernes mere konkrete opgaver er der ligeledes en del af de identificerede problemer, der har med relaterede tjenester og ikke direkte NemID at gøre. Som forventet har THT'erne afdækket mange af de åbenlyse *og* mere skjulte problemer, hvorimod spørgeskemaerne har leveret mere bred evidens for blandt andet testpersonernes holdninger til produktet.

6 Diskussion

I dette afsnit vil vi diskutere resultaterne fra analysen samt relatere dem til hinanden og sammenholde med de resultater, som firmaet Userminds er fremkommet med i deres testrapport. Vi vil også tage et kort sammenlignende kig på lignende undersøgelser og tiltag fra IT- og Telestyrelsen. Efter at have diskuteret resultaterne vil vi i afsnit 6.2 kritisk diskutere og vurdere de anvendte metoder og komme med retrospektive forslag til forbedringer. Strukturen i afsnittet vil være som følger:

- Diskussion af resultater
  - Identificerede problemstillinger
  - Testpersoners holdning til NemID
  - Sikkerhed og nøglekortet
  - Relation til Userminds' testrapporter
  - Opsummering
- Diskussion af metoder / opsamlede erfaringer
  - Tænke-højt test
  - Første spørgeskema
  - Opfølgende spørgeskema

6.1 Diskussion af resultater

Vi skal nu sammenfatte og kritisk diskutere de resultater, som er kommet frem gennem de foregående afsnit og de undersøgelser, som er foretaget i dette speciale. Målet er at komme med en form for samlet 'vurdering' af produktet NemID i forhold til særligt usability, set ud fra dets slutbrugeres synspunkt. Målet er *ikke* at komme med forbedringsforslag til systemet, hvilket ellers er formålet med de fleste usabilityevalueringer udført inden lanceringen af et bestemt produkt. I vores tilfælde *er* NemID allerede lanceret, og vi vurderer derfor retrospektivt dets kvalitet og potentiale.

6.1.1 Identificerede problemstillinger

I de udførte THT'er er der afdækket et forholdsvist stort antal usability-relaterede observationer (185 unikke), og hvis man fraregner *positive findings* samt de problemer, som ikke direkte har med NemID at gøre (single-signon), er vi nede på 139 deciderede usabilityproblemer. Nu er verden desværre sådan indrettet, at der ikke findes fejlfrie produkter, heller ikke dem som er blevet grundigt gennemtestet, så det var forventeligt, at NemID ville indeholde et antal problemer. Det er dog i vores øjne overraskende, at antallet er så stort og måske mere overraskende, at de fundne problemer er blevet klassificeret så højt, som de er. Et færdigt produkt må i vores øjne ikke indeholde 8-10 kritiske fejl og et meget stort antal alvorlige fejl.[33] En anden observation er her, at hvis det færdige produkt generelt afføder samme fejlrate, som vi har observeret i løbet af undersøgelserne, er der tale om et noget ustabilt og i visse henseender decideret uigennemtænkt system. Ni tilfældigt udvalgte personer kan ikke lede til et så stort antal usabilityprobemer, uden

[33] Vi minder læseren om Tabel 10 på side 60.

at de enten er en meget særlig gruppe, eller at systemets opførsel antageligt kan generaliseres til resten af befolkningen. Det er også bemærkelsesværdigt, at næsten 23% af respondenterne oplever problemer med systemet efterfølgende. Konkret må de væsentligste problemer siges at være forståelsen af koncepterne (særligt konsekvenser og brug af single-signon) og terminologien i NemID, det upraktiske i nøglekortet samt den store tvivl om sikkerheden af systemet.

Særligt single-signon er svær at placere i kritikken, fordi den som sådan ikke er nogen integreret del af NemID – det modsatte er derimod gældende. Der er dog identificeret en del problemer relateret direkte til single-signon, særligt overraskelsen og utrygheden ved systemets opførsel. Selvom de to ting principielt er adskilte, vil de uden tvivl blive set som én og samme ting af langt de fleste brugere, og vi må på denne baggrund konkludere, at der er behov for langt større information om single-signon til den generelle befolkning.

En anden bemærkelsesværdig faktor, som vi har set i både THT'erne og det indledende spørgeskema er raten for succesfuld aktivering. For THT'ernes vedkommende var 4 ud af 9 ikke i stand til selvstændigt at aktivere deres NemID og efterfølgende anvende det til at logge på en tjeneste. For spørgeskemaernes vedkommende var tallet ca. 10%. Her skal det bemærkes, at de to THT-testpersoner i gruppe A som sådan godt kunne aktivere deres NemID men ikke tilknytte et OCES-certifikat. Tæller man ikke disse to med som task failures, harmonerer antallet af fejlaktiveringer bedre for de to typer undersøgelser. Hvis man som et simpelt tankeeksperiment skalerer disse tal op til de ca. 3 mio. forventede brugere af NemID, vil dette betyde, at ca. 300.000 personer ikke selvstændigt kan aktivere deres NemID, hvilket bør anses som højst uacceptabelt og i strid med produktets navn og proklamerede styrker. Og her er ikke engang taget folk med funktionsnedsættelser eller synshandicap med i betragtningen.

Skal man supplere ovenstående med flere indikationer på, at noget er galt, kan man se på antallet af supporthenvendelser. Ifølge en undersøgelse, som sitet Version2.dk har lavet,[34] har mindst hver femte ringet til deres bank for at få support på NemID. Tallet er endnu højere, hvis man tæller opkald direkte til DanIDs support med. Selvom der er delte meninger om denne andels størrelse og betydning, er det ifølge DanID selv i omegnen af tre gange så stort som forventet. De fleste henvendelser drejer sig dog ikke om selve aktiveringen, men det er stadig en stor andel, som har brug for hjælp – flere end spørgeskemaerne indikerer, men i tråd med THT'erne. Mange af disse opkald vil efter vores opfattelse bunde i de konkrete identificerede usabilityproblemer, som belyses af dette projekt.

IT- og Telestyrelsen har i en redegørelse til Udvalget for Videnskab og Teknologi (se IT- og Telestyrelsen [16]) redegjort for mange af de kritikpunkter, som er blevet fremlagt i medierne og af flere folketingspolitikere. Redegørelsen bygger på arbejdet fra en taskforce nedsat af videnskabsministeren i oktober 2010 med det formål at *"sikre kvalitet i udrulningen af NemID"* (IT- og Telestyrelsen [16], s1). Taskforcen er primært kommet med forslag og har nedsat arbejdsgrupper, der skal sikre bedre oplysning og brugervenlighed for udsatte grupper, blandt andet handicappede og ældre. I redegørelsen er der desuden gengivet de fem største kategorier af spørgsmål stillet til DanIDs support. I perioden

[34]Se: http://www.version2.dk/artikel/17532-nemid-er-svaerid-hver-femte-har-brug-for-support

juli-oktober 2010 har 17.4% af de ca. 216.000 henvendelser drejet sig om "Aktivering af NemID". Der bliver ligeledes i redegørelsen stillet i udsigt, at der fremfor kun et enkelt forsøg til aktivering med den midlertidige adgangskode skal være tre – noget som muligvis kunne have afhjulpet en enkelt task failure i denne undersøgelse. Der vil desuden blive udarbejdet et revideret velkomstbrev, som bedre beskriver, at og hvordan brugeren skal aktivere sit NemID igennem den tjeneste, som vedkommende har modtaget det, ligesom en større indsats sættes i værk for at informere brugerne om korrekt log ud fra tjenester. Dette tiltag vil i vores øjne klart kunne gavne de borgere, der kommer i problemer med aktiveringsprocessen, hvilket vores undersøgelser tyder på, at en ganske stor vil.

I forhold til *learnability* har vi gennem THT'erne observeret, at testpersonerne generelt havde meget hurtigt ved at lære at bruge systemet, så snart det var blevet aktiveret og brugt første gang. Dette indikerer god *learnability* og er med til at styrke formodningen om, at NemID er hurtigt at vænne sig til og overordnet set let at bruge. Mange testpersoner og respondenter udtrykte sig dog kritisk om nøglekortet undervejs, men selve processen blev anset som ukompleks.

6.1.2 Testpersoners holdning til NemID

En anden interessant drejning på de udførte undersøgelser er testpersonernes holdninger til NemID generelt, med andre ord deres *subjective satisfaction*. Samtlige vurderingskriterier ligger omkring middel for et system, som har kostet lige i underkanten af 1 milliard kroner og skal bruges så ofte af så mange mennesker – man burde måske have forventet mere for pengene. Selvom det opfølgende spørgeskema i tråd med lignende undersøgelser (mere herom om lidt) viste en svagt positiv tendens i brugernes holdning til systemet, så er der i vores øjne stadig tale om en lunken modtagelse. Selvom resultater på SUS-skalaen kan diskuteres, kan en gennemsnitlig score på 54.4 under ingen omstændigheder anses som god.

I IT- og Telestyrelsens redegørelse citeres tre andre tilfredshedsundersøgelser af NemID. Finansrådet konkluderer, at *"den altovervejende del af netbankkunderne er tilfredse med den måde, som NemID fungerer på, men der er dog blandt netbankkunderne kritikere, særligt blandt de meget it-kyndige"* (IT- og Telestyrelsen [16], s5). Vi kan fint ud fra vores undersøgelser genkende dette billede, selvom vi ikke får nærmere defineret, hvad 'altovervejende' dækker over. En anden undersøgelse indikerer, at 77% er 'tilfredse', og en tredje beretter, at 81% er 'positive' overfor NemID (hvor mulighederne åbenbart kun er enten 'positiv' eller 'negativ'). Vi vil rent subjektivt mene, at i omegnen af 80% ikke er nok til at kalde noget en 'succes', ligesom tallenes tvetydighed gør det svært at sige noget konstruktivt ud fra.

En stort anlagt undersøgelse foretaget blandt Danske Banks kundepanel på 6.600 personer (se udvalgte resultater fra denne i Bilag R) konkluderer, at en fjerdedel (24%) overordnet set er *utilfredse* med NemID, mens hele 19% hverken er tilfredse eller utilfredse. Lige under halvdelen indikerer altså, at de ikke er positivt stemt overfor systemet. Undersøgelsen viser dog også (i tråd med vores egne undersøgelser), at jo længere tid folk har haft deres NemID, jo mere tilfredse bliver de, ligesom det også interessant nok observeres, at flere mænd end kvinder er særdeles utilfredse. Noget, som vores proportionstest i Tabel 15 i afsnit 5.2 også er med til at underbygge. Hele 85% indikerer i undersøgelsen,

at de ikke kommer til at bruge deres netbank oftere med NemID end før – noget som netbankerne ellers gerne havde set som en positiv sidegevinst ved indførelsen af systemet – ligesom 61% ikke er enige i, at det er blevet lettere med NemID (i vores undersøgelse er denne andel på 82%). Sidst, men ikke mindst, mener 24%, at det har været besværligt at komme igang med NemID.

En anden interessant betragtning i vores undersøgelser har været, at der umiddelbart ikke er de store forskelle at se, hvis man sammenligner på køn, alder, uddannelse, IT-færdigheder og lignende. Vi havde forventet, at der ville være forskelle i svarene givet på tværs af nogle af disse grupperinger, men hverken medierne eller IT- og Telestyrelsen har da heller givet dette nogen større opmærksomhed, måske lige bortset fra de særligt udsatte grupper (ældre og personer med handicap eller funktionsnedsættelser), som der ikke blev tid til at se nærmere på i dette projekt. Visse undersøgelser antyder dog også, at de unge har et mere stramt forhold til NemID end andre aldersgrupper, igen primært på grund af nøglekortets betydning.[35]

Nøglekortet Særligt selve nøglekortet, som NemID ofte identificeres med, har vakt stor furore og givet anledning til irritation blandt brugerne – noget som medierne i høj grad også har kørt med på. Det virker mærkeligt, at man i 2010 vælger at anvende papkort i stedet for elektroniske løsninger, som alligevel er på vej for NemIDs vedkommende i 2011. Hvor meget man end kritiserer de upraktiske og umiddelbart økonomisk tåbelige aspekter af en papkorts-løsning, har der ligget grundige overvejelser og tests forud for beslutningen:

> *"Kortet er jo ikke mere sikkert end diverse RSA-tokens og lignende - det er tilsvarende sikkerhed. Årsagen til, at det blev papkort, var, at vi havde testet det på mange brugere. Kan man finde ud af det, kan man forstå sikkerheden i det, kan man forstå at beskytte den del, som man skal beskytte? Og det bongede altså ud til, at et nøglekort er meget nemmere at forstå. Et godt eksempel er, at når de ser sådan en her [RSA-token, red.], så er de i tvivl om, hvordan de her seks cifre er kommet frem, og om det nu er sikkert. Og af en eller anden grund så synes de, det er mere sikkert, at tallene står på et papkort. Opfattelsen er så også, at hvis man får sådan et elektronisk device, efter man har fået et nøglekort, så har man nemmere ved at acceptere det, for det er jo bare et nøglekort, der er digitaliseret. [...] For at det overhovedet kunne lade sig gøre at rulle [NemID] ud på et halvt år, så var det nøglekortsløsningen...der var nogle produktionstekniske årsager til, at vi gjorde det på den måde."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Efter vores mening har DanID måske nok gjort sig overvejelser omkring papkort eller elektronisk løsning, men de efterfølgende undersøgelser har vist, at kritikken af papkortet er langt større end forventet og sandsynligvis er med til at trække helhedsvurderingen af NemID ned – noget som i vores øjne ville kunne have retfærdiggjort en længere udrulningsproces eller elektroniske nøglegeneratorer. Førnævnte undersøgelse fra Danske Bank konkluderer da også, at 32% synes, NemID er upraktisk. Det bliver altså interessant at se, hvor mange der kommer til at vænne sig til papkortet eller skifter til mere

[35] Se f.eks. http://www.version2.dk/artikel/17572-nemid-er-ikke-nede-med-de-yngre

moderne elektroniske nøglegeneratorer, når disse bliver lanceret senere på året, og hvorvidt dette vil have en positiv indflydelse på brugertilfredsheden. Nøglekortet er og bliver et irritationsmoment for mange brugere, men man skal her overveje, om en elektronisk kodegenerator ikke vil give anledning til samme irritation; den skal blot ikke udskiftes nær så ofte og vil virke mindre 'oldnordisk'. Vi rører her ved det velkendte problem om bekvemmelighed kontra sikkerhed. To faktorer, som oftest modarbejder hinanden.

Sikkerheden Særligt sikkerheden i NemID og single-signon-konceptet har fået drøje hug i vores undersøgelser. Folk er ikke glade for at skulle "lægge alle deres æg i én kurv", en kurv som er administreret af et privat selskab. En del testpersoner og respondenter har decideret konspiratoriske tanker i forhold til DanIDs rolle som udbyder, særligt i forhold til ikke at have mulighed for at opbevare sin egen private nøgle. Her beroliger DanID igen:

> *"ABP: Så i forhold til den gamle signatur vil du mene, at nøglerne er langt mere beskyttede nu?*
> *PLD: Helt afgjort! Jeg er sikker på, at hvis man vil angribe systemet, så er det, man skal have fat i brugeren - man ville ikke gå efter serveren. Vi beskytter brugerens nøgler efter samme principper, som vi også beskytter vores CA-nøgler, så hvis jeg var hacker og ville gå efter serveren... Lad os antage, at jeg ikke ville gå efter brugeren (hvilket ville være meget nemmere), men jeg ville gå efter serveren, fordi jeg ville opnå meget magt. Så tror jeg ikke, jeg ville gå efter den server, hvor brugeren havde deres private nøgler liggende.*
> *ABP: Du ville gå efter CA'ens?*
> *PLD: Ja, for så kan jeg lave mine egne identiteter, som jeg har lyst til. Det ville være meget mere magtfuldt end at gå efter brugernes nøgler. Så kan jeg opfinde mine egne identiteter, også på folk, som slet ikke er oprettet i systemet - dem kan jeg selv opfinde og lave certifikater til. Så jeg kan godt forstå kritikken, men det er måske mere ud fra en akademisk synsvinkel end ud fra den praktisk, sikkerhedsmæssige løsning. [...] Så hvis vi var onde, kunne vi bare bruge CA-nøglerne til at lave nogle nye private nøgler til eget brug, det havde vi ikke brug for brugernes private nøgler til at gøre."*
>
> Peter Lind Damkjær, DanID (Bilag P)

Så de konspiratoriske teorier og nervøsitet er nok mere et udtryk for holdninger end reel frygt for, at DanID skulle gå ud og misbruge identiteter. Det er dog interessant, at så mange har en forudindtaget holdning til NemID og sikkerheden i systemet. Vi begrunder i høj grad dette med mediernes dækning, som har været ganske massiv og mestendels negativt ladet. NemID er af medierne blevet døbt 'SværID' og mange andre navne, er blevet spået umiddelbar og hurtig fiasko, blevet beskyldt for at have en uhørt høj nedetid og lignende. Særligt i spørgeskemaundersøgelserne er mange respondenter gået særdeles hårdt til koncepterne, selvom de efterfølgende ikke har bedømt systemet specielt dårligt på de forskellige udstukne kriterier. Vi tror, at dette bunder meget i uvidenhed om de forholdsvis komplekse sikkerhedsmæssige mekanismer, der gemmer sig bag facaden, samt en udpræget 'det gamle virkede jo fint nok, hvorfor skal vi have noget nyt'-mentalitet. For langt de fleste brugere vil vi dog mene, at nøglerne nu er væsentlig bedre beskyttet end på brugerens egen PC.

Der er dog også faktorer i NemID, som uundgåeligt vil blive set som kritiske: man-in-the-middle-angreb (som vi så på i afsnit 3) og det faktum, at et nøglekort kan kopieres uden brugerens vidende – det kan en elektronisk kodegenerator ikke. Ligeledes har mange brugere allerede scannet deres nøglekort ind på computeren eller telefonen, så de ikke behøver finde det hver gang. Noget, som DanID på det kraftigste har frarådet, men som man bestemt ikke kan fortænke folk i at gøre. En anden overvejelse er, at mange mennesker sandsynligvis vil have deres nøglekort i pungen, ved siden af flere kort indeholdende deres CPR-nummer. Hermed er halvanden af de to faktorer i hus for eventuelle lommetyve eller mere målrettede elektroniske tyvebander.[36] Her vil en elektronisk nøglegenerator igen kunne sænke risikoen en smule. Hertil kommer de store problemer med midlertidige adgangskoder og velkomstbreve modtaget samme dag, fadæsen med nemid.dk og andre hændelser, som bestemt heller ikke har været med til at hæve NemIDs troværdighed. Uanset de ovenstående trusler mod sikkerheden i systemet må vi dog erkende, at NemID overordnet virker som en klar forbedring af sikkerheden, hvilket primært må tillægges den ægte to-faktorsikkerhed og den øgede uafhængighed af sikkerheden på den enkelte borgers PC.

6.1.3 Relation til Userminds' testrapporter

Som sidste punkt inden opsummeringen vil vi i dette afsnit se nærmere på den testrapport, som firmaet Userminds har udarbejdet, og som blev præsenteret i afsnit 3.5.2. Selvom vi ikke kender de direkte forudsætninger, som Userminds' test er baseret på, vil vi alligevel tillade os at relatere vores undersøgelser til deres. Dette gør vi primært for at illustrere nogle af de forudsætninger og teknikker, som muligvis kunne have været med til at undgå en stor del af de mange usabilityproblemer, som vi har identificeret i dette speciale.

Overordnet signalerer det udleverede materiale, at IT- og Telestyrelsen og DanID har gjort sig stor umage for at teste de forskellige aspekter af produktet inden lancering. Vi har en del kommentarer og indvendinger, som vi gengiver herunder. Mange af forholdene vil givetvis være fastlagt i de rammer, som Userminds har arbejdet under, og disse skal de selvfølgelig ikke klandres for. Skal vi relatere Userminds' opgaver til vore egne, har vi følgende overlap af opgaver:

- Oprettelse af NemID
- Oprettelse af adgangskode
- Oprettelse af brugernavn
- Tilknytning af digital signatur til NemID
- Logon med adgangskode og brugernavn
- Logon med nøgle

[36] DanID har dog efterfølgende tilføjet muligheden for at deaktivere login med CPR-nummer via selvbetjeningen på NemID.nu. De er dog gået meget stille med dørene omkring denne udvidelse, så det er uvist, hvor stor en effekt det rent faktisk vil have for den almindelige bruger.

På ovenstående liste mangler et- og to-faktorsignering med NemID, men af hensyn til testpersonernes privatliv og ønsket om ikke at tvinge dem til at foretage ufrivillige eller irreversible handlinger, har dette ikke været en mulighed i vores tests.

- **Lokation og forbehold for tænke-højt**
  På samme måde som i vores egne undersøgelser er konceptet for tænke-højt selvfølgelig også en smule kunstigt i Userminds' test. Den store forskel er dog, at der her er testet i et usability-laboratorium og med samme computer, styresystem (Windows) og browser (interessant nok Firefox) for samtlige testpersoners vedkommende. På dette punkt vil vi mene, at vores tests er mere virkelighedsnære, idet vi pånær et enkelt tilfælde har foretaget vores undersøgelser i testpersonens vante omgivelser og på egen computer. Userminds' procedure må anses for at være standarden inden for denne type test, men da det testede system skal anvendes i så bredt omfang og stiller så forholdsvis høje krav til softwaren på computeren, burde man måske nok have forsøgt sig med flere forskellige typer brugermaskinel. Vi må desuden antage, at der forud for disse brugervenlighedstests har ligget et stort arbejde med system-tests på forskellige platforme, og at dette selvfølgelig ikke er Userminds' primære fokus.

- **Snæver aldersgruppe**
  Selvom Userminds har testet på hele 22 testpersoner (hvilket er ret mange i THT-sammenhæng), er aldersgruppen indsnævret til 25-59 år. Igen har vi af omveje erfaret, at DanID tilsyneladende også har testet systemet på en gruppe ældre mennesker, men noget sådant har vi trods vores aktindsigt ikke fået udleveret materiale om. Det er dog næsten sikkert, at DanID har undladt at teste systemet på den yngste og måske umiddelbart mest kritiske målgruppe. Her fandt vi i vores undersøgelser nogle af de mest kritiske task failures.

- **Manglende "live-test"**
  De udførte tests er som sagt foretaget på en prototype, og derfor har systemet i sagens natur været nogenlunde fri for overbelastning og kryptiske fejlmeddelelser. Vi savner dog en test af systemet i mere realistiske omgivelser – hvordan kan DanID for eksempel have overset den særdeles skjulte måde at tilknytte OCES-certifikat for begge vores unge testpersoner, eller de manglende fejlbeskeder om genbrugt midlertidig adgangskode, som først kom frem det tredje sted, testpersonen forsøgte at starte aktiveringen.

- **Testleders involvering**
  Ud fra spørgerammen bagerst i testrapporten virker det i vores øjne, som om testlederen har spurgt og guidet testpersonerne mange gange undervejs i opgaveløsningen. Dette både for hele tiden af spørge ind til testpersonernes holdning til skærmbilleder eller systemets opførsel, samt for at guide testpersonen videre til næste trin i opgaven. Et eksempel fra spørgerammen:

  > "Hvilke muligheder har du på siden? *[Forstår man mulighederne?]*
  > Hvad synes du om mulighederne? Hvorfor? *[Synes man det er relevant at tilknytte den digitale signatur til sit NemID?]*
  > Hvad vil du gøre hvis du er i tvivl om hvad du skal gøre her, hvor vil du søge hjælp? *[Vil man bruge support linket der ligger udenfor appletten?]*

> Du må godt forsætte. *[...]* Jeg kan se at du valgte [Ja, Nej], hvorfor gjorde du det? *[...]*
> Hvad synes du om den information du får her? Beskriv, hvordan du forstår den information, du får her Hvad sker der når du vælger / ikke vælger at tilknytte digital signatur til dit NemID?
> Jeg vil bede dig om at vælge *[Ja eller nej – det modsatte af hvad brugeren lige har valgt]."*
>
> Pedersen og Holst [27], s50n

På denne måde går testforløbet fra at være klassisk observerende til at være meget mere styrende fra testleders side. Lidt ligesom at indlægge ultrakorte debriefinger efter hvert eneste trin i en opgave. Userminds har selvfølgelig fået til opgave at teste samtlige aspekter af systemet, men dette kan opnås ved at gøre opgaverne forskellige fra gang til gang – at indlægge mere styring i opgavens formulering og mål. Den anvendte metode er selvfølgelig med til at maksimere mængden af information givet fra testpersonen, men det er bestemt også medvirkende til at gøre forløbet mere instrumentelt og unaturligt – noget som vi bevidst har forsøgt at undgå i vores undersøgelser for at gøre dem mere realistiske og strømlinede.

- **Bestilling af ekstra nøglekort**
  I deres test af NemID.nu konstaterer Userminds, at "deltagerne har let ved at bestille ekstra nøglekort. Der er ikke identificeret problematikker specifikt knyttet hertil." Det kan Userminds jo have ret i, og vores tests viste da heller ikke større problemer med selve bestillingen – så snart testpersonerne fandt den rigtige hjemmeside. Dette er endnu et eksempel på, at de udførte tests kunne være foretaget tættere på det realistiske scenarie, hvor brugeren starter på en tilfældig hjemmeside. Vi anser det heller ikke for realistisk, at en bruger skal foretage 4-5 forskellige handlinger i træk på selvbetjeningen på NemID.nu (som testen har simuleret), men blot skal derind en sjælden gang for eventuelt at bestille et ekstra nøglekort.

- **Klassifikation og gruppering af problemer**
  Userminds har valgt at anvende deres egen klassifikation af de 56 fundne usabilityproblemer. De har ligeledes valgt at gruppere problemerne efter alvorlighedsgrad snarere end efter overordnet emne. Denne rangordning er selvfølgelig kendetegnende for 'vi ordner de kritiske først'-mentaliteten, men vi vil ikke vurdere, hvorvidt vores egen anvendte metode er bedre her dog blot pointere, at er der et tilstrækkeligt antal semi-kritiske problemer i en bestemt gruppe, kan man efter vores mening tillade sig at hæve gruppens overordnede klassifikation. Der er i Userminds' tilfælde tale om test af et isoleret system, hvorimod vi har testet mere bredt og derfor haft et større behov for overblik i form af problemgrupperinger.

- **Primære identificerede udfordringer**
  Med hensyn til de primære identificerede problemområder er Userminds' og vores resultater tæt på hinanden. Userminds fremhæver forståelse af konceptet, utilfredshed med nøglekort samt sikkerheden i nøglekortet som de væsentligste mangler ved systemet, hvilket harmonerer udmærket med vores identificerede problemområder.

6.1.4 Opsummering

På baggrund af ovenstående diskussion vil vi nu forsøge at indsnævre og redefinere spørgsmålet om, hvorvidt NemID indtil videre er en succes. For at komme svaret nærmere, er vi nødt til at se på forskellige kriterier for 'succes'. Hvis vi umiddelbart snakker om en succes med hensyn til udbredelse, kan NemID ikke undgå at blive en succes. Tallene for januar siger over 2.3 millioner brugere, og antallet af anvendelser indtil videre er nået godt over 100 millioner. På seks måneder har NemID formået, hvad den gamle digitale signatur ikke kunne på 7-8 år, så på dette punkt må NemID siges at være en succes – man har dog ikke kunnet nå det ambitiøse mål med 3 millioner brugere inden udgangen af 2010, men der er ikke lang vej igen. Bagsiden af medaljen er her for mange brugere, at NemID er en 'tvunget succes'. Bankernes centrale rolle i den korte og særdeles agressive udrulningsstrategi, samt det faktum at de fleste netbanker omkring årsskiftet udelukkende understøtter NemID som login, er med til at presse brugerne til at anvende systemet. Læg hertil at staten og de offentlige tjenester gør noget lignende inden for ganske kort tid, samt at de gamle digitale signaturer ikke længere bliver fornyet, og man har en fuldbyrdet påtvunget løsning for borgere, som gerne vil udnytte øget selvbetjening på nettet. Fordelene for de parter, der er involveret i udvikling og implementering af NemID, er klare: øget selvbetjening frigiver værdifulde og dyre medarbejdertimer, mindsker papirarbejde, forsimpler arbejdsgange og mindsker sagsbehandlingstider på grund af øget digitalisering. Indtil videre har den store uforudsete ulempe så til gengæld været markant øgede udgifter til support, men på lang sigt vil vi umiddelbart vurdere NemID som en fornuftig investering.

Og fordelene for forbrugerne? 'Tvangsløsning' lyder umiddelbart voldsomt, for hovedparten af forbrugerne er som sagt tilfredse med systemet eller gider ikke brokke sig, for hvad nytter det? I vores øjne gives forbrugerne klart en fordel, i og med at det nu er den samme løsning, der i princippet vil kunne anvendes til nærmest alle transaktioner eller beslutninger på nettet. *Learnability* vil naturligvis også gå stærkere, jo flere anvendelser der er tale om. Vel at mærke kun hvis brugerne bliver ved med at anvende selvbetjeningsløsninger i samme omfang som før – noget som tallene fra Danske Banks brugerpanel giver blandede udsigter for. dog vil der efter vores mening også komme flere private selskaber, der skifter til NemID, selvom der har været stor kritik af priserne for at koble sig på systemet. For nylig bekendtgjorde Den Blå Avis, at de fremover tilbyder logon med NemID på deres højtbesøgte site. Sådanne tiltag vil givetvis være med til at øge forbrugernes incitament til at anvende NemID.

Det ændrer dog ikke på, at selvom der er muligheder nok for anvendelse, så kan systemet på ingen måde siges at være vellidt. Både vores egne samt flere eksterne undersøgelser bekræfter, at NemID scorer middelmådigt på de fleste parametre, og nøglekortet bærer en stor del af skylden for dette. Havde DanID valgt den elektroniske løsning i stedet (uagtet deres indledende undersøgelser), havde der efter vores mening været en langt mere positiv holdning til systemet. DanID er også stået overfor den massive opgave at forklare danskerne, at systemet er sikkert, og at nøglekortet umiddelbart er den bedste løsning – både økonomisk, praktisk og sikkerhedsmæssigt. Denne opgave har DanID ikke kunnet løfte tilfredsstillende, i vores øjne til dels på grund af en for ambitiøs udrulningsperiode og for mange uheldige smuttere undervejs. Medierne har uheldigvis prompte fulgt trop i kritikken, sandsynligvis båret af dårlige erfaringer fra

tidligere offentlige IT-projekter, og har efter vores mening haft en afgørende betydning for NemIDs middelmådige *subjective satisfaction*. Systemet i sig selv lider af nogle fejl og uhensigtsmæssigheder, men er efter vores mening på de fleste punkter udmærket og usability-mæssigt og sikkerhedsmæssigt en klar forbedring i forhold til den gamle digitale signatur – vel at mærke så snart man kommer i gang med det. Det har været skræmmende og må anses som uacceptabelt at se det i vores øjne store antal testpersoner og respondenter, der ikke selvstændigt har kunnet aktivere deres NemID, men umiddelbart er det svært at genkende et mønster i fejlene. Sikkerhedsmæssigt er der ligeledes konstateret store mangler i kommunikationen og forståelsen for systemets virkemåde, og in mente skal selvfølgelig holdes pointen med alle æg i samme kurv. Den sidste vil vi lade være op til den enkelte forbrugers sikkerhedsopfattelse – under alle omstændigheder kan vi ikke tilslutte os konspirationsteoretikernes frygt. En god start ville her være at offentliggøre mere letforståelig dokumentation og forklaring af NemIDs virkemåde og sikkerhed. Noget, som vi har haft svært ved at opdrive til dette projekt.

Vi afrunder dette afsnit med en kommentar fra Peter Lind Damkjær på det afsluttende spørgsmål om, hvornår NemID bliver en fuldblodet succes:

> *"Det* er *for mig at se en fuldblodet succes. Vi har rullet en infrastruktur ud på et halvt år, hvor de undersøgelser, vi har lavet, IT- og Telestyrelsen har lavet, Danske Bank har også lavet en undersøgelse, viser, at langt hovedparten af brugerne er enten meget tilfredse eller tilfredse med denne her løsning. Tendensen er, at jo længere tid man har haft det, hvor man kan se synergieffekterne af, at man kan gå ind flere forskellige steder med den samme løsning, jo bedre. Så det er for mig at se en succes, hvor vi har rullet en løsning ud, som i store træk fungerer – vi har haft nogle driftsnedbrud i forbindelse med udrulningen her. Vi har haft nogle fejl, som jeg meget meget gerne ville have været foruden omkring forsendelse af PIN-breve og midlertidige adgangskoder. Men bortset fra det så har vi en infrastruktur, som faktisk er rimelig stabil og er accepteret af langt hovedparten af de brugere, der har fået det her, mere eller mindre uden at de har valgt det selv. Vi har folk, der kommer fra Japan for at se, hvad det her det er for noget og er vildt begejstrede over, at vi har kunnet lave det her."*
>
> Peter Lind Damkjær, DanID (Bilag P)

6.2 Diskussion af metode

Efter at have diskuteret de fremkomne resultater vil i dette afsnit for hver undersøgelse kritisk diskutere metodevalget og sammenholde dette med Userminds' fremgangsmåde. Vi vil også komme med gode råd og *best practice* baseret på de erfaringer, vi har oparbejdet undervejs i processen.

6.2.1 Tænke-højt test

Overordnet er vi tilfredse med planlægning og gennemførsel af de udførte THT'er. Som vi også nævnte tidligere, havde det været passende med en mere bred fordeling af testpersoner på de fire definerede grupper, men på trods af de små udsving i respondentgruppen, har testpersonerne i høj grad vist sig at kunne bidrage til en meget bred og forskellig

identifikation af usabilityproblemer i systemet. Jævnfør diskussionen om antallet af testpersoner ville de ni til denne test ikke være nok – der blev stadig hos P09 fundet nye usabilityproblemer, og man kunne derfor sagtens have inkluderet endnu 2-3 testpersoner. Vi står med andre ved vores valg af metode og kunne som eneste anke have tænkt os en bedre fordeling af testpersoner og gerne et større antal.

En almindelig anke ved brug af THT'er er den kunstige situation, som undersøgelsen uvægerligt vil præsentere testpersonen for. I dette tilfælde er det dog vores opfattelse, at dette ikke har haft nogen nævneværdig indflydelse på testpersonerne – tværtimod. Testleder og observatør har forholdt sig så neutralt som muligt for at give testpersonen plads til at være sig selv. En meget væsentlig faktor heri er også valget af, at testpersonerne i videst mulig omfang har deltaget i undersøgelsen på deres egne forudsætninger: i vante omgivelser og med deres egen computer til rådighed. Dette har efter vores mening skabt større tryghed hos testpersonen og fået vedkommende til at opfatte opgaven som mere personlig og relevant. Userminds har i deres tests placeret testpersonerne i et usabilitylab og bedt dem anvende en fremmed (men altid den samme) computer – noget som de også selv pointerer som forbehold. I nærværende undersøgelse er disse forbehold med succes forsøgt minimeret.

For at bidrage en smule til *best practice* kan vi ikke understrege nok, hvor stor en betydning forberedelsen af denne type test har. Selvom de udførte tests må siges at ligge et sted mellem *discount* og *full-blown*, har det været særdeles givtigt at kunne præsentere en velforberedt og struktureret undersøgelse for de deltagende testpersoner. Det har ligeledes været en fordel i den efterfølgende behandling af resultaterne, at der er blevet grundigt tænkt over brugen af de opsamlede data *inden* undersøgelsen, ligesom det i den forbindelse klart kan anbefales at være mindst to om at udføre sådanne undersøgelser – en enlig testleder havde ikke kunnet observere eller notere i nærheden af det antal problemer, som denne undersøgelse har kastet af sig.

En væsentlig forudsætning for vellykkede tests er testlederens opførsel og mængden af hjælp givet undervejs. For nogle af de ni testpersoners vedkommende har det været nødvendigt med flere og store interventioner fra testleders side, ligesom en del hjælp har været nødvendig undervejs. Rollen som testleder er særdeles svær, og arbejdet med at teste NemID har været uforudsigeligt – grundet systemets uventede opførsel er vi flere gange blevet kastet ud i situationer, som normalt ville have sat testen i stå uden at testpersonen havde kunnet gennemføre. Hverken Kasper Hansen eller undertegnede er rutinerede testledere, og det har været en del af udfordringen at få vores forskellige implementationer af rollen til at hænge sammen for at skabe konsistente resultater.

CUT-seancerne De indlagte CUT-seancer har vist sig særdeles givtige i forbindelse med at afdække forståelsesmæssige problemer og holdninger til koncepter, og som dermed kan skabe unødig bekymring eller overraskelse. Det havde været interessant at foretage tests med *og* uden CUT-seancer for at sammenligne det direkte udbytte af seancerne, men dette har ikke været vores primære fokus med de udførte tests. Desuden har vi på grund af CUT-seancernes særskilte natur nemt kunne adskille de i seancerne identificerede problemer fra de resterende, hvilket tydeligt har vist deres potentiale som problemidentificerende redskab i forlængelse af konceptuelt tunge, men tidsmæssigt

korte opgaver. Udover at have bidraget med unikke identificerede problemer, har CUT-seancerne efter vores mening bidraget med større præcision og uddybning set i forhold til en almindelig debriefing, ligesom de har været med til identificere problemer helt op i klasse Q.

6.2.2 Første spørgeskema

Til det første spørgeskema har vi ikke mange kommentarer til metoden, da den overordnet set har udført sin opgave fuldt ud. Spørgeskemaet har leveret de forventede resultater, og der har været overraskende få (hvis nogen) fejltolkninger af de stillede spørgsmål. Dette er uden tvivl på grund af de mange timer der er lagt i research og udarbejdelse af skemaet, men dette har i sidste ende betydet en minimering af tolkbare resultater og en mere direkte og ikke-arbejdskrævende efterbearbejdning af de modtagne data. Vi vil desuden igen komme med en kraftig anbefaling af Google Docs som redskab til udformning af skema og indsamling af data – værktøjet er gratis og er i konstant udvikling og udbredelse og har efter vores mening stort potentiale inden for gratis online-undersøgelser.

Best practice for elektronisk rekruttering Den nok største udfordring ved spørgeskemaet har været rekrutteringen af respondenter. Vi valgte efter nogle mislykkede forsøg med fagforeninger at distribuere skemaet udelukkende gennem sociale medier og internetbaserede fora. Dette viste sig at være en farbar, men ikke helt ukompliceret vej, og vi vil derfor nu opstille nogle forslag til huskeregler eller kriterier for elektronisk rekruttering af respondenter via sådanne medier:

- **Credibility:** Det skal i videst mulig omfang skinne igennem i invitationen, at der er tale om en seriøs undersøgelse fra en respektabel institution. Selvom det kortlink, som vi af praktiske årsager anvendte, gjorde det lettere at henvise til undersøgelsen, lagde det også et slør over, hvor det henviste til. Her ville det have været bedre at gengive det direkte link, og endnu bedre at have fået tildelt et 'officielt' link på KUs servere, så de potentielle respondenter havde mulighed for at gennemskue afsenderen af undersøgelsen og afkræfte enhver tvivl om lumske hensigter som f.eks. phishing. Credibility er et mindre problem på Facebook, hvor invitationen i første omgang udsendtes til en velkendt vennegruppe, som kunne bekræfte undersøgelsens autenticitet, og derefter kunne videresende invitationen og den nu tilknyttede autenticitet (meget i stil med hvordan PGP[37] fungerer) og opbygge en form for 'trusted network'. Flere forumbrugere udtrykte deres mistillid til det postede link, men så snart en anden forumbruger havde vovet sig ud i at gennemføre undersøgelsen og berettede herom, forstummede mistilliden.
- **Transparency:** Det skal tydeligt fremgå af invitationen, hvem afsenderen er, og hvad undersøgelsen går ud på. De fleste fora på nettet kræver oprettelse af en brugerprofil for at kunne poste indhold – her vil det være en fordel at give så fyldestgørende og officielle informationer som muligt, så man sætter den mulige respondent i stand til at vurdere ens baggrund nærmere. Det er omvendt selvfølgelig heller ikke meningen, at spørgeren skal udlevere samtlige personlige oplysninger

[37] PGP: Pretty Good Privacy. Et meget udbredt open-source standard for kryptering af f.eks. mails, oprindeligt baseret på et *web of trust*. Se mere på http://en.wikipedia.org/wiki/Pretty_Good_Privacy.

om sig selv. Igen har Facebook her en fordel på førstnævnte punkt, hvorimod offentlige fora har et hav af mere eller mindre ukendte brugere uden reel tilknytning til andre brugere på sitet.

- **Anonymity:** Respondenternes anonymitet skal garanteres tidligst muligt i forløbet, særligt når invitationen spredes på offentligt tilgængelige fora med navngivne (men ikke nødvendigvis indbyrdes velkendte) brugere.

- **Brevity:** Fat dig i korthed. De fleste brugere af offentligt tilgængelige tjenester vil hurtigt forbigå invitationen, hvis den kræver mere end 10-20 sekunder at gennemskue meningen bag. Kortfattetheden bør dog altid vægtes i forhold til de resterende kriterier. En måde at implementere kortfattetheden på er ligesom i denne undersøgelse at linke ikke til selve spørgeskemaet men til en hjemmeside med mere information, så besøgende ud fra denne mere fyldige beskrivelse kan vurdere, om de har lyst eller mulighed for at deltage i undersøgelsen.

- **Relevance of recruitment:** Invitationer til undersøgelser skal postes, hvor der forefindes relevante respondentgrupper. Det nytter ikke meget at poste en invitation til en læserundersøgelse af et mandemagasin på et forum for vordende mødre. Brug eventuelt en linie på at beskrive, hvorfor netop læseren af denne post er interessant for undersøgelsen. Her kan ligeledes beskrives eventuelle præmier eller kompensation for deltagelse i undersøgelsen.

- **Information usage:** Det skal fremgå tydeligt, til hvad og hvordan de indsamlede informationer bruges. Dette kriterium er vigtigt både af hensyn til respondentens tryghed men kan også være med til at illustrere relevansen af undersøgelsen – hvorfor det er vigtigt, at den enkelte respondent bør vælge at deltage. Det kan med andre ord være med til at 'overtale' den potentielle respondent.

- **Informed consent:** En sidste huskeregel er at gøre respondenten opmærksom på, hvad vedkommende giver sit samtykke til ved deltagelse i undersøgelsen. Dermed ikke sagt, at der skal udlægges en fuld samtykkeerklæring, men måske blot et kort "Jeg har læst, forstået og accepteret beskrivelsen og vilkårene for undersøgelsen og bekræfter hermed min deltagelse". En sådan samtykkeerklæring vil dog som oftest finde plads i starten af selve undersøgelsen eller på førnævnte introduktionsside.

Relevance of recruitment er i sig selv omfattende, da det er notorisk svært at finde fora på nettet, som giver mening for bestemte undersøgelser. Dette på trods af, at der rundt omkring findes fora (og Facebook-grupper) for næsten alle tænkelige emner – den store udfordring er at finde dem. Heldigvis har langt de fleste fora intet imod posts af denne art (de strider ikke imod forummets retningslinier), så længe invitationen udstråler høj *relevance of recruitment*. En anden overvejelse, man kan gøre sig her, er, hvordan man breder en sådan undersøgelse ud til folk, der kun sjældent anvender internettet.

I forhold til rekrutteringsmetoden er vi umiddelbart positivt overrasket over, at vi har kunnet få så forholdsvis mange besvarelser udelukkende fra Facebook og elektroniske fora på nettet. Der kunne sandsynligvis have været en del flere, hvis vi havde fået lov til at distribuere et link gennem en organisation eller virksomhed. Det har dog været et særdeles interessant og udfordrende projekt, som mange forhåbentlig vil afprøve senere

hen, også i mere struktureret form. De ovenstående pointer kan forhåbentlig være en hjælp hertil.

6.2.3 Opfølgende spørgeskema

Givet den forholdsvis høje svarprocent på 71.7 for det opfølgende spørgeskema, er vi også her ganske godt tilfredse. I forhold til det første spørgeskema har der her dog været visse tilfælde, hvor respondenterne formodentlig har misforstået det første spørgsmål (antallet af brugte koder), hvilket derfor skulle være formuleret anderledes. En anden væsentlig anke er, at vi, som tidligere nævnt, ikke har haft mulighed for at korrellere den enkelte respondenters svar, da vi ikke har kunnet 'genkende' respondenter mellem de to spørgeskemaer. Man kunne have gjort dette ved at tildele hver enkelt respondent i første spørgeskema et identifikationsnummer, som man så sendte ud som et særligt link til det opfølgende skema, men det ville næsten sikkert føre til mere arbejde på respondentens side eller uforholdsmæssigt meget fra vores side. Det blev derfor tidligt i forløbet fravalgt, men retrospektivt fortryder vi dette valg. Det opfølgende skema har ikke givet det udbytte, man kunne have forventet, og konklusionerne ud fra besvarelserne er da også i bedste fald tendentiøse og ikke statistisk underbyggede. De giver dog stadig interessante informationer omkring den daglige brug af NemID samt giver et hint om *learnability* og den mere langsigtede *subjective satisfaction* for produktet.

7 Konklusion

Vi har i dette speciale set nærmere på brugervenligheden af NemID. Indledningsvist har vi set, at der teknisk er stor forskel på, hvordan den digitale signatur og NemID fungerer, at NemID (umiddelbart) er sikrere, og at NemID på ganske kort tid har nået større, omend nærmest tvungen, udbredelse end den digitale signatur har kunnet på 7-8 år. NemID bør ikke anses som et fuldstændig sikkert system, og der er stadig mulighed for forholdsvis simple angreb, som dog er afhængige af brugerens generelle opførsel og kendskab til online-tjenester. Vi har ligeledes set på den kritik, der har været af NemID i medierne og forsøgt at udrede nogle af de misforståelser, som allerede har formet sig om dette komplekse system. I forhold til teknik, sikkerhed og udbredelse kan NemID efter vores mening derfor kaldes en succes og en forbedring.

Som specialets primære omdrejningspunkt har vi planlagt og gennemført to forskellige typer evalueringer af NemID: tænke-højt og elektroniske spørgeskemaer i to stadier. Undersøgelserne er bygget på solidt fagligt grundlag og anvender videnskabeligt anerkendte metoder til indsamling og bearbejdning af data. Vi er desuden kommet med best practice i forhold til udelukkende online distribution og udformning af spørgeskemaer. Undervejs i undersøgelserne har vi afprøvet *cooperative usability testing*, som i kraft af dens umiddelbare placering efter korte men konceptuelt komplekse opgaver har været i stand til at identificere usabilityproblemer, som ellers ikke ville være blevet opsamlet i debriefingen, hvor den konkrete opgave for testpersonen er tidsmæssigt længere væk. Vi har dermed vist, at CUT-teknikken kan anvendes som et godt supplement til mere traditionel tænke-højt, i direkte forlængelse af kortere opgaver og ikke kun i form af den længere efterfølgende seance, som teknikken oprindeligt lægger op til.

På trods af DanIDs egne tests har vi under analysen af resultaterne identificeret 185 unikke usability-relaterede observationer, hvoraf for mange i vores øjne må anses som alvorlige. De identificerede observationer viste sig under grupperingen primært at dreje sig om en mangel på forståelse af de bagvedliggende koncepter, single-signon, nøglekortet og en generelt middelmådig tilfredshed med systemet. En del kritiske problemer relaterede sig direkte til aktiveringen, som henholdsvis fire ud af ni (tænke-højt) og ca. 10% (spørgeskemaer) ikke selvstændigt kunne gennemføre. Begge tal indikerer i vores øjne generelle problemer med systemet – noget som lignende undersøgelser og tal bekræfter.

Særligt nøglekortet og det ikke direkte til NemID hørende single-signon er blevet vurderet som henholdsvis irriterende, uforståeligt og overraskende. DanID og IT- og Telestyrelsen står her overfor nogle store udfordringer, hvis de ønsker at forbedre folks holdning til systemet samt den konceptuelle forståelse af systemets sikkerhed og virkemåde. Interessenterne bag systemet virker dog umiddelbart særdeles lydhøre overfor kritik og har indledt en del initiativer, der forhåbentlig kan være med til at ændre befolkningens holdning inden for den kommende tid, ligesom en ikke uvæsentlig del af utilfredsheden må kunne tilskrives generel tilvænningstid. Vi anser dette speciale som et konstruktivt indspark i denne proces, men det ændrer dog ikke på, at sammenfatter man ovenstående problemstillinger, kan vi i forhold til brugervenlighed på nuværende tidspunkt *ikke* kalde NemID en succes.

Litteratur

[1] M. Bosnjak og T. L. Tuten. Prepaid and Promised Incentives in Web Surveys - An Experiment. *Social Science Computer Review*, 21(2):208–217, 2003.

[2] J. Brooke. SUS: A quick and dirty usability scale. I P. Jordan, B. Thomas, B. Weerdmeester, og I. McClelland, redaktører, *Usability Evaluation in Industry*. Taylor and Francis, London, U.K., 1996.

[3] J. Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, April 1960.

[4] K. A. Ericsson og H. Simon. *Protocol Analysis: Verbal Reports As Data*. MIT Press, Cambridge, MA, 1993.

[5] E. Frøkjær og K. Hornbæk. Cooperative Usability Testing: Complementing Usability Tests with User-Supported Interpretation Sessions. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, side 1383–1386, 2005.

[6] S. Greenberg og B. Buxton. Usability Evaluation Considered Harmful (Some of the Time). I *Proceedings of the CHI 2008 Conference: Human Factors in Computing Systems*, side 111–119. ACM Press, New York, 2008.

[7] K. Hansen. Usability-evaluering af NemID, November 2010. Kandidatprojekt ved Datalogisk Institut, Københavns Universitet.

[8] M. Hertzum, K. D. Hansen, og H. H. Andersen. Scrutinizing usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28(2):165–181, 2009.

[9] K. Hornbæk og E. Frøkjær. Comparison of techniques for matching of usability problem descriptions. *Interacting wih Computers*, 20(6):505–514, 2008.

[10] IT- og Telestyrelsen. Digital Signatur - Infrastrukturen til digital signatur, 2002. URL https://www.signatursekretariatet.dk/pdf/vejledninger/infrastruktur.pdf. Sidst tilgået: 31. januar 2011.

[11] IT- og Telestyrelsen. Digital Signatur - OCES – en fælles offentlig certifikat-standard, 2002. URL https://www.signatursekretariatet.dk/pdf/vejledninger/oces.pdf. Sidst tilgået: 24. februar 2011.

[12] IT- og Telestyrelsen. Digital Signatur - Forudsætninger og fordele, 2002. URL https://www.signatursekretariatet.dk/pdf/vejledninger/forud.pdf. Sidst tilgået: 24. februar 2011.

[13] IT- og Telestyrelsen. Undersøgelse af den danske befolknings holdning til it-sikkerhed, 2006. URL http://www.itst.dk/nyheder/filarkiv/befolkningsundersogelse%20til%20offentliggorelse%20-%20pdf-format.pdf. Sidst tilgået: 27. februar 2011.

[14] IT- og Telestyrelsen. Certifikatpolitik for OCES-personcertifikater (Offentlige Certifikater til Elektronisk Service), 2009. URL https://www.signatursekretariatet.dk/pdf/ca/POCES%20Certifikatpolitik%20version%204%20signeret.pdf. Sidst tilgået: 31. januar 2011.

[15] IT- og Telestyrelsen. Sikkerhedskonceptet vedrørende NemID, 2010. URL http://www.digitalsignatur.dk/visArtikel.asp?artikelID=1065. Sidst tilgået: 30. januar 2011.

[16] IT- og Telestyrelsen. Redegørelse fra taskforce til kvalitetssikring af NemID, 2011. URL http://www.ft.dk/samling/20101/almdel/uvt/bilag/79/941625.pdf. Sidst tilgået: 30. januar 2011.

[17] N. E. Jacobsen, M. Hertzum, og B. E. John. The Evaluator Effect in Usability Tests. *CHI 98 conference summary on Human factors in computing systems*, April 1998.

[18] G. Lindgaard og J. Chattratichart. Usability Testing: What Have We Overlooked? I *Proceedings of the CHI 2007 Conference: Human Factors in Computing Systems*, side 1415–1425. ACM Press, New York, 2007.

[19] Ministeriet for Videnskab, Teknologi og Udvikling. Lov om elektroniske signaturer, 2000. URL https://www.retsinformation.dk/Forms/R0710.aspx?id=6193. Sidst tilgået: 10. marts 2011.

[20] Ministeriet for Videnskab, Teknologi og Udvikling. Aktstk. 205: Indgåelse af kontrakt med DanID om digital signatur i 2008-2014, 2008. URL http://www.ft.dk/samling/20072/aktstykke/aktstk.205/aktstykket/595003.pdf. Sidst tilgået: 31. januar 2011.

[21] E. Møberg, A. H. Danielsen, T. T. Damgaard, S. Christensen, S. Weeke, N. Bundgaard, A. M. Christoffersen, E. Jørgensen, J. Primault, P. Nørregaard, M. Kandel, M. Eberhard, og M. Hartmann, redaktører. *IT i praksis 2010*. Rambøll Management Consulting A/S, 2010.

[22] R. Molich. User testing, discount user testing. http://www.dialogdesign.dk, 2003. Sidst tilgået: 17. marts 2011 (fra http://printfu.org/rolf+molich).

[23] R. Molich. Regler for standardtest af brugervenlighed, 2003. URL http://www.dialogdesign.dk/tekster/regler.pdf. Sidst tilgået: 3. februar 2011.

[24] R. Molich og J. S. Dumas. Comparative Usability Evaluation. *Behaviour & Information Technology*, 27(3):263–281, 2008.

[25] J. Nielsen. Why you only need to test with 5 users. http://www.useit.com/alertbox/20000319.html, Marts 2000. Sidst tilgået: 17. marts 2011.

[26] A. A. Ozok. Survey Design and Implementation in HCI. I A. Sears og J. A. Jacko, redaktører, *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, kapitel 58, side 1151–1170. Lawrence Erlbaum Associates, 2. udgave, 2008.

[27] C. D. Pedersen og T. Holst. Gennemførelse af prototype test af NemIDs logon applet. Technical report, Userminds, Juni 2009. Udleveret ved aktindsigt.

[28] C. P. Pfleeger og S. L. Pfleeger. *Security in Computing*. Prentice Hall, 4. udgave, 2007.

[29] R. Rivest, A. Shamir, og L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, 21:120–126, 1978.

[30] C. Robson. *Real World Research*. Blackwell Publishing, 2. udgave, 2002.

[31] J. Rubin og D. Chisnell. *Handbook of Usability Testing*. Wiley Publishing, Inc., 2. udgave, 2008.

[32] B. Schneiderman og C. Plaisant. *Designing the User Interface*. Addison-Wesley Publishing Company, USA, 4. udgave, 2005.

[33] B. Schneier. *Secrets & Lies: Digital Security in a Networked World*. John Wiley & Sons, 2000.

[34] D. H. Stone. Design a questionnaire. *British Medical Journal*, 307(6914):1264–1266, 1993.

[35] T. Tullis og B. Albert. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann Publishers, 2008.

[36] T. Tullis, S. Flieschman, M. McNulty, C. Cianchette, og M. Bergel. An Empirical Comparison of Lab and Remote Usability Testing of Web Sites. I *The Usability Professional's Association Annual Meeting*, Orlando, FL, 2002.

[37] T. S. Tullis og J. N. Stetson. A Comparison of Questionnaires for Assessing Website Usability. I *Proceedings of the UPA 2004*. UPA, Bloomingdale, IL, 2004.

[38] H. Udsen. Notat om anvendelsen af serverbaserede løsninger for kvalificerede elektroniske signaturer, 2009. URL https://danid.dk/export/sites/dk.danid.oc/da/dokumenter/henrik_udsen_notat.pdf. Sidst tilgået: 5. marts 2011.

[39] M. van Welie, G. C. van der Veer, og A. Eliëns. Breaking down usability. I *Proceedings of Interact '99*, 1999.

[40] T. Wu. The Secure Remote Password Protocol. I *Proceedings of the 1998 Internet Society Network and Distributed System Security Symposium*, side 97–111, Marts 1998.

[41] K.-P. Yee. User Interaction Design for Secure Systems. I *4th International Conference on Information and Communications Security*, side 278–290, 2002.

[42] T. Zhao og S. McDonald. Keep talking: an analysis of participant utterances gathered using two concurrent think-aloud methods. I *Proceedings of the NordiCHI 2010 Conference*, 2010.